Folivora: Ultra Low Power Microprocessor Design with Nano Electro-mechanical Relay and Nanotube Memory

Siyuan Lu, *Student Member, IEEE*, Yuanqing Cheng, *Senior Member, IEEE*, Ying Wang, *Senior Member, IEEE*, *Member, IEEE*, Rui Wang, *Senior Member, IEEE*

***Abstract*—In post-Moore era, CMOS technology scaling has encountered enormous design and fabrication challenges. "Power Wall" limits the further increase of integration density. Emerging AI computing and data center deployments aggravate the power consumption problem further. In the pursuit of efficient computing paradigm, Nano Electro-mechanical (NEM) relay and Nanotube Random Access Memory (NRAM) technology have attracted enormous attention and have ultra-low power consumption compared to CMOS counterparts. NEM relay is a kind of device based on electronic and mechanical interaction switching, characterized by remarkably low power consumption. This article explores the application of NEM relay and NRAM technology to build a complex RISC processor, aiming to achieve much lower power without degrading performance. The controller and data path can be implemented with primitive logic gates made of NEM relays, and on-chip cache can be implemented with NRAM. Experimental results show that the energy efficiency of the processor design based on NEM relay and NRAM can be improved by 88.2% and 78.9% compared with CMOS technology based in-order and out-of-order microprocessors, respectively. Meanwhile, the performance can be improved by 42.9% and the instruction execution time can be reduced by more than 17.9%, which implies the potentials of NEM relay and NRAM for emerging ultra-low power applications.**

***Index Terms*—Low power design, NEM relay, Nanotube memory, Performance**

I. Introduction

As CMOS technology driven by Moore's law has entered sub-10 nanometer regime, leakage power is now becoming a paramount roadblock to increase the integration density further due to the shrinking thickness of gate oxide. In addition, in deep sub-micron technology node, interconnect delay dominates the logic gate delay and is becoming the bottleneck of performance scaling. The pipeline technique has been used widely in modern high performance processor to reduce the critical path delay by overlapping instruction executions. However, the pipeline registers inserted introduce extra power consumption, which aggravates the "Power Wall" problem further. Therefore, the ultra low power microprocessor design is essential especially for emerging edge computing applications.

This work is partly supported by the Shenzhen Science and Technology Program under Grant No.SGDX20230116093303006 and KJZD20231023100201003. Siyuan Lu and Yuanqing Cheng are with the School of Integrated Circuit and Engineering, Beihang University, Beijing 100191, China and Shenzhen Institute of Beihang University, Shenzhen 518000, China (email: sy2341203@buaa.edu.cn, yuanqing@buaa.edu.cn). Rui Wang is with the School of Computer Science and Engineering, Beihang University, Beijing, 100191, China (email: wangrui@buaa.edu.cn). Ying Wang is with Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China. (email: wangying2009@ict.ac.cn). The corresponding author is Rui Wang.

In order to deal with the above problems, some emerging semiconductor devices are proposed to replace CMOS devices approaching ultra low power consumption. Among these devices, Nano Electro Mechanical (NEM) relay is a promising candidate due to their distinguished merits, such as zero leakage and one-cycle-for-all operation which means we can perform the instruction execution by simultaneous switching of NEM relays only once [1]. Besides, the Nanotube Random Access Memory (NRAM), which has the similar working mechanism as NEM relay, is a good candidate to replace SRAM as on-chip memory for leakage power reduction [2].

In prior research, Spencer *et al.* demonstrated how to construct basic logic gates and adders using a four-terminal NEM relay in [3]. Tang *et al.* showed that NEM relays can effectively implement not only quasi delay-insensitive (QDI) designs, but also bundled-data and power-gated circuits [4]. Li *et al.* optimized a 32-bit relay adder to show better performance compared to CMOS [5]. At the same time, various NEM relay variants have also been proposed: Rosendale *et al.* demonstrated a 4Mb nonvolatile memory with a carbon nanotube (CNT) storage element, fabricated with a $0.25\,\mu m$ CMOS process [2]; pull-in-free electrostatic NEM relays enabling stable switching was demonstrated in [6]; and compact single-contact four-terminal relay structures were recently reported in [7], [8].

While prior work emphasizes the functional versatility of NEM relays, their physical robustness has also become increasingly relevant for harsh-environment computing. Relay-based devices exhibit ultra-low leakage even at elevated temperatures and are intrinsically tolerant to radiation-induced charge effects [6], [9].

Meanwhile, NRAM complements NEM relays by providing bit-addressable non-volatility with similarly robust environmental tolerance. Saito *et al.* demonstrated a 16 Mb 1T1R NRAM macro integrating CNT resistive elements within intermediate metal layers in 55 nm CMOS technology [10]. Recent reliability reports from the same platform showed excellent thermal stability and endurance across cycles [10], while industry data highlighted resilience against heat, vibration, shock, magnetism, and radiation [11].

All these works show the promising benefits of NEM relay and NRAM in ultra low power circuit design. Unfortunately, they did not mention the ultra low power on-chip memory design, which occupies a large portion of leakage power. Furthermore, existing related works did not implement a complete microprocessor design with the nano-relays thus lacking of thorough Performance, Power, and Area(PPA) analyses compared to CMOS counterparts.

In this article, we present the design and implementation of the first microprocessor composed entirely of NEM relays and NRAM. This work pioneers the integration of NEM relays for logic operations and NRAM for on-chip storage, demonstrating a novel approach to ultra-low-power computing.

To enhance energy efficiency, we optimize the fundamental logic gate and flip-flop structures constructed using NEM relays, significantly reducing power consumption compared to conventional CMOS designs. To address the inherently large mechanical switching delay of NEM relays, we propose a one-cycle-for-all mechanical switching architecture that eliminates pipelining and enables instruction execution with only a single relay switching latency.

Furthermore, we integrate nanotube switches, which are compatible with the NEM relay fabrication process, to construct on-chip memory, effectively minimizing leakage power. Through extensive architectural and circuit-level simulations, we demonstrate that the proposed NEM relay and NRAM-based microprocessor achieves substantial power savings while maintaining performance levels comparable to CMOS implementations.

Our contributions can be highlighted as follows,

- In order to enable NEM relay to build large logic circuits while maintaining a single NEM switching delay, novel ANR-OR and OR-AND logic gates and multiplexer using NEM relay are proposed.
- We design an improved NRAM storage array structure to enhance data access speed and energy efficiency. This architecture optimizes storage density, power management, and reliability, making it suitable for high-performance computing and embedded systems.
- Building on the improved NRAM storage array, CACTI [12] has been extended and modified to accurately evaluate and optimize the PPA of NRAM-based cache architectures. Validated against prototype measurements, the enhanced CACTI tool provides precise modeling and analysis capabilities for NRAM-based cache design.

Experiment results showed that the energy consumption of NEM relay based processor is only 11.8% of the CMOS-based counterpart, and the performance is even 42.9% higher compared to out of order CMOS-based counterpart. Beyond performance and energy benefits, we further evaluate the proposed architecture from a physical design perspective through 3D integration-aware area analysis. Additionally, a detailed comparison against emerging memory technologies demonstrates the superior energy-leakage tradeoffs and CMOS compatibility of the NEM–NRAM system.

The rest of this paper is organized as follows. In Section II, we introduce the preliminaries of NEM relay and nano-tube based device followed by the introduction to related work. In Section III, we firstly describe the new structure of logic gates and flip-flops made by NEM relays in details. Then, we present the on-chip cache design with nanotube memory technology. Section IV evaluates the benefits of the NEM relay based microprocessor extensively, in terms of power and performance. Finally, Section V concludes this work.

II. Preliminaries and Related Work

A. Introduction to Nano Electro-mechanical Relay

NEM (Nano electro-mechanical) relay is an integrated, capacitively-actuated mechanical switches. As shown in Fig. 1, in a four-terminal NEM relay device structure, the gate resembles a spring, which is folded, bent, and suspended above several metal electrodes referred as the body, drain, and source respectively.

In an NEM relay, some depressions or cavities are formed on the drain and source electrodes, aligned with the raised portion of the gate. When a voltage is applied between the gate and the body, electrostatic forces are generated on the gate, making it move and deform the folded bending portion closer to the body. The downward mechanical motion during this adsorption phase exhibits a significant delay of up to $34\mu s$, primarily attributed to the gate structure's inertia and viscoelastic damping effects. When the conductive channel on the gate makes contact with the depressions on the drain and source, the movement stops, and a conducting path is formed between the source and drain. As the voltage between the gate and body decreases to a release voltage, the spring force rapidly pulls the gate back to its original position without substantial delay, thereby breaking the contact between the drain and source electrodes and restoring the off state.

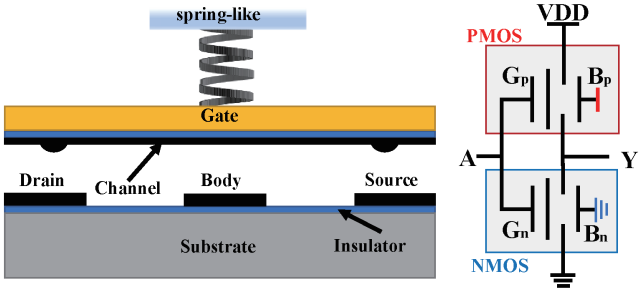

Fig. 1. Four terminal NEM relay structure and the schematic diagram of an NEM-based inverter.

The right part in Fig. 1 shows the structure of an inverter made of NEM relays. When the body is connected to VDD, the NEM relay forms a structure similar to a PMOS transistor. When the body is grounded, it forms a structure similar to a NMOS transistor. If a high voltage is applied to the gate, PMOS-like NEM relay is off and NMOS-like relay is on. So the gate outputs a low voltage. Otherwise, it produces a high voltage.

