Designing Low Power and Durable Digital Blocks Using Shadow Nanoelectromechanical Relays

Sadegh Yazdanshenas, Behnam Khaleghi, Paolo Ienne, and Hossein Asadi, *Senior Member, IEEE*

***Abstract*— Nanoelectromechanical (NEM) relays are a promising emerging technology that has gained widespread research attention due to its zero leakage current, sharp ON–OFF transitions, and complementary metal–oxide–semiconductor compatibility. As a result, NEM relays have been significantly investigated as highly energy-efficient design solutions. A major shortcoming of NEMs preventing their widespread use is their limited switching endurance. Hence, in order to utilize the low-power advantages of NEM relays, further device, circuit, and architectural techniques are required. In this paper, we introduce the concept of shadow NEM relays, which is a circuit-level technique to leverage the energy efficiency of the NEM relays despite their low switching endurance. This technique creates two virtual ground nodes in a block to allow: 1) a low power mode with functional NEM relays and 2) a normal mode with failed NEM relays. To demonstrate the applicability of this concept, we have applied it to a six-transistor SRAM cell as an illustrative example. We also investigate the applicability of this SRAM cell in field-programmable gate arrays and on-chip caches. Experimental results reveal that shadow NEM relays can reduce the power consumption of SRAM cells by up to 80% while addressing the limited switching endurance of NEM relays.**

***Index Terms*— Field-programmable gate arrays (FPGAs), nanoelectromechanical (NEM) relays, on-chip memory, shadow logic, static random access memory, switching endurance.**

I. Introduction

NANOELECTROMECHANICAL (NEM) relays are multiterminal mechanical switches that are electrically actuated. The most commonly studied NEM relays are four- and six-terminal laterally actuated relays, as shown in Fig. 1. The fundamental operation of all NEM relays is the same. A moveable beam is attracted (repelled) to (from) another terminal of the device by applying a particular voltage to the control terminal of each device to create (break) a connection between the beam and the pair terminal. The beam state in four- and six-terminal relays is solely determined by the voltage difference between the body and gate terminals and, hence, is independent of the source voltage [1], [2]. There is also a second group of NEM relays called vertically actuated relays in which the conducting channel is attached to the gate via an insulator. Applying specific voltage forces the gate to deviate downward and connect the channel to the drain and source terminals, resulting in a conductive path [3].

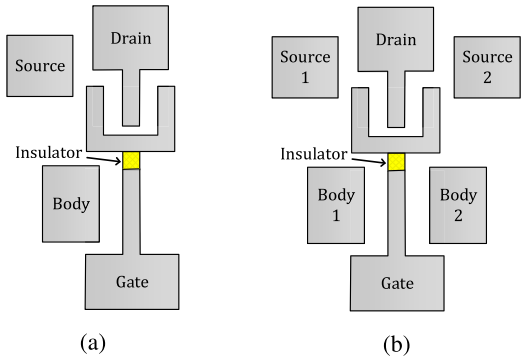


Fig. 1. (a) Four-terminal NEM relay. (b) Six-terminal NEM relay. The insulator causes the state of the relay to be independent of source voltage.

In comparison with the state-of-the-art complementary metal–oxide–semiconductor (CMOS) transistors, NEM relays exhibit zero leakage current, sharp ON–OFF transitions, CMOS compatibility, and low ON-state resistance [4]. Since these characteristics make the NEM relays promising for highly energy-efficient digital systems, there have been various efforts to employ them in a great variety of digital circuits, including combinational circuits [1], [5], [6], sequential elements [7], and field-programmable gate arrays (FPGAs) [8].

Despite their attractive features, NEM relays have two major drawbacks that prevent them from being used as reliable and high-speed alternatives for the CMOS technology, namely, limited switching endurance and slow switching speed. NEM relays suffer from a limited number of possible switching cycles. Potentially, the number of reliable switching operations in the NEM relays is reported to be 10^9 [9], which is the way too low for digital logic implementation. Existing prototypes exhibit even less reliable operations, which are reported to be about 10^6 [10]. In addition to the limited number of reliable switching operations, NEM relays also suffer from a slow switching operation due to their mechanical nature.

A majority of previous studies on employing NEM relays in digital circuits have mainly focused on using NEM relays in the delay-insensitive parts of the circuits, such as the configuration blocks of FPGAs [8], [11]–[14]. These studies have neglected the limited switching endurance of the NEM

Manuscript received November 4, 2015; revised February 6, 2016; accepted March 9, 2016. Date of publication April 29, 2016; date of current version November 15, 2016.

S. Yazdanshenas, B. Khaleghi, and H. Asadi are with the Department of Computer Engineering, Sharif University of Technology, Tehran 145888-9694, Iran (e-mail: syazdanshenas@ce.sharif.edu; behnam_khaleghi@ce.sharif.edu; asadi@sharif.edu).

P. Ienne is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: paolo.ienne@ep.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2016.2553106

relays, since the relays are employed as the configuration blocks. However, in many applications, FPGAs are dynamically reconfigured for numerous times [15], and hence, such FPGAs will have limited applicability. Another group of previous studies has designed the basic blocks of larger systems using NEM relays or has designed the hybrid CMOS-NEM blocks to be employed in larger systems. These designs include combinational logic blocks, SRAM and memory blocks, and other sequential blocks [16]. Such designs, however, suffer from the limited switching endurance of NEM relays.

This paper presents the concept of shadow NEM relays, which allows a digital block to operate in a low leakage mode with the functional NEM relays and a normal mode with the failed NEM relays. Shadow NEM adds an NEM relay on top of selected transistors in a particular design, so that the path to the ground signal can go either through the NEM relays or the selected transistors. The designed block is then controlled using an external circuitry to choose the route for the ground signal depending on the NEM relay switching capability. This can help leverage the energy efficiency of NEM relays despite their low switching endurance. The key design considerations required to apply shadow NEM relays to a circuit are detailed in this paper. To further elaborate this concept, a six-transistor (6T) SRAM cell enhanced with the shadow NEM relays is introduced. A low overhead external circuitry for this SRAM cell is also designed to show how it can be used in FPGAs and on-chip caches.

Our experimental results, carried out using HSPICE simulations, demonstrate up to 67% power savings in FPGA configuration cells and 80% in on-chip caches. While this amount declines as NEM relays start failing, it still allows the circuit to be fully functional despite failed switches. When employing the concept of shadow NEM relays in FPGAs, 5% area is added to each frame of configuration bits. This area overhead is 4% for a group of ten words in on-chip caches. Since NEM relays are stacked atop CMOS transistors, no area overhead is imposed to the original circuit after applying the idea of shadow NEM relays.

The rest of this paper is organized as follows. In Section II, we review the related work on a hybrid NEM-CMOS design. In Section III, the concept of shadow NEM relays is elaborated and it is further demonstrated with a design of an SRAM cell. The application of the SRAM cell in FPGAs and on-chip memory is also shown in this section. In Section IV, the experimental setup and the simulation results are detailed. Finally, the conclusion is drawn in Section V.

