
Floating-point Unit (FPU) Designs with Nano-electromechanical (NEM) Relays

by

Sumit Dutta

B.S., Electrical Engineering, University of Illinois at Urbana-Champaign, 2011

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

September 2013

© 2013 Massachusetts Institute of Technology
All Rights Reserved.

Signature of Author: _____

Sumit Dutta
Department of Electrical Engineering and Computer Science
August 29, 2013

Certified by: _____

Vladimir Stojanović
Associate Professor of Electrical Engineering
Thesis Supervisor

Accepted by: _____

Leslie A. Kolodziejcki
Professor of Electrical Engineering and Computer Science
Chair, Committee for Graduate Students

Floating-point Unit (FPU) Designs with Nano-electromechanical (NEM) Relays

by Sumit Dutta

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science

Abstract

Nano-electromechanical (NEM) relays are an alternative to CMOS transistors as the fabric of digital circuits. Circuits with NEM relays offer energy-efficiency benefits over CMOS since they have zero leakage power and are strategically designed to maintain throughput that is competitive with CMOS despite their slow actuation times. The floating-point unit (FPU) is the most complex arithmetic unit in a computational system. This thesis investigates if the energy-efficiency promise of NEM relays demonstrated before on smaller circuit blocks holds for complex computational structures such as the FPU. The energy, performance, and area trade-offs of FPU designs with NEM relays are examined and compared with that of state-of-the-art CMOS designs in an equivalent scaled process. Circuits that are critical path bottlenecks, including primarily the leading zero detector (LZD) and leading zero anticipator (LZA) blocks, are carefully identified and optimized for low latency and device count. We manage to drop the NEM relay FPU latency from 71 mechanical delays in a CMOS-style implementation to 16 mechanical delays in a NEM relay pass-logic style implementation. The FPU designed with NEM relays features 15x lower energy per operation compared to CMOS.

Thesis Supervisor: Vladimir Stojanović

Title: Associate Professor of Electrical Engineering

Acknowledgments

The long strides we have made with NEM relay integrated circuits would not be possible without the stalwart efforts of everyone on our collaborative research team with members from MIT, University of California at Berkeley, and University of California at Los Angeles. I appreciate the sincere efforts of Professor Elad Alon and Professor Dejan Marković to maintain a positive attitude and encourage us to continue trailblazing on this research journey. I am most indebted to my advisor, Professor Vladimir Stojanović, for cultivating my critical thinking in circuits research and guiding my studies on NEM relays.

Many thanks go to Ofer Shacham of Stanford University in helping me get set up with Genesis2 and FPGen. I also thank Andrew Wright for showing me how he had FPGen running within our research group.

One of the early champions of the NEM relays project, Matthew Spencer, has also been an invaluable mentor to me for our NEM relay chip tape-outs and testing efforts.

I could not have made it through without the personal support and friendship of Karen Ng, a physicist and former classmate, and George Garcia, an engineer whom I also befriended at UIUC. At MIT, it has been a real pleasure to be a part of the Sidney-Pacific Graduate Community and to have the close friendship of Fabián, Nate, Aditya, Sunny, JP, and Shawn when he was in town. These friends as well as many more have helped me remain level-headed and made my experience at MIT all the more better.

Last but not least, I thank my parents, Subijoy and Urmi Dutta, and my brother Sanjit, for holding it together as a supportive family, and I thank FIRST Robotics Team 1111 for propelling my initial interests in science and technology.

Contents

Abstract	3
Acknowledgments	4
List of Figures	9
1 Introduction	13
1.1 Thesis Outline	15
2 NEM Relay Devices and Circuit Design	17
2.1 Structure of NEM Relays	17
2.2 Logical Description of NEM Relays	22
2.3 Circuit Design with NEM Relays	22
2.4 Process Technology and Scaling	25
3 Experiments on NEM Relay Chips	27
3.1 Scaled NEM Relays	28
3.2 Rudimentary Circuits and Tests	30
4 Floating-Point Numbers and the Floating-Point Unit (FPU)	33
4.1 Floating-Point Numbers	33
4.2 FPU Design	34
4.3 Representing FPU Design with the Fused Multiply-Add Architecture . .	36
5 The Fused Multiply-Add (FMA) Designed with NEM Relays	39
5.1 Shortcomings of CMOS-Style Circuit Design	39
5.2 Implementing the FMA Architecture in Hardware	40
5.3 NEM Relays CAD Infrastructure for Synthesis	42
5.4 Decoders and Multiplexers with NEM Relays	45
5.5 Shifters with NEM Relays	47
5.6 The Leading Zero Detector (LZD) Optimized for NEM Relays	51
5.7 The Leading Zero Anticipator (LZA) Optimized for NEM Relays	58