Due to this special mechanical structure, the NEM relay is non-volatile and has very low leakage current. Its threshold voltage is only $0.04V$ and the operating voltage is $1.2V$, which is suitable for ultra low power applications [13].

Fig. 2 shows the electrical model of an NEM relay that addresses power consumption and signal delay considerations. The model considers various resistances such as the channel resistance (R_{ch}), contact resistance (R_{con}), surface resistance due to chemicals or oxides (R_{surf}), and trace resistance (R_{trace}), where R_{con} and R_{surf} are more significant. Contact resistance is influenced by material properties, electron mean

free path, and effective contact area, dependent on electronic force and material hardness. Surface resistance encompasses effects from coatings and oxides, notably impacting large-scale relay operations. The model also includes intrinsic and extrinsic capacitances. Key capacitances between the gate and other terminals are modeled as parallel plates, with adjustments for the dynamic changes during relay operation. The combined electrical and mechanical model has been validated experimentally and implemented in Verilog-A for circuit design, capturing essential dynamics like switching delay and electrical delay.

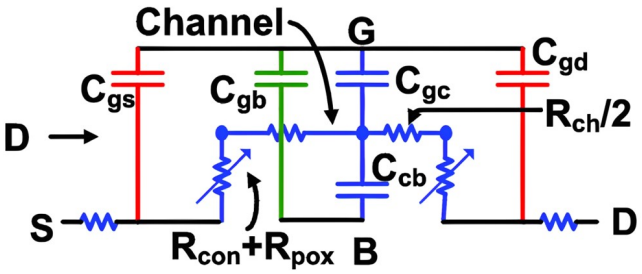

Fig. 2. The schematic of the NEM relay Verilog-A model.

NEM relays exhibit virtually zero off-state leakage current, resulting in negligible static power consumption, while their sharp on/off transitions support operation at very low supply voltages, reducing dynamic power usage. These features make them attractive for ultra-low-power digital logic [14], [15]. At the architectural level, techniques like shadow NEM relays can further amplify energy efficiency: applied to on-chip caches (*e.g.*, SRAM), they have achieved up to 80 % reduction in power consumption while mitigating NEM switching endurance limits [16].

B. Introduction to Nanotube Based Memory

Nanotube Random Access Memory, shorted as NRAM, works differently from the conventional SRAM. It is made from layers of carbon nanotubes that are grown from tiny particles of a catalyst, which is most commonly iron. Each NRAM cell has only one transistor consisting of a network of carbon nanotubes, and works in a similar way as other non-volatile memory technologies [17]. When the carbon nanotubes in a memory cell are close to each other, there is a repulsive force that can separate them apart. When they are far away, there is an attractive force that can contact each other by applying voltage, which allows NRAM to switch between low and high resistance states. Fig. 3 illustrates the working principle. When the nanotubes do not touch each other, the memory cell has a high resistance, representing the "off" or "0" state, and when nanotubes touch with each other, the memory cell has a low resistance, representing the "on" or "1" state.

NRAM is extremely durable, unlike SRAM and DRAM, which suffer from continuous leakage currents to retain data, NRAM consumes near-zero static power. This makes it particularly well-suited for low-power and energy-constrained applications, including edge computing, Internet of Things (IoT) devices, and battery-operated systems, approaching nearly unlimited read and write cycles [18].

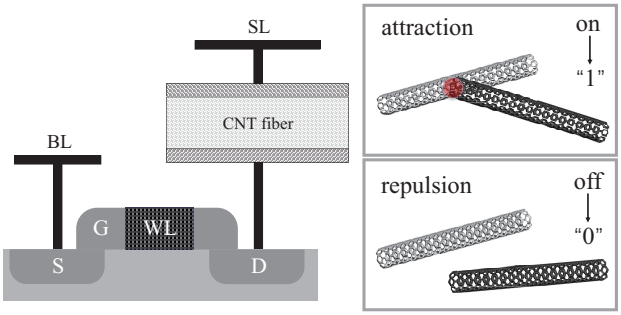

Fig. 3. The NRAM cell structure and on/off states.

They are also resistant to temperature variation, electromagnetic interference and radiation, which are threats of CMOS-based memory circuits. NRAM has very low read and write latency, and is comparable to that of DDR4 memory. At present, a few companies have built prototypes of NRAM and its advantages have been debuted in [19], [20]. Table I detailed comparisons of NRAM and CMOS memory [11].

TABLE I
MEMORY TIMING COMPARISONS

Operation Timing	DDR4 SDRAM(ns)	DDR4 NRAM(ns)	DDR5 SDRAM(ns)
Row cycle	47.00	46.25	50.18
Access time	17.14	13.50	18.18
Row to column	15.00	23.00	18.18
Precharge	15.00	14.25	18.18
Write recovery	15.00	23.00	30.00
Activate to precharge	32.00	32.00	32.00
Refresh	350.00	0.00	350.00

In addition, to facilitate an apples-to-apples comparison versus CMOS and to show the overlap with NEM relays, we add Table II summarizing common robustness benefits (temperature, EMI, radiation) and non-volatility across NEM relays and NRAM.

C. Related Work

NEM relay has proven to be a very energy-efficient alternative to CMOS transistors. Spencer *et al.* designed a 32-bit NEM relay-based adder, which was found to offer $10\times$ energy efficiency gain over an optimized CMOS adder with a moderate increase in area [3]. Tang *et al.* showed that a 64-bit C-element, 32-bit PCHB AND and 8-bit PCHB adder implemented with NEM relays can achieve over $16\times$, $25\times$ and $1.7\times$ better energy-efficiency respectively compared to the $90nm$ CMOS technology [4]. [30] used a combination of asynchronous and NEM techniques to implement bundled data and power gating circuits. Nair *et al.* proposed an NEM relay-based multi-bit RAM that can store multiple bits in a single cell. The NEM relay-based implementation can reduce power effectively, while the multi-bit storage technology helps achieve compact implementation with high storage

TABLE II
COMPARISON: NEM RELAY / NRAM / CMOS

Attribute	NEM Relay	NRAM (CNT)	CMOS
Temperature tolerance	up to $300\,^{\circ}C$ [21]	-55 to $+300\,^{\circ}C$ [22]	-40 to $+85/125\,^{\circ}C$ [23]
Radiation / SEU	radiation-hard; $\gg$ CMOS [21], [24]	tolerant [11]	SEU-susceptible [25]
magnetic susceptibility	low (no charge storage) [21]	low; magnetism resilient [11]	moderate–high [23]
Non-volatility	yes [6]	yes [11]	no [23]
Standby leakage	near zero [26]	near zero [27]	nonzero; rises with T [23]
Switching / write delay	tens–100 ns (mech.) [6]	≤ 5 ns [11]	few ns [23]
Integration / process	SOI foundry; BEOL-compatible [8]	BEOL-compatible; cross-point [27]	standard CMOS/BEOL [23]
Endurance (cycles)	$> 10^{10}$ [28]	$> 10^{12}$ [11]	10^4–10^6 [29]
Retention (years)	> 0.5 years measured off power [6]	> 300 years at $+300C$ [11]	loses data on power-down [29]

density [31]. Li *et al.* optimized the 32-bit NEM relay adder, which achieved $60\times$ less energy consumption per operation than its CMOS counterpart in $40nm$ technology [5]. Smith *et al.* proposed an efficient NEM relay logic gate designs, which were validated through a series of tests on 3-input NAND gate relay structures [32]. Lee *et al.* proposed to use a compact 2-to-1 multiplexer design based on a single six-terminal relay instead of Binary Decision Diagram (BDD) technique to implement arbitrary combinatorial logic functions [33].

In the field of NRAM research, a 4Mb non-volatile carbon nanotube memory manufactured in $0.25\mu m$ CMOS process was demonstrated by Rosendale [2], whose write endurance exceeded $10,000$ cycles showing strong data retention. Lamb *et al.* explored the key parameters, different device and memory array structures of CNT memory in $130nm$ process [34]. Saito *et al.* developed a 16Mb 1T1R NRAM that integrated CNT resistor elements into the intermediate metal level with $55nm$ CMOS process [10]. Veksler *et al.* investigated the memory refreshing characteristics of CNT-based memristors under circuit-relevant conditions, demonstrating their applicability for neuromorphic computing [17]. These advancements highlight the potential of NRAM as a high-performance, non-volatile memory solution.

A related demonstration involves a 3D CMOS–NEM FPGA architecture, where NEM relay logic is integrated atop the CMOS logic layer with face-to-face stacking. This approach achieves approximately 78% reduction in footprint, along with around 33% lower delay and 29% lower power compared to a 2D baseline [35]. In coarse-grained reconfigurable array (CGRA) designs, integration of multi-pole NEM relays in a 3D back-end-of-line stack improves area utilization by about 40%, yielding up to 19% reduction in area and 10% in power at iso-delay [36]. In hybrid 3D-stacked NEMFET–CMOS caches, despite a 55% larger footprint relative to 2D SRAM, the design delivers two orders of magnitude reduction in static energy and 38% average energy savings with minimal IPC loss [37].

While our implementation focuses on a single-core system, we investigated prior heterogeneous integration studies showing that NEM–CMOS systems scale well to multi-core or reconfigurable fabrics. For instance, a 3D CMOS–NEM FPGA leveraging face-to-face stacking of relay logic atop CMOS has been shown to nearly halve area and significantly reduce both dynamic and leakage power compared to CMOS-only baselines [38].