II. Related Work

NEM relays have attracted a significant amount of attention from the community. The most relevant works to ours are those creating hybrid CMOS-NEM blocks or create an NEM block to be used in a CMOS system. One of the most targeted areas in such designs is the design of FPGA configuration fabric [8], [11]–[14]. The main motivation behind such works is that the configuration fabric of FPGAs does not change during runtime, and hence, the shortcomings of endurance and high switching time are not an issue as long as the FPGA is not reconfigured frequently. These FPGAs, however, are unsuitable for the applications that demand dynamic reconfiguration. Dynamic reconfiguration is an important capability of FPGAs for which industrial tools have also been developed [15].

There have been several studies attempting to employ NEM relays in other application domains, such as combinational logic and memory blocks. Lee *et al.* [1] have proposed a method to implement combinational logic using 6T NEM relays. However, the proposed method suffers from the limited endurance of NEM relays and their mechanical switching delay. Its basic building block is also limited to a multiplexer and limits possible optimization that can be applied to circuits made using this scheme. Akarvardar *et al.* [5] and Chen *et al.* [6] had also previously proposed the design of logic gates using NEM relays, which also suffer from limited endurance. Venkatasubramanian *et al.* [7], [17] have designed the sequential and memory blocks using NEM relays, which have the same limitation. Chong *et al.* [18] have designed a hybrid CMOS-NEM SRAM cell that does not suffer from the mechanical switching delay, since it is not in the critical path of the design. The proposed cell still faces the problem of limited switching endurance. Other applications of NEM relays have also been proposed such as being used in the dc–dc converters [19] or being used as efficient sleep transistors [16].

In addition to various design attempting to create functional hybrid NEM-CMOS circuits, there have been various valuable efforts to fabricate novel NEM relays [2], [20]–[22]. Such studies can lead to high quality and industry-scale fabrication of NEM relays. Another category of studies is trying to provide accurate modeling for NEM relays [23]–[26]. These models can pave the way for future research by allowing more complex NEM-based circuits to be analyzed. The discussion of such studies is out of the scope of this paper.

III. Shadow Nanoelectromechanical Relays

In this section, we first detail the concept of shadow NEM relays along with the key design considerations required for applying it to any digital circuit. Afterward, an SRAM cell will be used as an illustrative example to further demonstrate the detail required to realize the concept of shadow NEM relays. Finally, the discussion on employing the proposed SRAM cell in FPGAs and on-chip caches will be presented.

A. Concept

The main idea behinds the concept of shadow NEM relays is adding an NEM relay on top of selected transistors in parallel with them in a particular design, so that the path to the ground signal can go either through the NEM relays or the selected transistors. We refer to such NEM relays as shadow NEM relays. The transistors are selected, such that they connect the rest of the circuit to the ground signal. Next, instead of connecting the subset of transistors or the NEM relays to the ground, they are connected to two separate virtual ground nodes. We refer to a node as the virtual ground if it could either be connected to the ground or be floating. The virtual ground nodes are then selectively routed to the actual ground based on NEM relays switching endurance. If the NEM

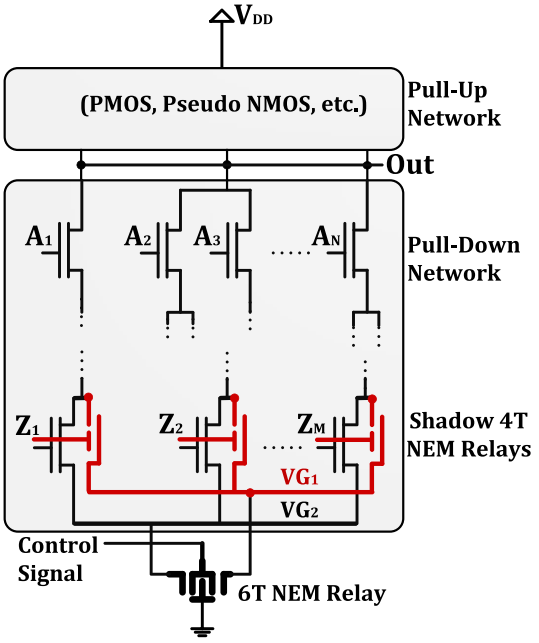


Fig. 2. Conceptual idea of shadow NEM relays.

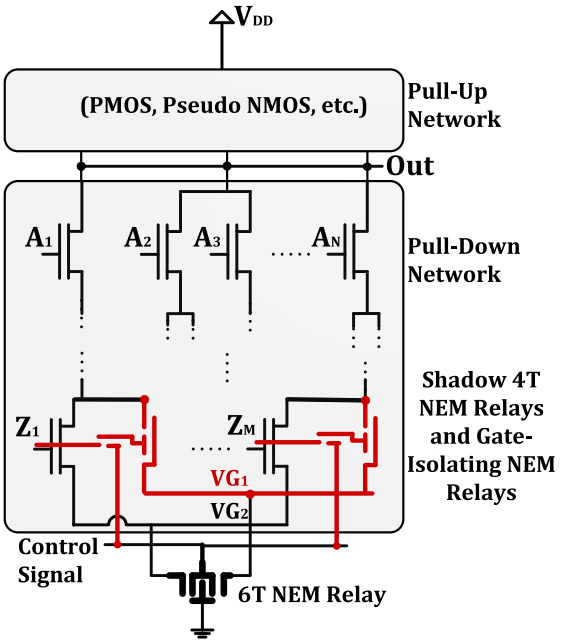


Fig. 3. Enhanced conceptual idea of shadow NEM relays along with additional gate-isolating NEM relays.

relays are still functional, the ground signal is connected to the NEM relays virtual ground. Otherwise, it will be connected to the transistors virtual ground. The main challenges here are the selection of the subset of transistors, the detection of NEM relays being worn out, and the design of an effective circuitry to control the connection of virtual grounds to the actual ground node. These challenges will be addressed in our illustrative example in Section III-B. The conceptual idea of shadow NEM relays is shown in Fig. 2.

The shadow NEM relay shown in Fig. 2 suffers from a major shortcoming. When the NEM relay fails due to exceeding its switching limit, two scenarios may happen. First, its gate, i.e., Z_i, which is also connected to the gate of its parallel nMOS transistor sticks to its source (i.e., VG_1); however, it does not cause any problem to the parallel nMOS transistor (which is enabled now), because when an NEM relay fails, the controller disconnects the VG_1 from the ground and makes it floating. However, the second scenario in which the common gate input Z_i sticks to NEM's drain causes a major problem, because the NEM's drain is shared with the nMOS transistor. In other words, in this scenario, the gate and the drain of nMOS will short. To overcome this limitation, we have added an additional NEM relay to isolate the gate of the shadow NEM relay from the gate of the transistor. The enhanced conceptual idea of shadow NEM relays is shown in Fig. 3. Upon failure, the same control signal that is used to change virtual ground connectivity will be used to isolate the gate terminals. Since this additional NEM relay only switches once upon the failure of the shadow NEM relay, its endurance will not become a bottleneck in the design.