5.8	Circuit Topology Analysis for Area, Energy, and Delay Trade-offs	58
6	Energy and Performance Analysis and Comparison with CMOS	61
6.1	Basis for Analysis	61
6.2	Process Technology	62
6.3	NEM Relays CAD Infrastructure for Verification and Analysis	63
6.4	Energy per Operation for Each Subcircuit	63
6.5	FMA Analysis and Comparison between CMOS and NEM Relays	65
6.6	Interconnect Model Including Scaling	65
7	Conclusions	69
7.1	Future Directions	69
	Bibliography	71

List of Figures

1.1	Energy per operation for circuits made with transistors is limited by a minimum energy point. Data courtesy of V. Stojanović.	14
1.2	Ideal switching characteristics exhibited by nano-electromechanical relays. Left: energy per operation can continue to scale down for circuits made with NEM relays due to no leakage energy. Right: the current voltage characteristic of NEM relays. Data courtesy of R. Nathanael [4].	14
2.1	Structure of the four-terminal (4T) nano-electromechanical (NEM) relay. When the voltage between the gate and body is low, the device is in the <i>off</i> -state with no current; when it is high, the device is in the <i>on</i> -state with a current I_{DS}	18
2.2	Static analysis of the 4T NEM relay whose dimensions are in Table 2.2. The electrostatic force F_{elec} is plotted for different voltages V against the spring force F_{spring} for all displacements of the movable gate, where $z = 0$ is the <i>off</i> -state. This analysis is with zero surface adhesion force. . . .	20
2.3	Seesaw NEM relay [8].	21
2.4	Equivalent nMOS and pMOS drop-in replacements with appropriately biased NEM relays.	23
2.5	Inverter and buffer designed with NEM relays. The input is A, the inverter output is A_inv, and the buffer output is A_buf. 4T, 6T, and seesaw NEM relay designs are shown.	23
2.6	Electrical vs. mechanical delay for a long stack of NEM relays in current ($1 \mu\text{m}$) and scaled (90 nm) NEM relay process technology, courtesy of H. Fariborzi [5].	24
2.7	The same circuit is built with CMOS and NEM relays. Less NEM relays are required than transistors. Schematics are courtesy of F. Chen [1]. . .	24
2.8	Schematic and measured voltage transfer curve of a NEM relay NOT or XOR gate showing hysteresis and full rail swing at the output. Data from [2].	25