Early work shows that applying a thin SiO_2 coating to contact electrodes can extend endurance to over 400 switching cycles, while using ruthenium contacts with structural stress-matching further stabilizes ON-state resistance and enhances lifetime [28].

Although large-scale manufacturability of CNT-based NRAM is challenging, Nantero has demonstrated a highly engineered, 300 mm wafer spin-on CNT process fully compatible with BEOL CMOS fabs. Moreover, NRAM has been installed in multiple production fabs and these features, along with demonstrated 300 mm wafer-level integration, indicate that CNT-based NRAM is no longer limited to research labs, but is moving toward commercial usage in large-scale systems [22], [39].

The above related works show the ultra low power advantages of NEM relay and NRAM in logic and memory design respectively, but none of them combines the advantages of both to build a complete microprocessor. So an extensive PPA can not be evaluated from the system perspective. In this work, we design a MIPS-like microprocessor based on the NEM relay and NRAM technologies to achieve ultra-low power consumption. NEM–NRAM hybrid design separates compute and memory domains, allowing each technology to operate within its optimal endurance envelope.

III. MICROPROCESSOR DESIGN WITH NEM RELAYS AND NRAM

In this section, we provide a detailed description of the microprocessor architecture using NEM relays and NRAM. The overall design is illustrated in Fig. 4. This microprocessor architecture integrates two distinct technologies to optimize both performance and power consumption. Specifically, the control unit, register file and the arithmetic logic unit (ALU) are implemented using NEM relays, while the instruction memory and on-chip cache are based on NRAM technology.

From an architectural perspective, this design follows the classical five-stage pipeline model, which consists of Instruction Fetch (IF), Instruction Decode (ID), Execute (EX), Memory Access (MEM), and Write Back (WB). During the IF stage, the program counter (PC) controls instruction retrieval and selects the next instruction address via a multiplexer (MUX). Additionally, the instruction memory is implemented using NRAM, which provides higher storage density and lower power consumption.

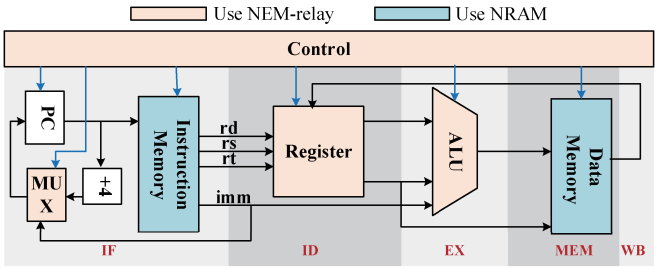

Fig. 4. Overview of the NEM relay based microprocessor architecture.

A. Controller Design with NEM Relays

The logic control unit of a RISC microprocessor is a critical component responsible for orchestrating and managing the operations of the entire processor, ensuring the accurate execution of instructions. As a purely combinational circuit, the control unit generates control signals based solely on the current instruction without relying on sequential elements or memory states.

To minimize the number of NEM relays used in the controller and achieve power and area efficiency, it is crucial to fully exploit the inherent strong pull-up and pull-down characteristics of NEM relays. This allows for the simplification of logic expressions governing control signals, thereby reducing circuit complexity, switching delays, and overall power consumption.

In previous research, NEM relays have primarily been utilized to implement fundamental logic gates such as AND, OR, and inverter. However, for complex microprocessor design, relying solely on primitive gates is insufficient due to area constraints and power efficiency requirements. To address these challenges, it is necessary to develop more powerful and intricate compound logic gates that can optimize circuit design by reducing relay count and minimizing power consumption.

The left part of Fig. 5(a) shows the structure of a compound OR-AND gate. The OR-AND gate is a cascading structure, with each stage comprised of relay-based multi-input OR gate structures. The output of each OR gate is connected to the next stage to implement the logic AND. The path from V_{dd} to first stage output can be established as long as A or B is turned on, realizing the OR gate function. The right part of Fig. 5 (a) shows the implementation of an AND-OR compound gate. It resembles the CMOS implementation with the same function. On the path to V_{dd}, multiple NEM relays are connected in parallel, while on the path to GND, multiple NEM relays are connected in series. Different from CMOS circuits which require an inverter to obtain a "NOT" signal, NEM relay can invert the input signal by simply changing the body voltage as shown in Fig. 1.

The use of NEM relays for constructing multi-input AND and OR gates provides several advantages over conventional CMOS-based designs. In traditional CMOS technology, when the fanout of a gate exceeds four, it is often necessary to restructure the circuit and introduce additional logic stages to mitigate excessive loading effects. This restructuring not only increases transistor count but also results in higher power consumption and longer signal propagation delays. In contrast, NEM relays have been demonstrated to possess strong driving capabilities, allowing them to handle significantly larger fanout values without requiring additional buffering [4]. Simulation results confirm that even when driving up to ten fanout gates, NEM relays exhibit negligible delay induced by load capacitance. The simulation waveform shown in Fig. 5 (b) further verifies the correctness and robustness of these compound gate structures, demonstrating their effectiveness in real-world microprocessor applications.

B. ALU Design with NEM Relays

The ALU is a fundamental component of a microprocessor, responsible for executing arithmetic and logical operations essential for instruction processing. In our microprocessor design, the ALU supports key operations, including addition, bitwise AND, and bitwise OR, ensuring compatibility with common computational tasks. To achieve efficient and compact logic implementation using NEM relays, the design primarily relies on an optimized adder and a high-speed multiplexer, in addition to elementary logic gates.

In order to complete instruction execution in a single NEM mechanical delay, we adopt the Manchester Carry Chain structure to implement the adder [3]. The C_{in} and $\overline{C_{in}}$ signals need to be propagated among multiple carry chains. As shown in Fig. 5(c), in our design, when A and B are inputs, all NEM relays have open/off transitions within one NEM switching delay. Note that since the carry signal path does not act on NEM devices again, it only introduces nanosecond of delay.

The multiplexer is another crucial component of the ALU, enabling the selection of different inputs for computation based on control signals. In our design, a dual-channel multiplexer is constructed using two NEM relays with opposite body voltages. When the selection signal is activated, one of the relays is guaranteed to be in an open state while the other remains in an off state, effectively sending the correct input to the output.

This multiplexer structure can be cascaded to create a larger multi-input selection block, allowing for more complex operand selection scenarios. Since all switching operations occur simultaneously at the gate terminals of the relays, the overall delay introduced by the multiplexer is also limited to one NEM mechanical delay. This delay characteristic is identical to that of the adder, ensuring that the entire ALU system operates with a uniform single NEM mechanical delay, thereby maintaining synchronization across all arithmetic and logical operations.

In our ALU design, each single-bit ALU unit consists of 12 NEM relays for the adder and 6 NEM relays for the 4-to-1 multiplexer. This results in a significant reduction in hardware compared to traditional CMOS designs, which typically require 24 and 12 transistors for adder and multiplexer respectively. Then, 32 single-bit ALU units can be used to build a 32-bit ALU unit.

C. Register Design with NEM Relays

In our microprocessor design, it includes 32 general-purpose registers, which are used for storing data and intermediate

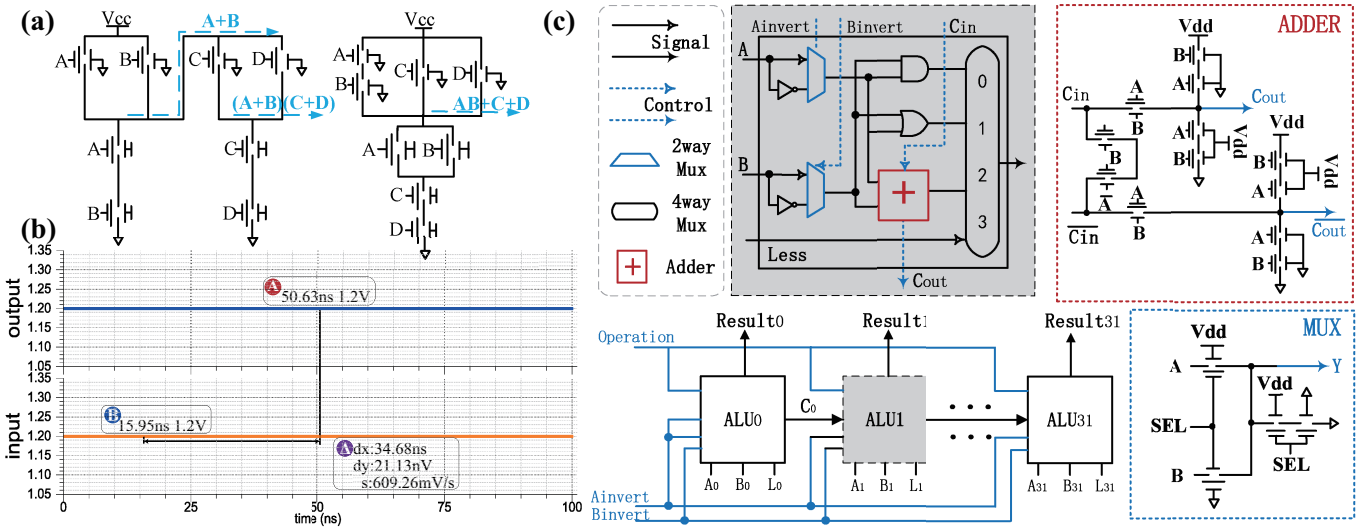

Fig. 5. (a) Two compound logic gates OR-AND and AND-OR. (b) The signal slew rate is negligible even the gate driving 10 fanout gates. (c) ALU ADDER and MUX structures built by NEM relays.

results. These registers play a crucial role in microprocessors. Registers are typically constructed using flip-flops, which may be D flip-flops or JK flip-flops.