The motivation behind the proposed idea of shadow NEM relays is providing the ability to leverage energy efficiency of NEM relays in the hybrid CMOS-NEM designs while preventing their low switching endurance from bringing about a failure in the designs. With the proposed scheme, the circuit begins its normal operation with NEM relays being functional at a minimum leakage power consumption. Upon a failure in one of the NEM relays, the subset of transistors that the NEM relay was acting as a shadow for takes over the normal operation in the absence of the NEM relay. This prevents the circuit failure with slight increase in power consumption. This can continue up to the point, where all NEM relay groups have a failed relay. At this point, the circuit has its highest leakage power consumption. Another aspect that adds the value to this scheme is the variations of endurance among different NEM relays. That is, a single NEM relay with lower switching endurance than average due to fabrication variations will not result in early failure of the circuit or substantial increase in power consumption. Rather, the subset of transistors to which that particular NEM relay belongs takes over and the circuit still benefits from the energy efficiency of other NEM relays. The concept of shadow NEM relays is orthogonal with the device-level improvements in NEM relays. Any future processing technique that further enhances NEM relay switching endurance or other characteristics can work along with this concept unless it completely annihilates the problem of switching endurance for all the NEM relays in a design.

B. SRAM Cell as an Illustrative Example

A typical 6T SRAM cell is shown in Fig. 4(a). As mentioned before, the first step to apply the shadow NEM relays to any arbitrary circuit is choosing a subset of the pulldown network that is directly connected to the ground node. For the SRAM cell, there are two such transistors. Next step is adding an NEM relay on top of each such transistor in a parallel fashion creating two separate virtual ground nodes. Finally, two additional NEM relays are added to isolate the

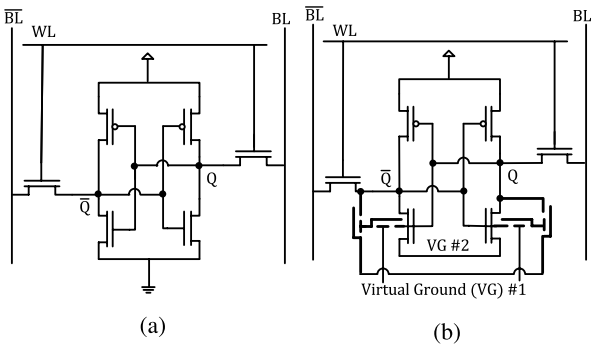


Fig. 4. (a) Typical 6T SRAM cell. (b) SRAM cell enhanced with shadow NEM relays.

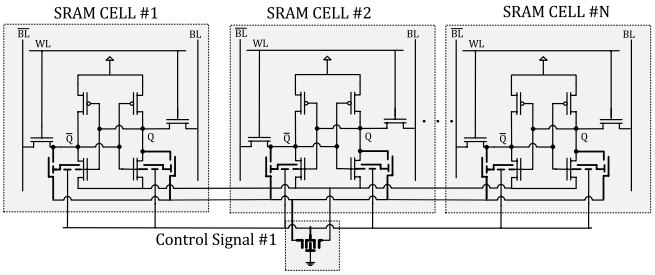


Fig. 5. Routing the actual ground to virtual grounds using a 6T NEM relay.

shadow NEM relay gates from the transistor gate upon the shadow NEM failure. The resulting SRAM cell is shown in Fig. 4(b).

In order to select which virtual ground node to connect to the actual ground, we group the SRAM cells and add a small NEM logic to each group to perform the selection. The grouping is carried out based on the circuit in which the SRAM cell is designed for and will be discussed in Sections III-D and III-E. The additional NEM logic will not be worn out, since it will switch much less frequently than the other NEM relays. There are two different methods that such grouping can be done. The first method, as shown in Fig. 5, uses a 6T NEM relay as a multiplexer to select which virtual ground to be connected to the actual ground. This method requires one of the virtual grounds to be grounded, while the other one is disconnected from the actual ground. The other ground routing method, as shown in Fig. 6, uses two 4T NEM relays to select which virtual ground is to be connected to the actual ground. This method is more flexible and allows either or both of the virtual grounds to be connected to the actual ground. This, however, requires two control signals and makes the controlling circuitry more complex. The advantage of this flexibility will be detailed in Section III-D along with the control circuitry used to generate control signals for both types of routing circuitries. It is worth mentioning that the routing NEM relay shown in both schemes could be more than a single relay. In either scheme, the controlling signal that indicates failure is connected to the gate-isolating NEM relays to isolate the shadow NEM relays upon its failure. It can be several NEM relays in parallel due to the current demand

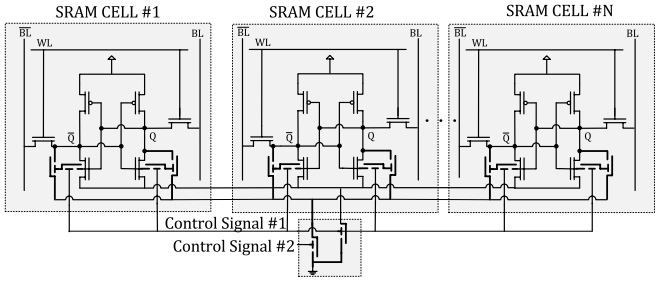


Fig. 6. Routing the actual ground to virtual grounds using two 4T NEM relays.

of the ground of SRAM cells without making the controlling circuitry more complex.

One may argue that the proposed scheme looks similar to using sleep transistors to eliminate the leakage current in part of a circuit. The distinction, however, is significant, because by using sleep transistors, a part of the circuit is unable to function and, hence, is considered as turned OFF. However, shadow NEM relays significantly alleviate the leakage current of transistors without turning OFF the circuit. It is worth mentioning that the shadow NEM relays cannot be considered as an alternative for sleep transistors. For example, in SRAM cells, sleep transistors should disconnect the voltage node from voltage supply and connect it to the ground node. Otherwise, a floating voltage node might lead to unwanted shorts in FPGAs [27].

C. Key Design Considerations

There are several key design considerations that should be addressed when using any design block enhanced with the shadow NEM relays.

1) The granularity of grouping of the designed cells with the common virtual ground nodes (as shown in Fig. 5) has to be determined depending on the circuit.
2) The detection mechanism for wear out of NEM relays has to be determined, so that it has a minimum impact on the overall circuit.
3) The method of routing the virtual ground nodes to the actual ground has to be chosen from the two options discussed earlier.
4) Finally, an efficient controller has to be designed to control the routing of virtual ground nodes.

In the remainder of this section, the above-mentioned design considerations will be demonstrated for FPGAs and on-chip caches.

D. Proposed SRAM in FPGAs

FPGAs are a high-speed platform for fast prototyping of digital systems that are configured by loading a specific bitstream into their configuration SRAM cells. Hence, SRAM cells are one of the key components in the state-of-the-art FPGAs. As mentioned before, the first point in utilizing the SRAM cell in a digital system is determining the granularity of grouping of the designed SRAM cells with the common virtual grounds. Since there is an embedded error detection

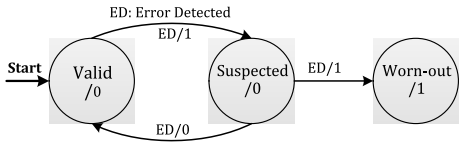


Fig. 7. Control FSM for the routing circuitry using a 6T NEM relay.