3.1	On the top left is the full adder output courtesy of F. Chen, on the top right is the (7:3) compressor output courtesy of H. Fariborzi, and on the bottom is a micrograph of the CLICKR2 test chip containing the circuits.	28
3.2	From left to right are a high V_{pi} NEM relay and the CLICKR4 test chip containing those devices.	29
3.3	From left to right: the transfer characteristic for a low V_{pi} NEM relay with long L , that for a high V_{pi} NEM relay with short L , and a micrograph of the CLICKR3 test chip containing those devices. Measured parameters for those devices are tabulated.	29
3.4	Microcontroller test chips, CLICKR5 (left) and CLICKR6 (right).	31
3.5	Flip-flop with oxide breaker devices on CLICKR5, courtesy of C. C. Wang (top). Partial buffer test on a CLICKR6 flip-flop (bottom). In the CLICKR6 buffer test, CLK2 is high and the pulse on D is followed at the output MID.	32
4.1	Generic single-precision FPU design.	35
4.2	Single-precision FPU design with a fused multiply-add.	35
5.1	Reducing the number of mechanical delays for the fused multiply-add (FMA). Adder (ADD) and multiplier (MUL) optimizations are recommended by [2, 5]. Additional decoder (DEC) and multiplexer (MUX) optimizations are the beginning of further optimizations introduced in this work.	40
5.2	AND gate, OR gate, full adder, and 32-bit adder designed with 6T and seesaw NEM relays. All circuits produce their result within a single mechanical delay.	41
5.3	Fused multiply-add (FMA) designed with CMOS.	43
5.4	Generic optimal design of a decoder with NEM relays.	46
5.5	2:1 multiplexers designed with 4T, 6T, and seesaw NEM relays. The 6T NEM relay implementation is a dual 2:1 multiplexer.	46
5.6	8:1 multiplexers designed with 4T, 6T, and seesaw NEM relays. An 8:1 multiplexer comprises 2:1 multiplexers.	48
5.7	The 8:1 multiplexer of Fig. 5.6 may have a constant signal (e.g. 0 in this figure) on some inputs. For this case, the multiplexer tree is pruned to use the least number of NEM relays.	49
5.8	Barrel shifters used for right shift (top) and left shift (bottom) operations.	50
5.9	Complexity of LZDs. These are synthesized LZDs before pruning unnecessary relays.	52
5.10	Tree design used as the RTL for area-efficient LZDs. This particular design is for an LZD with 8 inputs.	53
5.11	Generic modular design of an LZD with n input bits using two LZDs with half the number of input bits [31].	54

5.12 LZD with 2 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices.	54
5.13 LZD with 4 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices.	55
5.14 LZD with 8 input bits designed with 6T NEM relays. The circuit runs in one mechanical delay. This circuit design is synthesized and dummy NEM relays introduced are not removed from this schematic, making it less intuitive.	55
5.15 LZD with 8 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices, and it provides the intuition for higher-bit single- t_{mech} LZD designs.	56
5.16 LZD with 16 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices, and it provides significant intuition for higher-bit single- t_{mech} LZD designs.	57
5.17 Fused multiply-add (FMA) designed with NEM relays.	60
6.1 Comparison of the switching power estimation and the circuit power consumption from a physical-model-based simulation for the same circuits designed in the present scaled 6T NEM relay process. Both the functional Verilog simulation and Verilog-A-based simulation use $f = 100$ kHz. This shows that switching power estimation is a reasonable approximation in most cases.	64
6.2 Energy and performance trade-off for equivalent fused multiply-add circuits made in the 90nm CMOS technology node [24] and predictive 90nm equivalent NEM relays technology node (top). The total latency of the NEM relay FMA depends on V_{DD} (bottom).	66
6.3 A placed and routed LZD2 circuit in the scaled 6T process technology, whose physical data is scaled to the predictive 90nm equivalent 6T NEM relays process technology and used alongside a predictive 90nm equivalent 6T NEM relays process interconnect capacitance table file to estimate post-route switching power.	67

Introduction

NANORELAYS, short for nano-electromechanical (NEM) relays, are electrostatically actuated mechanical switches. They are an alternative to transistors as the building block for digital circuits. Digital circuits are traditionally implemented with complementary metal-oxide-semiconductor (CMOS) transistor logic, and such logic can also be implemented with NEM relays [1]. While the process technology for CMOS continues to scale down to smaller sizes, its unwanted leakage power has also been increasing exponentially. On the other hand, NEM relays offer zero leakage power for any type of digital circuit, and its process technology can be scaled down like that of transistors [2].

CMOS scaling has enabled the energy per digital computation to go down by reducing the power supply, V_{DD} , and the threshold voltage, V_T . However, this scaling approach reaches a minimum energy point as the once dominant dynamic energy $E_{dynamic}$ begins competing with leakage energy $E_{leakage}$, as seen in Fig. 1.1 [3]. The reduction of V_T leads to subthreshold conduction that contributes to $E_{leakage}$. This has forced a move to running digital circuits in parallel as in multi-core processors, but eventually this will become ineffective [1].

NEM relays are nearly ideal switches that have an infinite subthreshold slope and do not exhibit leakage power as seen in the current-voltage characteristic in Fig. 1.2 [4]. Since the energy per circuit operation is then constrained only by dynamic switching power, the energy per operation can continue to scale down beyond that for CMOS. The potentially higher energy-efficiency of digital circuits made of NEM relays compared to those made of transistors makes NEM relays worth exploring.