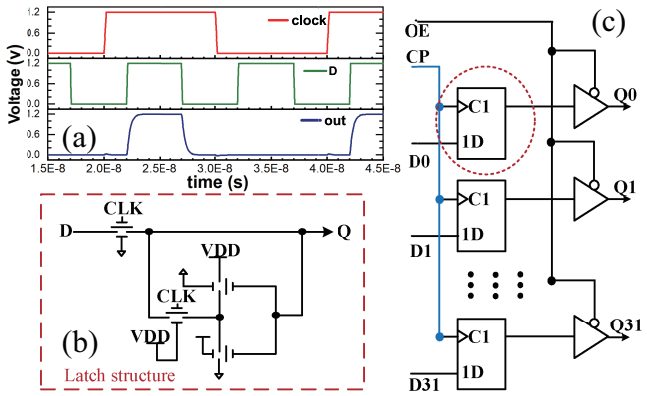

Fig. 6. (a) Simulation waveforms of NEM relay based D flip-flop. (b) The structure of NEM relay based D flip-flop. (c) The 32-bit register structure.

Flip-flops are fundamental sequential elements in microprocessor design, serving as essential building blocks for registers and storage elements. In traditional CMOS circuits, one of the most widely used structures of flip-flops is the transmission gate-based design. A conventional CMOS transmission gate typically requires a dual-MOS structure, comprising both NMOS and PMOS transistors, to ensure proper signal transmission and logic-level retention.

However, when utilizing NEM relays, the design constraints and operational principles differ significantly from those of MOS-based circuits. Due to the strong pull-up and pull-down characteristics inherent in NEM relays, as well as their lower on-resistance, the need for transmission gates is effectively eliminated. Instead, a single NEM relay is sufficient to achieve the functionality of a transmission gate. This simplification not only reduces the number of required devices but also minimizes power consumption and circuit complexity, making it highly advantageous for low-power applications.

We use D flip-flops to implement the register file in our design. Each transmission gate is replaced by a single NEM relay as mentioned above. Due to the strong pull-up and pull-down characteristics of NEM relays, the design does not require traditional transmission gates, reducing circuit complexity and improving efficiency.

To ensure that there is only one NEM mechanical delay in the timing path, the relay buffer is relocated to the feedback path of the latch, effectively hiding an NEM switching delay. This adjustment ensures that register file operations remain within a single NEM relay operation cycle. The structure is shown in Fig. 6, where the latch configuration and overall 32-bit register implementation are illustrated.

From the simulation waveforms in Fig. 6(a), we observe that the signal passing through the flip-flop exhibits minimal delay, confirming that the register file access completes within one NEM mechanical delay. The clock signal (red) drives the D input (green), and the output (blue) follows with a small propagation delay. The alignment of these signals verifies the correct timing behavior of the flip-flop.

The latch structure is depicted in Fig. 6(b), where a single NEM relay replaces the conventional transmission gate. The clock controlled switching mechanism ensures efficient data retention while the feedback path helps mitigate mechanical delays. The simplified design results in lower hardware overhead while maintaining precise logic operation.

The 32-bit register design is illustrated in Fig. 6(c), where multiple D flip-flops are arranged in parallel. Each flip-flop consists of a clocked latch (C1) and a data storage unit (1D). The clock pulse (CP) controls state transitions, while the output enable (OE) ensures proper data retrieval. This design efficiently handles data storage and retrieval while maintaining balanced timing constraints among instruction

execution stages.

The implementation eliminates redundant components and leverages NEM relay properties to optimize the register file design for low-power, high-speed applications.

D. NRAM-based On-chip Memory Modeling

The on-chip memory of the MIPS architecture comprises two primary components, *i.e.*, the instruction cache and data cache. Implementing 64MB cache consumes a large amount of NEM relays, which may introduce significant power and area overheads. So we choose NRAM as the substrate of on-chip cache, which has the similar structure as NEM introduced in Section II-B and is compatible with the NEM relay fabrication process.

Each memory cell has two states, representing "0" and "1". The commonly used NRAM cell structure is 1T1R as shown in Fig. 3. Compared with SRAM, NRAM has lower energy consumption, close to zero power consumption in standby mode, faster writing speed and unlimited scalability [18].

In order to evaluate the power/area/speed of NRAM, we revised the open source SRAM simulator CACTI [12], and added the support of NRAM devices. Although Miwa *et al.* developed a CNT-based cache simulator extended from CACTI, their work focused on CNFET-based memory cell structure similar to 6T CMOS SRAM [40]. In this work, however, our target technology is nanotube-based memory or NRAM, which is fundamentally different from CNFET-based SRAM. In the revised CACTI, the 6T-SRAM memory cell is changed to 1T1R NRAM cell. The voltage amplifier used in SRAM is changed to a current amplifier referring to [2]. The writing of "0" or "1" is determined by the switching current magnitude, and data are read out through a voltage sensing amplifier (SA). The SA structure is optimized for power consumption, and is shown in Fig. 9. Power consumptions of NRAM cell, SA, bitline/wordline interconnects and other peripherals were obtained by circuit simulation with parameters detailed in Section IV in $90nm$ technology node. The simulation results were calibrated and verified with publication data from [18], [41].

1) NRAM-based Memory Architecture: Fig. 7 illustrates how to convert conventional CACTI-modeled SRAM architecture into a 1T1R-based NRAM architecture. The NRAM storage array is composed of 1024 rows and 512 columns, resulting in a total of 524,288 memory cells. Each memory cell leverages the electrical conductivity variation of CNTs to store binary states, either "0" or "1". The storage cells are arranged in a cross-bar structure, where access is controlled through word lines (WL) and bit lines (BL).

At the top and bottom of the array, the *Y Selection and Precharge* (Y SEL & Precharge) circuits facilitate bit-line selection and precharging operations. Surrounding the memory array, multiple sense amplifiers and write drivers (SA&WD) are positioned to manage read and write operations. Specifically, SA&WD0 to SA&WD3 handle the lower section of the array, while SA&WD4 to SA&WD7 manage the upper section. These sense amplifiers detect the stored data, whereas the write drivers enable data writing by applying appropriate voltage levels to alter the conductive state of the CNT network.

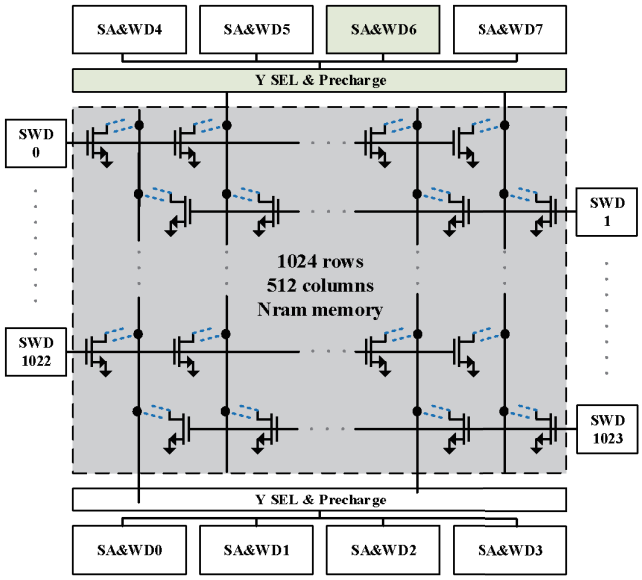

Fig. 7. NRAM array structure

Within the memory cells, each storage device consists of a nanotube network accessed via control transistors. The writing process applies an external voltage to reversibly alter the conductive state of the nanotube bundle, ensuring reliable data storage. The reading process involves detecting the electrical conductivity state of a memory cell and amplifying the output signal using sense amplifiers. This structure enables high-density storage while maintaining low power consumption and excellent durability.

2) NRAM memory cell structure modeling: Different from the traditional CACTI model for SRAM, which utilizes bitlines and wordlines optimized for rapid charge dynamics to support volatile memory functions, NRAM introduces significant structural modifications tailored to its non-volatile nature and unique resistive switching mechanism. Unlike SRAM that maintains data integrity through continuous power supply, NRAM leverages carbon nanotube technology where data storage is governed by the physical change in resistance states. To accommodate this, NRAM design halves the capacitance of its bitlines compared to that of SRAM, effectively reducing the energy consumption for memory state transitions by about 30%. This reduction in bitline capacitance is quantitatively represented as:

$$C_{\text{bitline, NRAM}} = \frac{C_{\text{bitline, SRAM}}}{2} \tag{1}$$

This critical enhancement in NRAM technology not only reduces the operational energy but also enables faster memory access, thereby enhancing the efficiency and performance of memory operations. These targeted adjustments in the bitline and wordline configurations are indicative of NRAM's engineered approach to meet the specific demands of high-density, low-power memory applications.

Furthermore, the carbon-nanotube based memory cell introduces new operational constraints on its control transistor,

which demand recalibration of the working voltage and programming current. Since the NRAM read and write mechanisms are fundamentally different from SRAM's charge-based storage, its programming operation requires a well-defined set/reset current to switch between high and low resistance states. The calibrated values are derived from [18], [41] and incorporated into the modified CACTI model to ensure accurate estimations of power and performance.