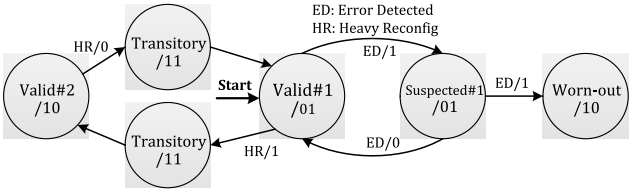


Fig. 8. Control FSM for the routing circuitry using two 4T NEM relays.

mechanism for the configuration cells inside the state-of-the-art FPGAs [28]–[30], we choose a group to be equal to the size of the error detection block, which is equal to a frame size. For instance, a frame size is 1312 bit in Virtex-4 FPGAs [28]. This enables us to switch to the transistors once a permanent error is detected on such a block. This is closely tied with the second design consideration, namely, the error detection mechanism. Since there is an embedded error detection mechanism inside most FPGAs for configuration cells, no additional circuitry will be needed. However, the controller should be able to detect whether an error is a soft error or a permanent error. We address this issue in the controller design.

The routing circuitry using 4T NEM relays makes the controller design much more intricate, since it requires two control input signals. However, it allows the device to have one or several of its frames be reconfigured many times without wearing out the NEM relays. This can be useful in dynamic reconfiguration. If a great degree of reconfiguration takes place in a relatively short time, the frame that receives such a heavy load can switch the routing path to go through its transistors rather than the NEM relays. This will allow the frame to be reconfigured several times without being worn out at the cost of increased static power consumption. Once the heavy reconfiguration phase has passed out, the controller should restore its state back and route the ground signal via the NEM relays. This scheme is not possible in a 6T NEM relay controller design, because it will loose the frame content when switching back to the NEM routing, since at one point, both the NEM and the nMOS transistors will be floating.

Both of the required controllers can be designed as simple FSMs. The FSM for controller used along with the 6T NEM relay routing circuitry is shown in Fig. 7. The only input of the FSM comes from the error detection mechanism in each frame. The output of the FSM determines the state of the 6T NEM relay in the routing circuitry. An output of zero means that the NEM relays are routing the ground signal to the virtual ground, whereas an output of one means that the NEM relay is being routed by the transistors. The controller begins at the valid state in which the output is zero. Once an error has been detected, the controller moves to the suspected status. Usually, once an error is detected in FPGAs, it gets fixed by either scrubbing [28], [31] or error correction [28], [32] to prevent soft errors from affecting the configuration bits. If the reconfiguration in scrubbing fixes the error signal, the FSM goes back to the valid state and determines the error to be of a soft error type. However, if scrubbing or error correction is unable to restore FPGA to a fault-free state, the FSM goes to the worn-out state, issuing an output of one.

The case is more complicated in the controllers designed with 4T NEM relays (shown in Fig. 8). The inputs to the circuit come from the error detection mechanism once an error is detected and they come from the reconfiguration platform once a phase with a heavy load of reconfiguration is about to take place. There are two outputs for the circuit, each determining whether the controlled 4T NEM is short or open. The controller begins at the valid state in which the output is zero for the transistor virtual ground and one for the NEM virtual ground. Similar to the other controller, once an error has been detected, the controller moves to the suspected state and if the error is persistent it goes to the worn-out state. Once a heavy load of reconfiguration is about to take place, the controller transitions to another state, which sets both the paths to one. This allows the current value of the frame to be preserved. This is more valuable in caches where updates to such groups are partial. Afterward, the NEM virtual ground becomes floating, so that the heavy reconfiguration can take place without wearing out the NEM relays. Once the heavy load of reconfiguration has passed away, this process is reversed.

E. Proposed SRAM in On-Chip Cache

As demonstrated in Section III-D, the first design consideration, i.e., the selection of granularity of grouping of the designed cells with the common virtual ground nodes, can be closely tied to the second design consideration if the target circuitry includes some error detection mechanisms. Similar to FPGAs, on-chip caches also include built-in error detection and correction circuitry [33]. This potentially eliminates the need for adding additional error detection circuitry. The error detection is applied to each word (64 bit) in on-chip caches [33]. This sets the limit for the minimum size of a group. However, since this size is too small and controller power consumption might be dominant, we group every ten words together as a single group. Hence, this problem reduces to the design of a controller capable of distinguishing between a soft error and a permanent error.

As discussed before, the routing circuitry using 4T NEM relays makes the controller design more complex, yet it allows more flexibility during the write operation. This controller is less useful in the on-chip caches. Since heavy write operations are not known in advance in on-chip caches, using the routing circuitry with 6T NEM relays is preferable in on-chip caches. However, in very limited cases where the processor is being used in an embedded system running a fixed application, the 4T NEM relay can be useful. However, it requires the target application to be profiled for write-intensive phases. Once the application gets to those write intensive phases, the controller should change the routing to pass through the transistors instead of the NEM relays, so that the NEM relays do not

TABLE I
ELECTRICAL PARAMETERS OF THE SIMULATED NEM RELAY MODEL

Parameter	Value	Parameter	Value
V_{dd}	1 v	R_{on}	1 $K\Omega$
V_{pi}	0.8 v	V_{po}	0.2 v
C_{gd}	20 aF	C_{gs}	20 aF
W	260 nm	L	65 nm

change their state too frequently in a small period of time. This can help further enhance NEM relay endurance. Since error detection also exists in the on-chip caches, the controller design is identical to that explained in FPGAs with the same signals.

F. Read/Write Latency

As for the read latency of the proposed SRAM cells, it is not affected neither in FPGAs nor in on-chip caches. FPGA SRAM cells provide a constant read (node Q in Fig. 4 is directly connected to a logic gate), and hence, the notion of read latency does not apply here. The read operation in the SRAM cells used in the on-chip caches involves precharging the bitlines, connecting the desired memory cell to the bitlines, and sensing the change in the bitline voltage. Using an NEM relay does not affect its ability to drive a bitline, i.e., NEM relays are only slower in switching their internal state and even offer better drive current due to their low ON-state (i.e., not switching) resistance ($\sim$1 kΩ compared with nMOSs $\sim$20 kΩ). Thus, it has no negative impact on the read latency of on-chip caches. The write latency is not negatively affected in SRAM cells, as explained in [18]. This is due to the fact that the typical write duration in an SRAM cell is long enough to set voltage node values, such that the relay switches with existing charge. In addition, the change in a static noise margin of an SRAM cell after using NEM relays is also reported in the same work and is not a contribution of this paper; SRAM hold and read static noise margins are improved by 110% and 250%, respectively, when using NEM relays [18].

IV. EXPERIMENTAL RESULTS

In this section, we first detail the experimental setup used to carry out the simulations. Afterward, we report the simulation results used for an FPGA configuration structure similar to Virtex 4, including the additional controlling circuitry. Finally, the simulation results for a last level on-chip cache of 4 MB are reported.

A. Experimental Setup

We used HSPICE to obtain the characteristics of the proposed SRAM cell. We also used the predictive technology model for the CMOS transistor model [34]. To model an NEM relay in HSPICE, we developed a model with the characteristics shown in Table I [18]. We purposefully use laterally actuated NEM relays due to their small footprint and their advantage to implement six-terminal relays [1]. As demonstrated in [18], stacking two lateral NEMs in a small 300 F^2 SRAM cell adds another metal layer into the layout (NEM relays fit in layer 2 and shifts up the other layers by one) but does not incur any area overhead. Notice that the reported area for a vertically driven NEM relay is 12 μm^2 [3], which equals to 27 SRAM cell. The designed controllers are synthesized with NanGate 45-nm open cell library using Synopsys Design Compiler, and the output netlist is converted into HSPICE to include the NEM relays. RapidSmith was used to obtain the statistics regarding the partial reconfiguration of FPGAs [35].