The relatively high yields of fabricated NEM relays have enabled the design and test of increasingly complex NEM relays circuits. A NEM relay based inverter, latch, and full adder have been demonstrated [2]. Additionally, the (7:3) compressor needed for a NEM relays based multiplier circuit has also been demonstrated, making it the largest working NEM relays based circuit with 98 relays [5]. This opened the door to explore very large scale integrated (VLSI) circuits with NEM relays.

Thus far, prior work demonstrated the energy-efficiency advantage of NEM relays over CMOS, with NEM relay adders and multipliers as the most complex blocks. These are mostly realized through circuit design techniques that balance out the slow mechanical delay with the fast electrical delay in the NEM relay. The question remains if such

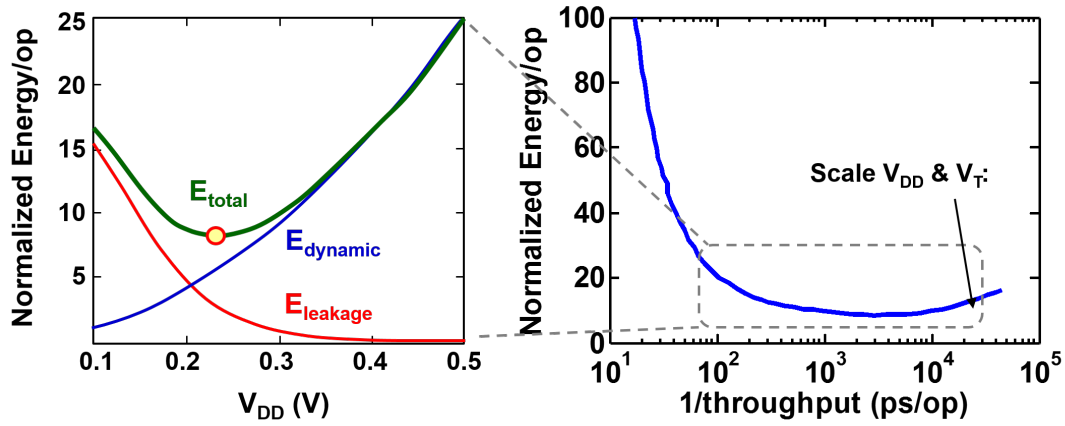


Figure 1.1. Energy per operation for circuits made with transistors is limited by a minimum energy point. Data courtesy of V. Stojanović.

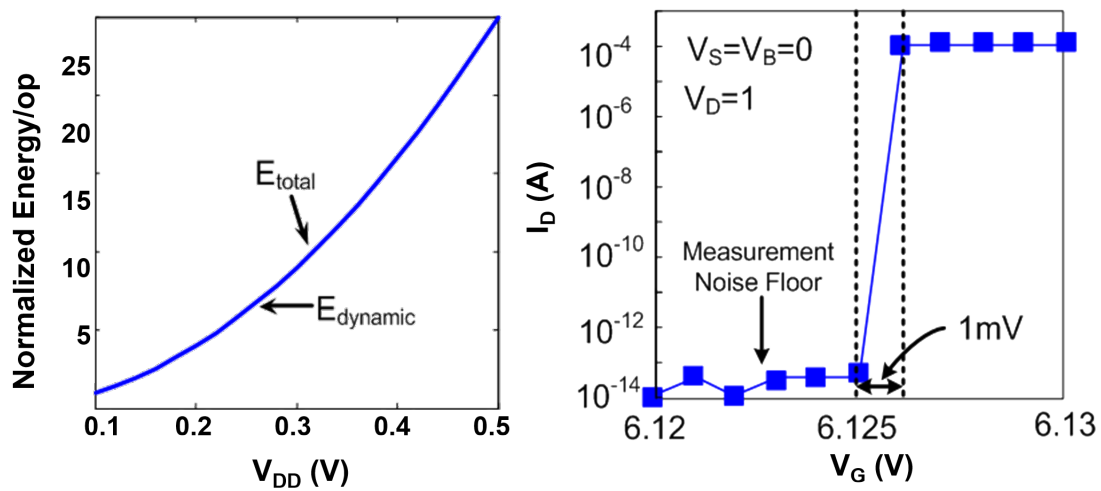


Figure 1.2. Ideal switching characteristics exhibited by nano-electromechanical relays. Left: energy per operation can continue to scale down for circuits made with NEM relays due to no leakage energy. Right: the current voltage characteristic of NEM relays. Data courtesy of R. Nathanael [4].

techniques can be carried over successfully to the largest computational blocks in a digital system. To answer this question, this thesis looks at the design of the floating-point unit (FPU) with NEM relays, as the most complex arithmetic structure in the heart of every modern processor.