3) Peripheral circuit modeling: To accurately model NRAM within CACTI, the integration of a Current-Sensing Amplifier (CSA) is essential. This adaptation addresses the incapacity of traditional SRAM architecture to accommodate the unique resistive behaviors inherent to NRAM's operation, primarily due to the different nature of the memory's read and write mechanisms that rely on a bias voltage. The read current is expressed as:

$$I_{\text{read}} = \frac{V_{\text{bias}}}{R_{\text{NRAM}}} \tag{2}$$

Here, R_{NRAM} behaves in a manner consistent with resistive memory technology, where the ON state resistance, R_{on}, is approximately $10k\Omega$ and the OFF state resistance, R_{off}, is greater than $1M\Omega$. Such characteristics necessitate a different approach for the sense amplifier, transitioning from voltage-based to current-based sensing mechanisms. This adaptation not only allows for reduced power consumption but also increases the efficiency of the peripheral circuitry as shown in Fig. 8.

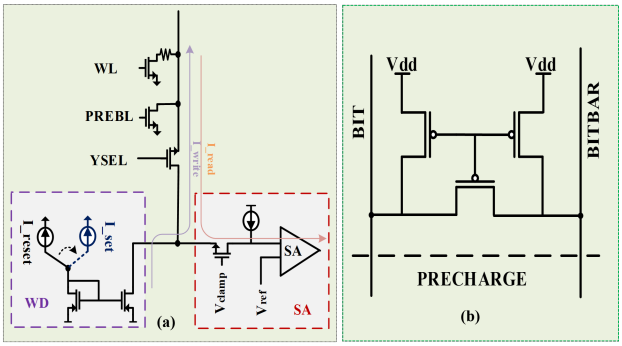

Fig. 8. The peripheral circuit structure: (a) Improved NRAM Current Read/Write Structure. (b) Traditional CMOS Voltage Read/Write Structure.

In the WD circuit, traditional CACTI implementation employs precharge and discharge mechanisms to force a high or low voltage onto the bitcell storage node [42]. However, for NRAM, the write process is fundamentally different, as it requires controlling the resistive state of the memory cell by injecting precise programming currents. Our modified WD circuit introduces a current-steering write driver, capable of delivering the necessary write currents while minimizing overshoot and energy loss. Additionally, to reduce write latency, a self-timed write termination mechanism is introduced, ensuring that the NRAM state switching is completed efficiently without unnecessary energy dissipation [43].

One of the key differences in the sense amplifier (SA) design is the transition from a voltage-sensing differential amplifier in conventional SRAM, to a current-based sensing scheme tailored for NRAM. The conventional CACTI SA model assumes low-swing bitline sensing, where a small differential voltage is generated and amplified using strong positive feedback latches. However, in NRAM, the read operation is resistance-based, meaning that the sensing process depends on current difference rather than voltage difference. To accommodate this, our modified SA architecture incorporates a high-gain current-sensing amplifier with clamping mechanisms to stabilize bitline voltage swings. This ensures robust read operations with minimal leakage and higher resilience to process variations as shown in Fig. 9.

By integrating these modifications, our CACTI-NRAM model achieves a more accurate estimation of read/write latency, power consumption, and area overhead, enabling comprehensive simulations and comparisons between conventional SRAM and NRAM.

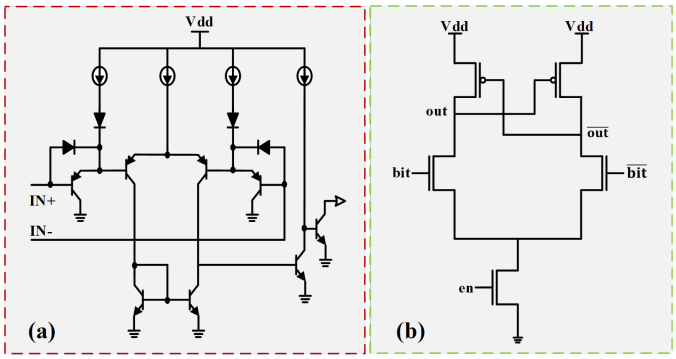

Fig. 9. The sense amplifier structure: (a) Improved Low-Power Sense Amplifier. (b) Traditional CMOS Sense Amplifier

4) Leakage Power modeling: Due to the fact that SRAM storage cells are composed of bistable flip-flops, they must constantly be supplied with a stable source of power, leading to high leakage power consumption in SRAM. The operating voltage of SRAM is typically set between $0.8V \sim 1.2V$ to ensure reliable inversion of CMOS inverters. However, NRAM stores data based on the physical contact and separation of carbon nanotubes, fundamentally making it a non-volatile memory that can retain data even when power is off. As such, NRAM static power consumption is significantly lower than that of SRAM. In this study, we have modified the static power consumption module of CACTI, replacing the leakage power component of SRAM with the retention power of NRAM, calculated as follows:

$$P_{\text{static, SRAM}} = I_{\text{leakage}} \times V_{\text{dd}} \tag{3}$$

$$P_{\text{static, NRAM}} \approx 0 \tag{4}$$

where I_{leakage} represents the leakage current of an SRAM storage cell, typically at the nanoampere (nA) level, while NRAM, due to its physical switching characteristics, exhibits virtually no static current loss in its quiescent state, with only minimal leakage currents under extreme temperatures, which are negligible ($< 1nW/cell$). Owing to this characteristic, NRAM saves nearly 100% of the static power consumption

in long-duration data retention scenarios (such as cache sleep mode) compared to SRAM.

IV. Experimental Results

A. Experimental Setup

At the circuit level, a Verilog-A model of the NEM relay was adopted from [3], [26] and implemented in Cadence Virtuoso for simulation. The NEM relay parameters were based on established models to ensure accuracy, including mechanical gaps, contact and channel resistance and capacitance. These parameters were carefully selected to accurately reflect the physical behavior of NEM relays, balancing switching performance and power efficiency. The $90nm$ NEM relay model parameters are listed in Table III, the read and write voltage and current parameters of $90nm$ NRAM are shown in Table IV, $90nm$ CMOS technology was used for comparisons.

TABLE III
NEM RELAY MODEL PARAMETERS IN $90nm$ TECHNOLOGY NODE

Parameter	Value
g_0 $[nm]$	10
g_d $[nm]$	5
R_{ch} $[\Omega]$	25.6
$I_{\text{on}}(V_p)$ $[\Omega]$	3880
R_{surf} $[\Omega]$	500
C_{ge} $[fF]$	0.9
C_{gb} $[fF]$	1.46
$C_{gd,s}$ $[aF]$	0.6

TABLE IV
NRAM PARAMETERS IN $90nm$ TECHNOLOGY NODE

Parameter	Value
Cell area	$0.36\mu m \times 0.36\mu m$
Set Condition of Cell Array	$2.5v$, $200ns$
Reset Condition of Cell Array	$3.5V$, $100ns$
Read Voltage	$0.5V$
Write Endurance	1×10^{12} cycles
Switching Speed	$0.5ns$

The logic synthesis for the MIPS controller were performed by "Logisim" [44]. The supply voltage of NEM relay was set to 1.2V and its input capacitance was set to $25fF$ [4].

For architectural evaluations, the microprocessor design was simulated using Gem5 [45] and McPAT [46], leveraging the SPEC2000 and SPEC2017 benchmark suite [47], [48] to analyze execution performance. In addition, we incorporated the Genann machine learning benchmark [49] to evaluate the processor's efficiency on lightweight neural network workloads. The architecture parameter settings for a MIPS-like microprocessor are shown in Table V. The branch predictor was set to a bimodal type with a specified branch target buffer size. Miss-prediction penalty, load-store queue size, and register update unit size were tuned to match realistic microprocessor configurations.

TABLE V
GEM5 PARAMETER SETTINGS

Parameter	Value
Instruction fetch queue size	4
Size of register files	1024×512
Extra branch mis-prediction latency	3
Branch predictor type	bimod
Bimodal predictor BTB size	2048
Instruction decode B/W	4
Instruction issue B/W	4
LSQ size	8
RUU size	16
Warmup instruction count	10000000
Checkpoint restore warmup	true
Total simulation instructions	100000000
Run pipeline with in-order issue	false
Number of integer ALU's	4
Number of floating point ALU's	4

B. Power, Performance and Area Evaluations on NEM relay-based MIPS Microprocessor

We conducted extensive PPA comparisons on different components of the NEM relay-based MIPS microprocessor.

1) Evaluations of the Controller: The PPA results are presented in Fig. 10, which compares the power consumption, area, and delay of the NEM relay-based controller with a conventional CMOS controller. The y-axis on the left represents power consumption in joules per second (w), plotted on a logarithmic scale axis, highlighting the significant difference in power efficiency between the two implementations. The y-axis on the right measures delay in nanoseconds (ns), allowing for a direct visual performance comparison.

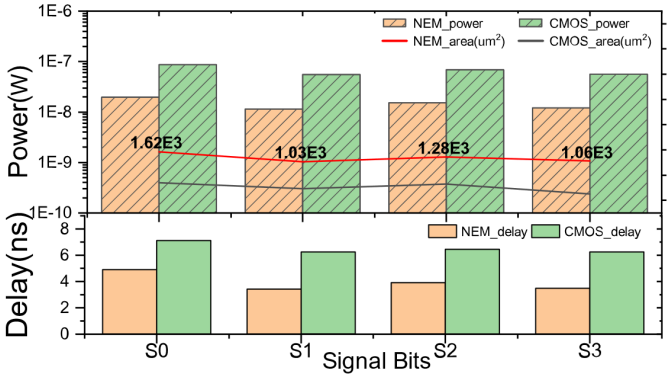

Fig. 10. Power, area and delay comparisons of NEM relay based controller and CMOS controller.

From the experimental results, it is evident that the NEM relay-based controller achieves a substantial power reduction, ranging from 77% to 79.1%, compared to the CMOS counterpart. Additionally, the delay is reduced by 31.1% to 45.1%, confirming the efficiency of the NEM relay-based design in terms of both energy consumption and performance.