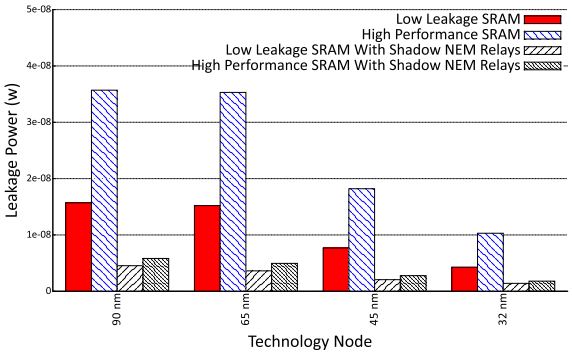


Fig. 9. Leakage power of high-performance and low-leakage SRAM cell designs with and without shadow NEM relays at various technology nodes.

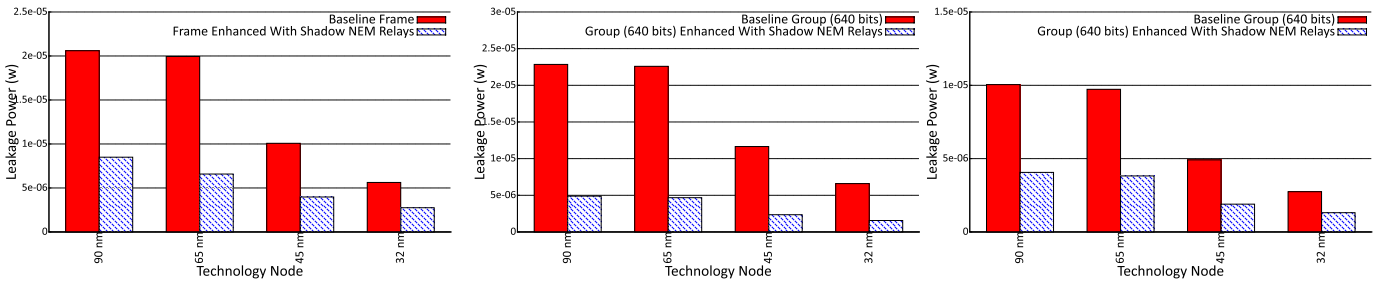
B. Simulation Results

For an individual SRAM cell, the power consumption for various technology nodes is shown in Fig. 9. These SRAM cells include the low-leakage SRAM cell for FPGAs and high-performance SRAM cell for on-chip caches. The low-leakage and high-performance SRAM designs are sized according to [36]. Fig. 9 shows the interesting fact that the leakage power reduction is much greater in the high-performance SRAM cells than the low-leakage SRAM cells. The leakage power of the low-leakage SRAM cells at a 45-nm technology is reduced by 73% by applying the shadow NEM relays, whereas the leakage power of the high-performance SRAM cells is reduced by 85%. This is due to the fact that the transistors of high-performance cell are wider and leak larger current, so the impact of adding an NEM relay with high OFF-resistance is greater in these cells. Other technology nodes also follow the same trend.

C. FPGA

In FPGAs, one controller is required for every frame to maximize the power efficiency. At the 45-nm technology node, the design compiler reported the area of the ground routing FSM circuitry designed with two 4T NEM relay is 29.792 μm^2. This area overhead is considered as 5% for a frame of 1312 low-leakage SRAM cells, which has an area of 577.936 μm^2 (1312 μm $\times$ 0.440 μm [36]). Such low area overhead is due to NEM relays being stacked on top of a CMOS layer and, hence, not contributing to the total area.

The power consumption of the whole frame along with the additional circuitry is shown in Fig. 10(a). As shown



(a) (b) (c)

Fig. 10. Power consumption of a frame of FPGA configuration bits and a group of 640 cache bits using different SRAM cells with routing circuitries at various technology nodes. (a) FPGA frame using low-leakage SRAM cells and two 4T NEM routing circuitry. (b) 640 cache bits using high-performance SRAM cells and a 6T NEM routing circuitry. (c) 640 cache bits using low-leakage SRAM cells and a 6T NEM routing circuitry.

in Fig. 10(a), the power consumption of a frame is reduced to 41%, 33%, 39%, and 49% for 90-, 65-, 45-, and 32-nm technology nodes, respectively. Upon the detection of a failed frame, the power consumption for that frame will increase to its normal amount along with the overhead imposed by the added circuitry. At the 45-nm technology node, any amount of failure more than 83% of frames results in power overhead. Any value below 83% results in power savings proportional to the frames with the working NEM relays. Upon the failure of all frames, 13% power overhead is added to the device as a result of the additional circuitry.

D. On-Chip Cache

As discussed before, one controller is required for every group to maximize the power efficiency. We choose every group to consist of ten words (640 bit). At the 45-nm technology node, the area of the ground routing circuitry designed with a 6T NEM relay is 14.896 μm^2. Such area overhead is considered as 4% area overhead for a group of high-performance SRAM cells (640 μm $\times$ 0.566 μm [36]). The power consumption of the whole group along with the additional circuitry is shown in Fig. 10(b). As shown in Fig. 10(b), the power consumption of a group of ten words is reduced to 21%, 21%, 20%, and 24% for 90-, 65-, 45-, and 32-nm technology nodes, respectively. The power saving is more significant than for FPGAs for two reasons. First, the 6T NEM routing controller was used, which requires a more complex control circuitry. Second, high-performance SRAM cells are used in on-chip caches, which consume more power than the low-leakage cells in the FPGAs. If the processor uses the low-leakage SRAM cells, the results will be different, as shown in Fig. 10(c). In this case, the power consumption of a group of ten words is reduced to 40%, 39%, 38%, and 48% for 90-, 65-, 45-, and 32-nm technology nodes, respectively. Upon the detection of a failed group, the power consumption for that group will increase to its normal amount along with the overhead imposed by the added circuitry. If a smaller number of words had been grouped into a group, more power could be saved once a single NEM relay fails. However, it would increase the number of controllers and, hence, the power overhead of the controllers. At the 45-nm technology node, any amount of failure more than 94% of groups results in power overhead when compared with a baseline. Any value below 94% results in power savings proportional to the group with the working NEM relays. Upon the failure of all groups, 5% power overhead is added to the device as a result of the additional circuitry.

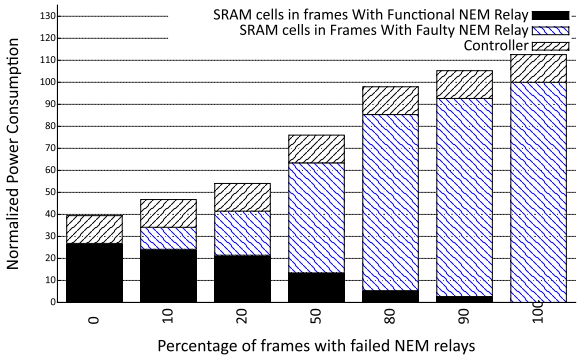


Fig. 11. Breakdown of power consumption for FPGA frames with a particular percentage of frames working with failed NEM relays.