Modern processors typically run arithmetic computations on an FPU. The FPU can be designed with NEM relays to realize energy-efficiency benefits offered by the technology. Although the NEM relays circuit design is already understood for many FPU components, several other components are designed in this work to optimize the use of NEM relay circuit area, energy, and performance. A complete analysis of the fused multiply-add (FMA) operation, representative of FPU logic, as implemented with NEM relays, is presented.

■ 1.1 Thesis Outline

This work begins with a background on NEM relay devices and the design of circuits with them in Chapter 2. Experiments on NEM relay test chips are discussed in Chapter 3. The FPU is described in Chapter 4, along with the motivation for representing the variety of FPU designs with the design of an FMA. The FMA, designed with NEM relays, is described in Chapter 5. The energy and performance analysis of the FMA and comparison with CMOS is presented in Chapter 6. The work concludes with Chapter 7.

NEM Relay Devices and Circuit Design

THERE are several flavors of NEM relays with which we can design circuits. Although some of the first computers were built with relays, vacuum tube and transistor technology proved more efficient in that era [6]. The recent revival of the relay stems from advances in micro-electromechanical systems (MEMS) that have enabled the fabrication of miniaturized relays with precise features whose size ranges from micrometers to nanometers [4, 7, 8, 9]. Current relays are called NEM relays because they have sub-100nm features, though they may still be referred to as micro-electromechanical (MEM) relays due to the remaining micron-scale features.

Each type of NEM relay has a physical model based on its electromechanics and a logical model based on switching characteristics. These models provide the basis for circuit design with NEM relays.

■ 2.1 Structure of NEM Relays

The different types of NEM relays shown in Table 2.1 are fabricated by our collaborators working under Professor Tsu-Jae King Liu at the University of California at Berkeley. The first MEM relay design considered had three terminals (3T) and was a single cantilever as described in [10], but the cantilever design gave way to a symmetric movable structure as shown in Table 2.1 since it suffers less from the impact of residual stress and strain gradient.

It was desirable to keep the actuation by the gate independent from the conduction enabled by actuation. Hence, a four-terminal (4T) NEM relay was designed where the gate to body voltage determines whether the device is actuated and the channel connects the source and drain terminals when the device is actuated [4]. The 4T NEM relay is pictured in Fig. 2.1.

Since the gate structure consumes the most area, another pair of drain and source electrodes could be added to the 4T NEM relay, yielding a six-terminal (6T) NEM relay [9]. The 6T NEM relay enables logic gates to be designed with less area since any two paths sharing the same gate to body control can be placed under the same gate structure.