Despite the increase in area, the significant reduction in power consumption makes the NEM relay-based controller

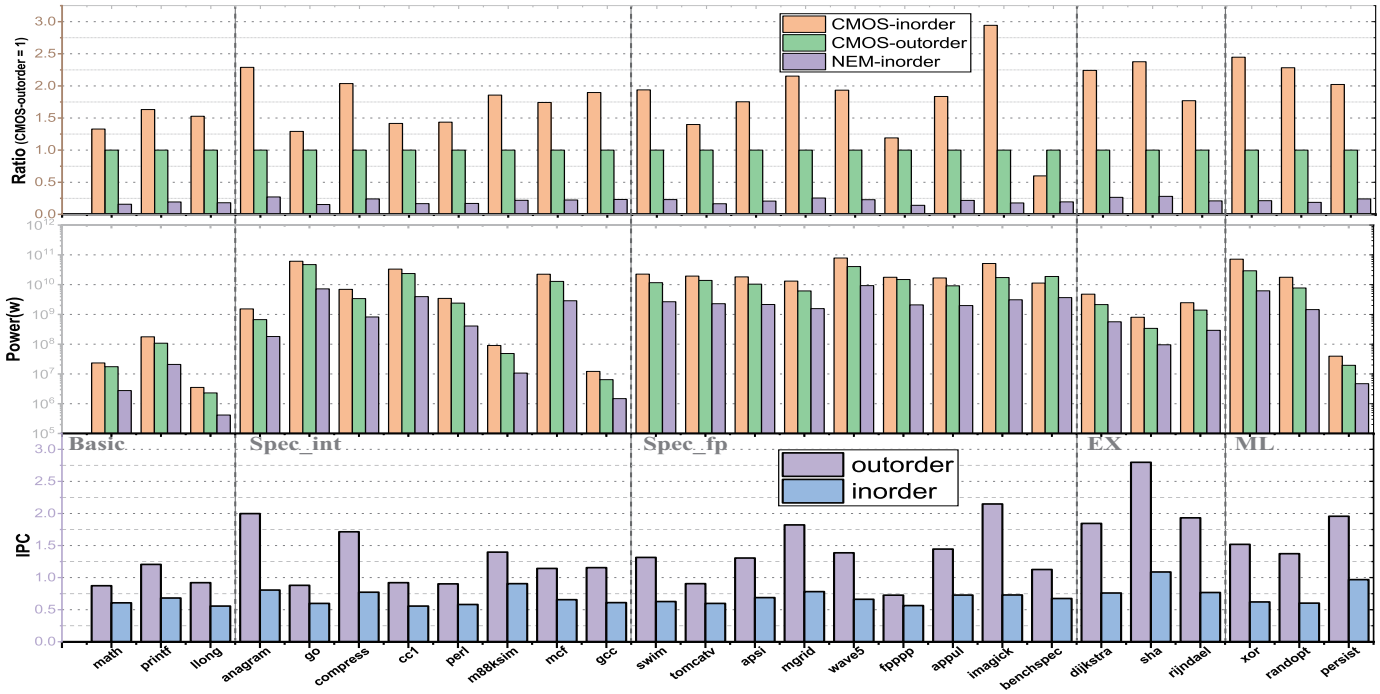

Fig. 11. Power and performance comparisons of NEM relay based processor and CMOS based processor.

a promising solution for ultra-low power edge computing applications. Moreover, 3D stacking [50], [51] and 6T-NEM techniques could help mitigate the area overhead [52], [53].

2) Evaluations of ALU Unit: We performed PPA evaluations of ALU unit from 1 bit to 32 bits, the results are shown in Fig. 12. For 32-bit, the power consumption of the NEM relay-based ALU is $1.441nw$, whereas the CMOS ALU consumes $10.637nw$ This indicates that the NEM ALU achieves an 86.5% power reduction compared to the CMOS implementation, demonstrating its significant energy efficiency.

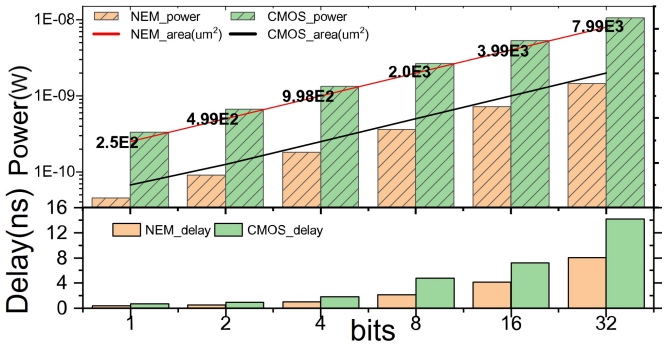

Fig. 12. Power, area and delay comparisons of NEM relay based ALU and CMOS ALU.

The delay for the NEM relay-based ALU at 32-bit is $8.04ns$, while the CMOS ALU delay is $14.15ns$, reflecting a 42.6% reduction in delay. This confirms that the NEM relay-based design not only reduces power consumption but also enhances computational speed by reducing signal propagation delay.

3) Evaluations of Register file: We also performed PPA evaluations of register file from 1 to 32 bits, as shown in Fig. 13.

For a 32-bit register file, the NEM relay-based one consumes $0.2445nw$, while the CMOS register consumes $1.8017nw$, demonstrating a nearly 87% power reduction. In terms of delay, the NEM relay-based register file achieves $0.18ns$, while the CMOS-based counterpart is $0.23ns$, showing a 19.9% delay improvement.

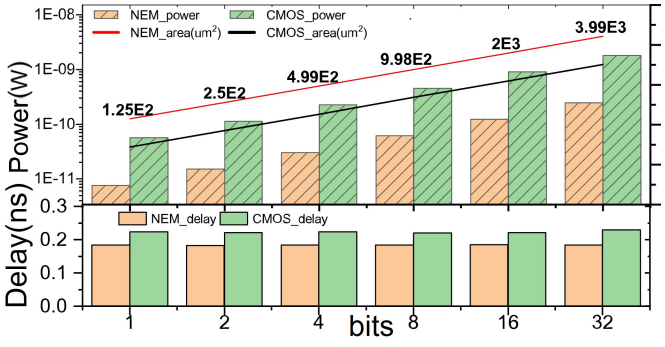

Fig. 13. Power, area and delay comparisons of NEM relay based register file and CMOS based register file.

4) Evaluations as a Whole Microprocessor System: The results shown in Fig. 11 compare power consumption and performance between NEM relay-based and CMOS-based processors on different benchmarks. The top graph shows the power ratio, the middle graph represents absolute power consumption on a logarithmic scale, and the bottom graph displays Instructions Per Cycle (IPC) across different workloads, including Basic tasks, SPEC Integer (Spec_int), SPEC Floating Point (Spec_fp), extra (EX) benchmarks, and machine learning (ML) workloads based on Genann.

The NEM relay-based processor consumes much less power than the CMOS processor in all cases. It uses only 11.8% of

TABLE VI
COMPARISON OF MEMORY TECHNOLOGIES

Memory Type	Area(mm²)	Read Latency(ps)	Write Latency(ns)	Read Energy(pJ)	Write Energy(pJ)	Leakage(mW)
NEM–NRAM	5.208	986.86	20.4	112.073	184.152	1.975
SRAM	1.532	656.974	0.641	375.267	356.734	92.738
STTRAM	3.145	1129.0	10.546	557.606	1033.0	16.444
PCRAM	0.299	492.239	95.406	232.493	138807.0	7.495
SLCNAND	22.479	1.21e7	7.25e5	1.85e6	2.12e7	1.170

the power compared to a non-pipelined microprocessor and saves over 78.9% power compared to an out-of-order CMOS processor. Notably, this trend persists across ML benchmarks, where NEM in-order execution maintains ultra-low energy profiles even under neural workload execution.

We also observed that out-of-order execution consumes less power than in-order CMOS execution in some cases, due to better utilization of instruction-level parallelism (ILP), reduction of pipeline stalls, and lower per-instruction energy. This effect is particularly noticeable when the issue width is set to 4, as the processor can dispatch multiple instructions per cycle, leading to improved execution efficiency. In the middle graph, CMOS in-order and out-of-order processors both consume more power, while the NEM in-order processor remains significantly more energy-efficient across all benchmark types.

In terms of performance, the bottom graph shows IPC comparisons. The out-of-order processor achieves higher IPC in most cases, but for a four-issue microprocessor running at 50 MHz with an energy limit of 9.26×10^9 J, the NEM–NRAM processor maintains an IPC of 4, which is 42.9% faster than the out-of-order CMOS microprocessor. For Spec_int benchmarks, which involve integer computing intensive tasks like compression and encryption, out-of-order execution performs better, but the NEM relay-based in-order processor remains competitive. In Spec_fp benchmarks, which focus on floating-point calculations, NEM processors deliver good performance with much lower power consumption. Similarly, under ML workloads, NEM processors provide balanced IPC and superior energy efficiency, suggesting strong potential for lightweight neural inference in edge environments.

Overall, these results show that the NEM relay-based processor is much more power-efficient while still maintaining strong performance, making it well-suited for emerging low-power applications such as edge AI, neural inference, and other energy-constrained domains.

C. Area Evaluation and 3D Integration

To address the relatively higher area cost of NEM and NRAM devices compared to scaled CMOS, we adopted a 3D integration strategy to improve system compactness and performance. Our architecture comprises three vertically stacked tiers: a non-volatile NRAM memory layer and two logic tiers built with NEM relays. This separation of storage and logic enables high integration density, mitigates footprint overhead, and facilitates energy-efficient data movement.