It is worth mentioning that Chong *et al.* [18] have shown that the SRAM cells using the NEM relays in their internal structure will not switch slower than the typical SRAM cells. This is due to the fact that the typical write duration in the SRAM cell is long enough to set voltage node values, such that the relay switches with existing charge. This fact may not be applicable to other digital blocks, which are enhanced with the shadow NEM relays and requires a case-by-case study in every digital block.

E. Power Consumption Upon Failure of NEM Relays

The breakdown of power consumption for both FPGAs and on-chip caches for various failed fractions of the NEM relays in the 45-nm technology node is shown in Figs. 11 and 12, respectively. As can be seen, the power consumption growth with failing NEM relays is much more significant in on-chip caches, since the SRAM cells used are high performance and, hence, consume more power. It can also be seen that the overhead of the simple 6T routing controller used in on-chip caches is much less than that of two 4T routing controller used in FPGAs.

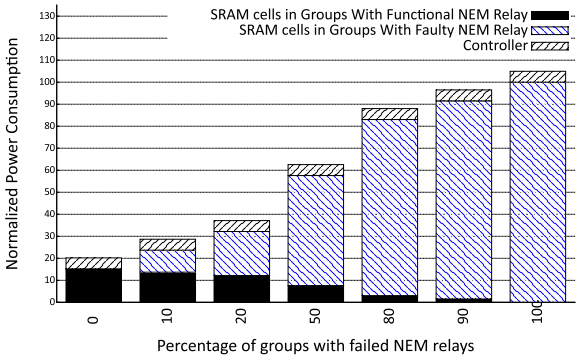


Fig. 12. Breakdown of power consumption for on-chip caches with a particular percentage of groups working with failed NEM relays.

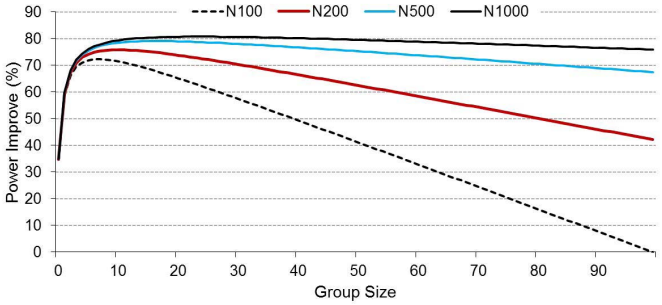


Fig. 13. Power improvement of a cache word with respect to different group sizes and hotspot probabilities.

To further elaborate the effectiveness of the proposed shadow NEM relays scheme upon the failure of several relays, we have investigated two partial reconfiguration design samples from Xilinx. The first design is named Color2 and is detailed in Xilinx UG702 [37]. Once this design is mapped into the device, 944 of frames are configured with data. Depending on the reconfiguration, some of these frames are changed, and at worst 220. Hence, after significant number of reconfigurations, eventually 220 of 944 frames will be worn out. At this point, the proposed scheme would provide 44% power saving with the low-leakage SRAM cells. The second design is a processor described in Xilinx UG744 [38]. In this design, 161 of 3967 configured frames are reconfigured during the partial reconfiguration. The NEM relays used in these 161 frames will eventually fail resulting in 57% power savings with the low-power SRAM cells.

As the worst case, if we assume that the whole FPGA is being reconfigured once per minute, still it will take about four years for the NEM relays having 2×10^6 reliable switching to wear out. It is worth to remind that in the case of heavy reconfigurations, i.e., once per minute, our controller circuitry temporarily switches to the normal nMOS mode and prevents the relay from being worn out.

F. Group Size and Error Detection

As for the cache example, it may inferred that choosing groups with higher number of words compensates the controller power overhead more efficiently. However, note that the virtual ground node in such case is shared between all SRAM cells within a group, and in the case of a failing a single NEM relay, all of the cells should switch to the nMOS state, which consumes higher power. For example, if a group size with ten words is used and one of the words becomes hotspot, it will wear out quickly and cause the other NEM-based words become useless, which means only one-tenth of NEMs' life cycle could be utilized, by average. The power gain of a single cache word within a group of G_S words can be modeled as follows:

$$\text{Gain}_{\text{word}} = 1 - \frac{P_{\text{NEM}} + \frac{P_{\text{CTRL}}}{G_S}}{P_{\text{MOS}}} \tag{1}$$

in which P_{NEM}, P_{MOS}, and P_{CTRL} are the power of cache word using NEM relays, cache word with nMOS transistors, and the controller, respectively. As mentioned before, different words within a group do no wear out uniformly. Assuming that there is one hotspot word within every N words, we can rewrite (1) as follows:

$$\text{Gain}_{\text{word}} = \left(1 - \frac{P_{\text{NEM}} + \frac{P_{\text{CTRL}}}{G_S}}{P_{\text{MOS}}}\right) \times \left(1 - \frac{G_S}{N}\right). \tag{2}$$

This formula can be simply extended for any other design (e.g., frames of FPGA) sharing the controller within a group. Fig. 13 shows the power saving of a single cache word with respect to different N and group size values. Maximum power saving is achieved for large N, i.e., when the probability of hotspot is low. For the smaller values of N, choosing large group sizes deteriorates the power saving due to early worn out of hotspot relays.

As mentioned in Section III, when the error detection circuitry discovers an error within a bunch of bits, the configuration circuitry rewrites the corresponding content, and then, the error detector checks again to decide whether it was a transient or permanent error. Therefore, a simple error detection algorithm as parity-checker (e.g., XOR) can be used for designs lacking an internal error detection system. In the case of choosing groups with a larger size than the granularity of the detector, the output of multiple error detectors can be ORed and fed into the shared controller.

V. Conclusion

Emerging memory technologies, such as NEM relays and nonvolatile memories, are being explored in academia to help resolve the issues faced in industry. As for our proposed method, it helps with the employment of NEM relays in FPGAs that require frequent dynamic partial reconfiguration. In the partial reconfiguration, a small part of the design is frequently changed in each configuration, while the rest of the design remains intact. Hence, after a significant number of reconfigurations, the NEM relays in that part get stuck and the NEM-based FPGAs become useless. Using our proposed method, however, NEM relays can be used in FPGAs that are subject to the partial reconfiguration in conjunction with parallel nMOS transistors as the spares. In the proposed method, once NEM relays become useless, the power consumption will

return to its nominal value, i.e., when no NEM relays are used in the FPGA device. In order to demonstrate the efficiency of the proposed method, we have shown that in the Xilinx partial reconfiguration designs, only a small fraction of the FPGA is frequently reconfigured without regard to the number of reconfigurations, providing the opportunity for the majority of NEM relays to continue functioning, which consequently provides a considerable power savings for such applications. As for on-chip caches, our proposed method can be used in the embedded processors that keep executing the same application. Hence, the designer can make sure that only parts of the cache are frequently updated, and hence, even after a very long period of time, one part of the cache consumes nominal power, while the rest of the cache is working in the power-efficient mode.