Our area estimation is based on characterized or fabricated devices in this work:

- NRAM cell area = $0.36 \times 0.36\,\mu m^2 = 0.1296\,\mu m^2$
- NEM relay area = $12\,\mu m^2$ per device

Layer 1 consists of a 16 KB 1T1R NRAM array, with an estimated 20% area overhead from sensing and peripheral circuitry. Layer 2 implements the ALU and controller: the ALU requires 576 relays, while the controller uses approximately 800 relays. An additional 10% routing overhead is assumed. Layer 3 is a 1024-bit register file, implemented with 2048 relays in total, plus interconnect overhead.

The total area across tiers is balanced, with logic and memory layers sized comparably. This facilitates compact vertical stacking, minimizes interconnect energy, and enables better thermal balance.

TABLE VII
AREA BREAKDOWN OF 3D STACKED SYSTEM

Layer	Components	Area(mm^2)
Layer1(NRAM)	Array & Periphery	20.4
Layer2(NEM1)	ALU & Controller	18.2
Layer3(NEM2)	Register File	27.0

D. Comparisons with Emerging Low-Power Device Alternatives

To assess the competitiveness of our NEM–NRAM architecture, we perform a comparative evaluation against both conventional and emerging memory classes, including SRAM, STT-MRAM, PCRAM, and SLC-NAND. These technologies have been explored for low-power or non-volatile applications: for instance, STT-MRAM provides fast access but incurs high switching energy [54], PCRAM achieves high density but suffers from limited endurance [55], and SLC-NAND offers large capacity with relatively high latency [22]. By contrast, NEM–NRAM uniquely combines the ultra-low leakage of nano-electro-mechanical relays [16] with the non-volatility and bit-addressability of NRAM [22], [39]. As summarized in Table VI, it achieves competitive latency, improved write energy efficiency (184.2 pJ), favorable read energy (112.7 pJ), and minimal leakage (1.98 mW), making it a compelling solution for scalable low-power integrated systems.

V. CONCLUSION

In this work, we introduce two emerging semiconductor devices, NEM and NRAM, which have ultra low power consumption, low latency and high endurance. Then, we propose a RSIC architecture built with these devices, which can complete instruction in a single NEM switching delay. Experimental

results show that compared to its CMOS counterpart, the proposed design reduces power consumption by 88.2% in in-order processors and by 78.9% in out-of-order pipelined designs. The performance can be improved by 42.9% compared to out of order CMOS microprocessors.

Through detailed area breakdown and 3D integration planning, we demonstrate that NEM–NRAM architectures can mitigate footprint overhead and achieve compact logic-memory stacking. Compared to emerging low-power memory alternatives, our design achieves favorable energy-leakage tradeoffs. Future research will focus on fine-grained 3D integration strategies, and interconnect co-design to further reduce area and routing overhead.

REFERENCES

[1] H. Kam, T.-J. K. Liu, V. Stojanović, D. Marković, and E. Alon, "Design, optimization, and scaling of mem relays for ultra-low-power digital logic," *IEEE Transactions on electron devices*, vol. 58, no. 1, pp. 236–250, 2010.

[2] G. Rosendale, S. Kianian, M. Manning, D. Hamilton, X. H. Huang, K. Robinson, Y. W. Kim, and T. Rueckes, "A 4 megabit carbon nanotube-based nonvolatile memory (nram)," in *2010 Proceedings of ESSCIRC*. IEEE, 2010, pp. 478–481.

[3] M. Spencer, F. Chen, C. C. Wang, R. Nathanael, H. Fariborzi, A. Gupta, H. Kam, V. Pott, J. Jeon, T.-J. K. Liu *et al.*, "Demonstration of integrated micro-electro-mechanical relay circuits for vlsi applications," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 308–320, 2010.

[4] B. Z. Tang, S. A. Bhave, and R. Manohar, "Low power asynchronous vlsi with nem relays," in *2014 20th IEEE International Symposium on Asynchronous Circuits and Systems*. IEEE, 2014, pp. 85–92.

[5] R. Li and H. Fariborzi, "Design and demonstration of mem relay-based arithmetic and sequential circuit blocks," *IEEE Transactions on Electron Devices*, vol. 68, no. 12, pp. 6415–6421, 2021.

[6] S. Rana, J. Mouro, S. J. Bleiker, J. D. Reynolds, H. M. Chong, F. Niklaus, and D. Pamunuwa, "Nanoelectromechanical relay without pull-in instability for high-temperature non-volatile memory," *Nature communications*, vol. 11, no. 1, p. 1181, 2020.

[7] J. D. Reynolds, S. Rana, E. Worsey, Q. Tang, M. K. Kulsreshath, H. M. Chong, and D. Pamunuwa, "Single-contact, four-terminal micro-electromechanical relay for efficient digital logic," *Advanced Electronic Materials*, vol. 9, no. 1, p. 2200584, 2023.

[8] Y. Li, E. Worsey, S. J. Bleiker, P. Edinger, M. K. Kulsreshath, Q. Tang, A. Y. Takabayashi, N. Quack, P. Verheyen, W. Bogaerts *et al.*, "Integrated 4-terminal single-contact nanoelectromechanical relays implemented in a silicon-on-insulator foundry process," *Nanoscale*, vol. 15, no. 43, pp. 17 335–17 341, 2023.

[9] D. Pamunuwa, E. Worsey, J. D. Reynolds, D. Seward, H. M. Chong, and S. Rana, "Theory, design, and characterization of nanoelectromechanical relays for stiction-based non-volatile memory," *Journal of Microelectromechanical Systems*, vol. 31, no. 2, pp. 283–291, 2022.

[10] H. Saito, J. Watanabe, J. Seino, T. Tamura, N. Sashida, K. Hara, K. Kawabata, A. Fujii, J. Ohno, A. Nakakubo *et al.*, "Development of 16 mb nram aiming for high reliability, small cell area, and high switching speed," in *2021 IEEE International Memory Workshop (IMW)*. IEEE, 2021, pp. 1–4.

[11] B. Gervasi, "Will carbon nanotube memory replace dram?" *IEEE Micro*, vol. 39, no. 2, pp. 45–51, 2019.

[12] N. Muralimanohar, R. Balasubramonian, and N. P. Jouppi, "Cacti 6.0: A tool to model large caches," *HP laboratories*, vol. 27, p. 28, 2009.

[13] F. Chen, H. Kam, D. Markovic, T.-J. K. Liu, V. Stojanovic, and E. Alon, "Integrated circuit design with nem relays," in *2008 IEEE/ACM International Conference on Computer-Aided Design*. IEEE, 2008, pp. 750–757.

[14] W. S. Lee, *Nanoelectromechanical Relays for Low Power Digital Systems*. Stanford University, 2013.

[15] M. K. Kulsreshath, Q. Tang, E. Worsey, M. B. Krishanan, Y. Li, S. J. Bleiker, F. Niklaus, and D. Pamunuwa, "Digital nanoelectromechanical non-volatile memory cell," *IEEE Electron Device Letters*, vol. 45, no. 4, pp. 728–731, 2024.

[16] S. Yazdanshenas, B. Khaleghi, P. Ienne, and H. Asadi, "Designing low power and durable digital blocks using shadow nanoelectromechanical relays," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 12, pp. 3489–3498, 2016.

[17] D. Veksler, G. Bersuker, A. W. Bushmaker, M. Mason, P. Shrestha, K. P. Cheung, J. P. Campbell, T. Rueckes, L. Cleveland, H. Luan *et al.*, "Memory update characteristics of carbon nanotube memristors (nram®) under circuitry-relevant operation conditions," in *2020 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 2020, pp. 1–4.

[18] S. Ning, T. O. Iwasaki, K. Shimomura, K. Johguchi, G. Rosendale, M. Manning, D. Viviani, T. Rueckes, and K. Takeuchi, "23% faster program and 40% energy reduction of carbon nanotube non-volatile memory with over 1011 endurance," in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*. IEEE, 2014, pp. 1–2.

[19] P. Pandey *et al.*, "Role of nanotechnology in electronics: A review of recent developments and patents," *Recent patents on nanotechnology*, vol. 16, no. 1, pp. 45–66, 2022.

[20] R. Sen, T. Kocab, J. Black, J. McDermott, S. K. Pal, S. Buffat, D. C. Gilmer, and T. Rueckes, "Engineering nanomaterials and nanostructures for electronic applications: A case study of carbon nanotubes for memory devices," in *2020 4th IEEE Electron Devices Technology & Manufacturing Conference (EDTM)*. IEEE, 2020, pp. 1–4.

[21] D. Pamunuwa, E. Worsey, Q. Tang, M. K. Kulsreshath, V. Marot, Y. Li, and S. Bleiker, "Hardware platform for edge computing based on nanoelectromechanical relays," in *2023 22nd International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers)*. IEEE, 2023, pp. 537–541.

[22] D. C. Gilmer, T. Rueckes, and L. Cleveland, "Nram: a disruptive carbon-nanotube resistance-change memory," *Nanotechnology*, vol. 29, no. 13, p. 134003, 2018.

[23] M. Gupta, K. Gupta, and N. Pandey, "Comparative analysis of the design techniques for low leakage srams at 32nm," *Microprocessors and Microsystems*, vol. 85, p. 104281, 2021.

[24] H. R. Shea, "Effects of radiation on mems," in *Reliability, Packaging, Testing, and Characterization of MEMS/MOEMS and Nanodevices X*, vol. 7928. SPIE, 2011, pp. 96–108.

[25] A. Keys, J. Adams, J. Cressler, M. Johnson, and M. Patrick, "A review of nasa's radiation-hardened electronics for space environments project," in *AIAA Space 2008 Conference & Exposition*, 2008, p. 7673.

[26] M. E. Spencer, *Design Considerations for Nano-Electromechanical Relay Circuits*. University of California, Berkeley, 2015.

[27] S. Ning, T. O. Iwasaki, D. Viviani, H. Huang, M. Manning, T. Rueckes, and K. Takeuchi, "Nram: High performance, highly reliable emerging memory," 2016.

[28] T. Qin, S. J. Bleiker, S. Rana, F. Niklaus, and D. Pamunuwa, "Performance analysis of nanoelectromechanical relay-based field-programmable gate arrays," *IEEE Access*, vol. 6, pp. 15 997–16 009, 2018.

[29] Y. Yokoyama, "Design optimization for low-power and highly reliable embedded srams on advanced cmos platforms," *Kyoto Institute of Technology*, 2021.

[30] R. Venkatasubramanian, S. K. Manohar, and P. T. Balsara, "Nem relay-based sequential logic circuits for low-power design," *IEEE transactions on nanotechnology*, vol. 12, no. 3, pp. 386–398, 2013.

[31] N. M. Nair and R. M. Kuruvithadam, "Nem relay based multi-nary ram a memory with optimized power, area and speed," in *2015 International Conference on Control Communication & Computing India (ICCC)*. IEEE, 2015, pp. 563–568.

[32] B. Smith, M. A. Mamun, B. Horstmann, Ü. Özgür, and V. Avrutin, "Multi-gate in-plane actuated nems relays for effective complementary logic gate designs," *Journal of Microelectromechanical Systems*, 2023.

[33] D. Lee, W. S. Lee, C. Chen, F. Fallah, J. Provine, S. Chong, J. Watkins, R. T. Howe, H.-S. P. Wong, and S. Mitra, "Combinational logic design using six-terminal nem relays," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 5, pp. 653–666, 2013.

[34] J. E. Lamb, S. Gibbons, Y. Jiang, K. Mangelson, K. Kremer, D. Janzen, J. Bledsoe, and M. Boeser, "Advancement of microelectronics-grade carbon nanotube materials for nram® device manufacture," in *25th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC 2014)*. IEEE, 2014, pp. 125–129.

[35] C. Li, "Architecture study of a 3d cmos-nem fpga," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2013.

[36] A. Levy, M. Oduoza, A. Balasingam, R. T. Howe, and P. Raina, "3-d coarse-grained reconfigurable array using multi-pole nem relays for programmable routing," *Integration*, vol. 88, pp. 249–261, 2023.
[37] M. Lefter, M. Enachescu, G. R. Voicu, and S. D. Cotofana, "Energy effective 3d stacked hybrid nemfet-cmos caches," in *Proceedings of the 2014 IEEE/ACM International Symposium on Nanoscale Architectures*, 2014, pp. 151–156.
[38] C. Chen, R. Parsa, N. Patil, S. Chong, K. Akarvardar, J. Provine, D. Lewis, J. Watt, R. T. Howe, H.-S. P. Wong *et al.*, "Efficient fpgas using nanoelectromechanical relays," in *Proceedings of the 18th annual ACM/SIGDA international symposium on Field programmable gate arrays*, 2010, pp. 273–282.
[39] T. Rueckes, "High density, high reliability carbon nanotube nram," *Flash Memory Summit*, 2011.
[40] P. S. Kanhaiya, C. Lau, G. Hills, M. D. Bishop, and M. M. Shulaker, "Carbon nanotube-based cmos sram: 1 kbit 6t sram arrays and 10t sram cells," *IEEE Transactions on Electron Devices*, vol. 66, no. 12, pp. 5375–5380, 2019.
[41] S. Ning, T. O. Iwasaki, K. Shimomura, K. Johguchi, E. Yanagizawa, G. Rosendale, M. Manning, D. Viviani, T. Rueckes, and K. Takeuchi, "Investigation and improvement of verify-program in carbon nanotube-based nonvolatile memory," *IEEE Transactions on Electron Devices*, vol. 62, no. 9, pp. 2837–2844, 2015.
[42] S. Li, K. Chen, J. H. Ahn, J. B. Brockman, and N. P. Jouppi, "Cacti-p: Architecture-level modeling for sram-based structures with advanced leakage reduction techniques," in *2011 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE, 2011, pp. 694–701.
[43] D. Jiang, Q. Chen, Z. Li, Q. Shan, Z. Wei, J. Xiao, and S. Huang, "The design of a low noise and low power current readout circuit for sub-pa current detection based on charge distribution model," *Electronics*, vol. 11, no. 11, p. 1791, 2022.
[44] C. Burch, "Logisim: A graphical system for logic circuit design and simulation," *Journal on Educational Resources in Computing (JERIC)*, vol. 2, no. 1, pp. 5–16, 2002.
[45] J. Lowe-Power, A. M. Ahmad, A. Akram, M. Alian, R. Amslinger, M. Andreozzi, A. Armejach, N. Asmussen, B. Beckmann, S. Bharadwaj *et al.*, "The gem5 simulator: Version 20.0+," *arXiv preprint arXiv:2007.03152*, 2020.
[46] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi, "Mcpat: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proceedings of the 42nd annual ieee/acm international symposium on microarchitecture*, 2009, pp. 469–480.
[47] J. L. Henning, "Spec cpu2000: Measuring cpu performance in the new millennium," *Computer*, vol. 33, no. 7, pp. 28–35, 2002.
[48] J. Bucek, K.-D. Lange, and J. v. Kistowski, "Spec cpu2017: Next-generation compute benchmark," in *Companion of the 2018 ACM/SPEC International Conference on Performance Engineering*, 2018, pp. 41–42.
[49] H. Xu, J. Xiang, Z. Huang, G. Chen, Y. Meng, and H. Zhu, "Latte: Layered attestation for portable enclaved applications," in *2025 IEEE 10th European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2025, pp. 339–354.
[50] F.-K. Hsueh, H.-Y. Chiu, C.-H. Shen, J.-M. Shieh, Y.-T. Tang, C.-C. Yang, H.-C. Chen, W.-H. Huang, B.-Y. Chen, K.-M. Chen *et al.*, "Tsv-free finfet-based monolithic 3d+-ic with computing-in-memory sram cell for intelligent iot devices," in *2017 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2017, pp. 12–6.
[51] C. Dong, C. Chen, S. Mitra, and D. Chen, "Architecture and performance evaluation of 3d cmos-nem fpga," in *International Workshop on System Level Interconnect Prediction*. IEEE, 2011, pp. 1–8.
[52] H. Fariborzi, F. Chen, V. Stojanovic, R. Nathanael, J. Jeon, and T.-J. K. Liu, "Design and demonstration of micro-electro-mechanical relay multipliers," in *IEEE Asian Solid-State Circuits Conference 2011*. IEEE, 2011, pp. 117–120.
[53] A. Y. Zhixin, S. F. Almeida, U. Sikder, X. Hu, T. K. Esatu, K. Le, J. Jeon, and T.-J. K. Liu, "Body-biased multiple-gate micro-electro-mechanical relays," *IEEE Electron Device Letters*, vol. 42, no. 3, pp. 402–405, 2021.
[54] N. Chaturvedi, "A novel low power non-volatile sram cell with self write termination," 2019.
[55] X. Wang, S. Song, H. Wang, T. Guo, Y. Xue, R. Wang, H. Wang, L. Chen, C. Jiang, C. Chen *et al.*, "Minimizing the programming power of phase change memory by using graphene nanoribbon edge-contact," *Advanced Science*, vol. 9, no. 25, p. 2202222, 2022.

Siyuan Lu received his B.S. degree from Wuhan University of Technology, Wuhan, China. He is persuing his M.S. degree of Integrated Circuits at Beihang University. His research include design and optimization of low-power circuits based on carbon nanotube devices.

Yuanqing Cheng (S'11, M'13, SM'20) is a tenured associate professor of the School of Integrated Circuit and Engineering, Beihang University, Beijing, China. He received his Ph.D. degree from the Key Laboratory of Computer System and Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. After spending one year postdoc study at LIRMM, CNRS, France, he joined Beihang University. His research interests include design automation for 3D integrated circuits, as well as low power and architectural optimization of emerging semiconductor technologies. He is currently an editorial board member of Elsevier Integration, the VLSI journal, and a senior member of the IEEE.

Ying Wang Dr. Ying Wang is a professor in Institute of Computing Technology, Chinese Academy of Sciences, Beijing, and his current research interest includes the chip design automation, reliable computer architecture and memory system. He has published over 100 papers on IEEE/ACM conferences and journals, including TC, DAC, MICRO, HPCA and ICCAD. He has received several awards from international conferences, including the championship of IEEE LPIRC contest at DAC 2016, the championship of the System Design Contest at DAC 2018, 2024 IEEE Top Picks on IC testability and reliability, the Best Paper Award at ITC-Asia, GLSVLSI and ICCD. He is also the recipient of IEEE/ACM DAC under 40 innovator at 2021.

Rui Wang (M'10, SM'21) received the B.S. and M.S. degrees in computer science from Xi'an Jiaotong University, Xi'an, China, in 2000 and 2003, respectively, and the Ph.D. degree in computer architecture from Beihang University, Beijing, China, in 2009. He is an Associate Professor with the School of Computer Science and Engineering, Beihang University. His current research interests include computer architecture and deep learning compiler. Dr. Wang is a senior member of IEEE.