Our next step toward utilizing the shadow NEM relays is applying it to a complete standard cell library at the layout level. This standard cell library can then be used to implement any digital circuit. This will also require adding other detection schemes to such digital circuits, since there exist various circuits that do not have embedded error detection or correction mechanisms.

REFERENCES

[1] D. Lee *et al.*, "Combinational logic design using six-terminal NEM relays," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 32, no. 5, pp. 653–666, May 2013.

[2] R. Nathanael, V. Pott, H. Kam, J. Jeon, and T.-J. K. Liu, "4-terminal relay technology for complementary logic," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2009, pp. 1–4.

[3] M. Spencer *et al.*, "Demonstration of integrated micro-electro-mechanical relay circuits for VLSI applications," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 308–320, Jan. 2011.

[4] D. A. Czaplewski *et al.*, "A nanomechanical switch for integration with CMOS logic," *J. Micromech. Microeng.*, vol. 19, no. 8, p. 085003, Jul. 2009.

[5] K. Akarvardar *et al.*, "Design considerations for complementary nanoelectromechanical logic gates," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2007, pp. 299–302.

[6] F. Chen, H. Kam, D. Marković, T.-J. K. Liu, V. Stojanović, and E. Alon, "Integrated circuit design with NEM relays," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2008, pp. 750–757.

[7] R. Venkatasubramanian, S. K. Manohar, and P. T. Balsara, "NEM relay-based sequential logic circuits for low-power design," *IEEE Trans. Nanotechnol.*, vol. 12, no. 3, pp. 386–398, May 2013.

[8] C. Chen *et al.*, "Efficient FPGAs using nanoelectromechanical relays," in *Proc. 18th Annu. ACM/SIGDA Int. Symp. Field Program. Gate Arrays (FPGA)*, New York, NY, USA, 2010, pp. 273–282. [Online]. Available: http://doi.acm.org/10.1145/1723112.1723158

[9] H. Kam, V. Pott, R. Nathanael, J. Jeon, E. Alon, and T.-J. K. Liu, "Design and reliability of a micro-relay technology for zero-standby-power digital logic applications," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2009, pp. 1–4.

[10] N. Sinha, Z. Guo, A. Tazzoli, A. DeHon, and G. Piazza, "1 volt digital logic circuits realized by stress-resilient ALN parallel dual-beam MEMS relays," in *Proc. IEEE 25th Int. Conf. Micro Electro Mech. Syst. (MEMS)*, Jan./Feb. 2012, pp. 668–671.

[11] C. Chen *et al.*, "Nano-electro-mechanical relays for FPGA routing: Experimental demonstration and a design technique," in *Proc. Design, Autom., Test Eur. Conf. Exhibit. (DATE)*, Mar. 2012, pp. 1361–1366.

[12] S. Han, V. Sirigiri, D. G. Saab, and M. Tabib-Azar, "Ultra-low power NEMS FPGA," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2012, pp. 533–538.

[13] C. Dong, C. Chen, S. Mitra, and D. Chen, "Architecture and performance evaluation of 3D CMOS-NEM FPGA," in *Proc. 13th Int. Workshop Syst. Level Interconnect Predict. (SLIP)*, Jun. 2011, pp. 1–8.

[14] Y. J. Kim and W. Y. Choi, "Nonvolatile nanoelectromechanical memory switches for low-power and high-speed field-programmable gate arrays," *IEEE Trans. Electron Devices*, vol. 62, no. 2, pp. 673–679, Feb. 2015.

[15] P. Sedcole, B. Blodget, T. Becker, J. Anderson, and P. Lysaght, "Modular dynamic reconfiguration in Virtex FPGAs," *IEE Proc.-Comput. Digit. Techn.*, vol. 153, no. 3, pp. 157–164, May 2006.

[16] H. F. Dadgour and K. Banerjee, "Hybrid NEMS-CMOS integrated circuits: A novel strategy for energy-efficient designs," *IET Comput. Digit. Techn.*, vol. 3, no. 6, pp. 593–608, Nov. 2009.

[17] R. Venkatasubramanian, S. K. Manohar, V. Paduvalli, and P. T. Balsara, "NEM relay based memory architectures for low power design," in *Proc. 12th IEEE Conf. Nanotechnol. (IEEE-NANO)*, Aug. 2012, pp. 1–5.

[18] S. Chong *et al.*, "Nanoelectromechanical (NEM) relays integrated with CMOS SRAM for improved stability and low leakage," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2009, pp. 478–484.

[19] S. K. Manohar, R. Venkatasubramanian, and P. T. Balsara, "Hybrid NEMS-CMOS DC-DC converter for improved area and power efficiency," in *Proc. 25th Int. Conf. VLSI Design (VLSID)*, Jan. 2012, pp. 221–226.

[20] S. Chong *et al.*, "Integration of nanoelectromechanical (NEM) relays with silicon CMOS with functional CMOS-NEM circuit," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2011, pp. 30.5.1–30.5.4.

[21] J. O. Lee *et al.*, "A sub-1-volt nanoelectromechanical switching device," *Nature Nanotechnol.*, vol. 8, no. 1, pp. 36–40, Nov. 2012.

[22] W. W. Jang *et al.*, "Fabrication and characterization of a nanoelectro-mechanical switch with 15-nm-thick suspension air gap," *Appl. Phys. Lett.*, vol. 92, no. 10, pp. 103110-1–103110-3, Mar. 2008.

[23] A. Bazigos *et al.*, "Analytical compact model in Verilog-A for electrostatically actuated ohmic switches," *IEEE Trans. Electron Devices*, vol. 61, no. 6, pp. 2186–2194, Jun. 2014.

[24] S. Rana *et al.*, "Modelling NEM relays for digital circuit applications," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 805–808.

[25] V. Ranganathan, S. Rajgopal, M. Mehregany, and S. Bhunia, "Analysis of practical scaling limits in nanoelectromechanical switches," in *Proc. 9th IEEE Int. Conf. Nano/Micro Eng. Molecular Syst. (NEMS)*, Apr. 2014, pp. 471–476.

[26] S. H. Roh, K. Kim, and W. Y. Choi, "Scaling trend of nanoelectro-mechanical (NEM) nonvolatile memory cells based on finite element analysis (FEA)," *IEEE Trans. Nanotechnol.*, vol. 10, no. 3, pp. 647–651, May 2011.

[27] S. Yazdanshenas and H. Asadi, "Fine-grained architecture in dark silicon era for SRAM-based reconfigurable devices," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 61, no. 10, pp. 798–802, Oct. 2014.

[28] C. Carmichael and C. W. Tseng, "Correcting single-event upsets in Virtex-4 FPGA configuration memory," Xilinx Corporation, San Jose, CA, USA, Application Note no XAPP1088, Oct. 2009.

[29] K. Chapman and L. Jones, "SEU strategies for Virtex-5 devices," Xilinx Corporation, San Jose, CA, USA, Application Note no XAPP864, Apr. 2010.

[30] "Error detection and recovery using CRC in Altera FPGA devices," Altera Corporation, San Jose, CA, USA, Application Note no 357, Jul. 2008.

[31] I. Herrera-Alzu and M. Lopez-Vallejo, "Design techniques for Xilinx Virtex FPGA configuration memory scrubbers," *IEEE Trans. Nucl. Sci.*, vol. 60, no. 1, pp. 376–385, Feb. 2013.

[32] B. F. Dutton and C. E. Stroud, "Single event upset detection and correction in Virtex-4 and Virtex-5 FPGAs," in *Proc. CATA*, 2009, pp. 57–62.

[33] J. Kim, N. Hardavellas, K. Mai, B. Falsafi, and J. Hoe, "Multi-bit error tolerant caches using two-dimensional error coding," in *Proc. 40th Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO)*, Dec. 2007, pp. 197–209.

[34] W. Zhao and Y. Cao, "Predictive technology model for nano-CMOS design exploration," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 3, no. 1, 2007, Art. no. 1.

[35] C. Lavin, M. Padilla, J. Lamprecht, P. Lundrigan, B. Nelson, and B. Hutchings, "RapidSmith: Do-it-yourself CAD tools for Xilinx FPGAs," in *Proc. Int. Conf. Field Program. Logic Appl. (FPL)*, Sep. 2011, pp. 349–355.

[36] V. Gupta and M. Anis, "Statistical design of the 6T SRAM bit cell," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 1, pp. 93–104, Jan. 2010.

[37] *Xilinx Partial Reconfiguration User Guide*, Xilinx Corporation, San Jose, CA, USA, User Guide UG702, Oct. 2010.

[38] *Xilinx Partial Reconfiguration of a Processor Peripheral Tutorial*, Xilinx Corporation, San Jose, CA, USA, User Guide UG744, Apr. 2013.

Sadegh Yazdanshenas received the B.Sc. degree from the Iran University of Science and Technology, Tehran, Iran, in 2012, and the M.Sc. degree from the Sharif University of Technology (SUT), Tehran, in 2014.

He has been with the Data Storage Systems and Networks Laboratory, Department of Computer Engineering, SUT, as a Research Assistant for two years. His current research interests include reconfigurable computing, fault-tolerant design, and emerging nonvolatile memory technologies.

Behnam Khaleghi received the B.Sc. degree in computer engineering from the Sharif University of Technology (SUT), Tehran, Iran, in 2013, where he is currently pursuing the master's degree.

He was a Research Assistant with the Chair for Embedded Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2014 and 2015. He is also a Research Assistant with the Data Storage Systems and Networks Laboratory, SUT. His current research interests include reconfigurable architectures and computer-aided design.

Paolo Ienne was an Undergraduate Researcher with Brunel University, Uxbridge, U.K., from 1990 to 1991. From 1992 to 1996, he was a Research Assistant with the Microcomputing Laboratory and the MANTRA Center for Neuro-Mimetic Systems, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. In 1996, he joined the Semiconductors Group, Infineon Technologies AG, Munich, Germany. After working on datapath generation tools, he became the Head of the Embedded Memory Unit with the Design Libraries Division, Munich, Germany. He has been a Professor with EPFL since 2000, where he is currently the Head of the Processor Architecture Laboratory. His current research interests include various aspects of computer and processor architecture, electronic design automation, computer arithmetic, field-programmable gate arrays (FPGAs) and reconfigurable computing, and multiprocessor systems-on-chip.

Dr. Ienne is a regular member of program committees of international workshops and conferences in the areas of design automation, computer architecture, embedded systems, compilers, FPGAs, and asynchronous design. He was a recipient of the best paper award at the 40th Design Automation Conference (DAC) in 2003, the International Conference on Compilers, Architectures, and Synthesis for Embedded Systems in 2007, the 19th International Conference on Field-Programmable Logic and Applications in 2009, and the 20th ACM/SIGDA International Symposium on FPGAs in 2012. In 2008, he was the General Co-Chair of the Sixth IEEE Symposium on Application Specific Processors and a Guest Editor of a Special Section on Application Specific Processors of the IEEE TRANSACTIONS ON VLSI SYSTEMS. In 2010, he was the Program Subcommittee Chair of the DAC on High-Level and Logic Synthesis. From 2010 to 2012, he was a Topic Co-Chair of Design Automation and Test in Europe for Architectural and High-Level Synthesis. In 2011, he was a Program Co-Chair of the 20th IEEE Symposium on Computer Arithmetic and a Program Co-Chair of the 22nd IEEE International Conference on Application-Specific Systems, Architectures and Processors. In 2012, he was a Guest Editor of the Special Section on Computer Arithmetic of the IEEE TRANSACTIONS ON COMPUTERS. In 2014, he was the Chair of the Program Committee of the 23rd International Workshop on Logic and Synthesis and a Guest Editor of the IEEE Micro Special Issue on Reconfigurable Computing. In 2015, he was a Guest Editor of a Special Section on Emerging Memory Technologies in Very Large Scale Computing and Storage Systems of the IEEE TRANSACTIONS ON COMPUTERS and the Special Issue on Methods and Models for System Design of the *ACM Transactions on Embedded Computing*. He has been an Associate Editor of *ACM Transactions on Design Automation of Electronic Systems* since 2011, and of *ACM Computing Surveys* since 2014.

Hossein Asadi (M'08–SM'14) received the B.S. and M.S. degrees in computer engineering from the Sharif University of Technology (SUT), Tehran, Iran, in 2000 and 2002, respectively, and the Ph.D. degree in electrical and computer engineering from Northeastern University, Boston, MA, USA, in 2007.

He was a member of the Sharif Rescue Robots Group from 2001 to 2002. From 2002 to 2003, he was a member of the Dependable Systems Laboratory at SUT, where he was involved in research on hardware verification techniques. He was with EMC Corporation, Hopkinton, MA, USA, as a Research Scientist and Senior Hardware Engineer, from 2006 to 2009. He has been with the Department of Computer Engineering, SUT, since 2009, where he is currently a Tenured Associate Professor. He is the Founder and Director of the Data Storage Systems Laboratory at SUT. He spent three months in 2015 as a Visiting Professor with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. He has co-founded the first start-up company in the Middle East, called HPDS, Tehran, Iran, designing and fabricating midrange and high-end data storage systems. He has authored or co-authored over 60 technical papers in reputed journals and conference proceedings. His current research interests include data storage systems and networks, solid-state drives, operating system support for I/O and memory management, and reconfigurable and dependable computing.

Dr. Asadi was a recipient of the Technical Award for the Best Robot Design from the International RoboCup Rescue Competition, organized by The Association for the Advancement of Artificial Intelligence and RoboCup, the best paper award at the 15th CSI Internation Symposium on Computer Architecture and Digital Systems (CADS), and the Distinguished Lecturer Award from SUT in 2010, one of the most prestigious awards in the university. He was also a recipient of Extraordinary Ability in Science visa from the U.S. Citizenship and Immigration Services in 2008. He served as the Publication Chair of several national and international conferences, including CSI International Symposium on Computer Networks and Distributed Systems 2013, CSI International Symposium on Artificial Intelligence and Signal Processing 2013, and CSI International Symposium on Computer Science and Software Engineering 2013, for three years. Most recently, he served as a Guest Editor of the IEEE TRANSACTIONS ON COMPUTERS and a Program Co-Chair of the 18th International Symposium on CADS in 2015.