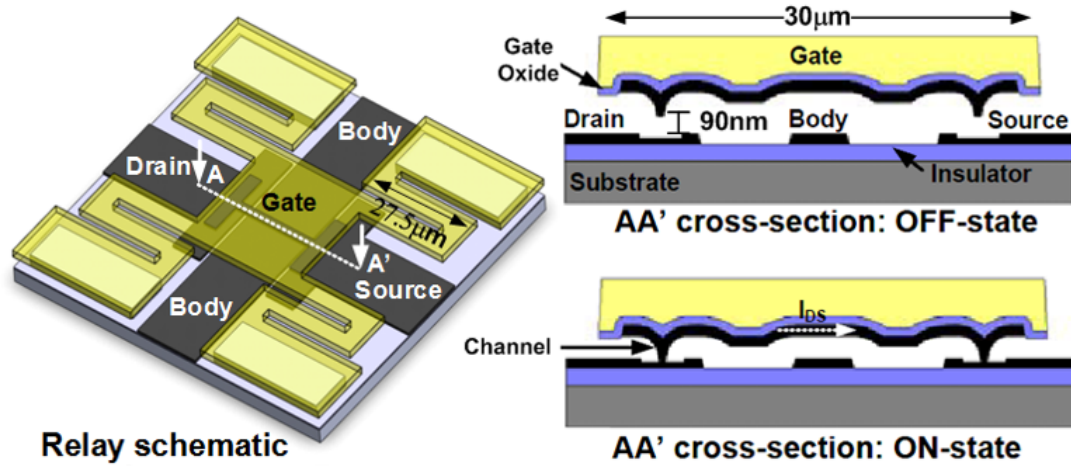


Figure 2.1. Structure of the four-terminal (4T) nano-electromechanical (NEM) relay. When the voltage between the gate and body is low, the device is in the *off*-state with no current; when it is high, the device is in the *on*-state with a current I_{DS} .

Device	3T	4T	6T	Seesaw
Original				
Enhanced				

Table 2.1. Evolution of NEM relay devices. The movable gate (G) structure is outlined in green, body (B) electrodes are filled in solid blue, drain (D) and source (S) electrodes are filled with a blue gradient, and the channels are in transparent red. The 3T, 4T, and 6T devices are also called crab-leg NEM relays, whose physics are different from seesaw NEM relays.

Details for the fabrication of crab-leg NEM relays are outlined in [4, 7].

Table 2.1 shows the evolution of the NEM relays with a crab-leg gate structure from 3T to 4T to 6T. Experiments with each of these types of NEM relays have enabled some learning as to how to enhance the device structure. Table 2.1 shows how the 4T NEM relay was enhanced by reducing the channel area and increasing the body electrode area to facilitate actuation, and how the 6T NEM relay was enhanced with an even larger actuation area and holes in the gate structure to facilitate the device's release in HF vapor.

Newton's second law is used to describe the motion of crab-leg NEM relays. The law takes the form of Equation 2.1, where the first term describes the mass, the second

W (μm)	h (μm)	g_d (nm)	g (nm)	A_d (μm^2)	A (μm^2)	V_{pi} (V)
5	1	100	200	8	810	10.5

Table 2.2. Dimensions and V_{pi} for a 4T NEM relay in a 5 μm process used for an illustration of device actuation in static analysis.

term describes the damping force, the third term describes the spring force F_{spring} , and the right side of the equation is the force of electrostatic attraction F_{elec} . m_{eff} is the mass of the movable structure, k_{eff} is the spring constant of the movable structure, Q is the quality factor, V is the gate to body voltage, g is the gap between the gate and body, and z is the displacement of the movable structure defined as 0 in the *off*-state.

$$m_{eff} \frac{d^2 z}{dt^2} + \frac{\sqrt{k_{eff} m_{eff}}}{Q} \frac{dz}{dt} + k_{eff} z = \frac{\epsilon A V^2}{2(g-z)^2} \quad (2.1)$$

The pull-in voltage is the voltage at which the NEM relay switches from *off* to *on*. This is determined through static analysis with derivations presented in [11].

$$V_{pi} = \sqrt{\frac{8k_{eff}g^3}{27\epsilon_0 A}} \quad (2.2)$$

Likewise, the pull-out voltage, also known as the release voltage, is determined as the voltage at which the NEM relay switches from *on* to *off*. It is lower than V_{pi} .

$$V_{po} = V_{rl} = \sqrt{\frac{2(k_{eff}g_d - F_A)(g - g_d)^2}{\epsilon_0 A}} \quad (2.3)$$

F_A is the surface adhesion force which also lowers V_{po} further. Since $V_{po} < V_{pi}$, there is a hysteresis gap in the transfer characteristic of a NEM relay. This gap is illustrated in Chapter 3 and can be used to an advantage because it can be used to avoid crowbar current by getting a device with the opposite body voltage to turn off before a device connected to the same output turns on [12].

A 4T NEM relay with specific dimensions in Table 2.2 is used for a static analysis to determine the behavior of its pull-in and pull-out. The device dimensions match a device presented in [13].

The static analysis shown in Fig. 2.2 involves setting the velocity and acceleration of the movable gate to zero to observe the remaining F_{spring} that F_{elec} must overcome for pull-in and also the F_{elec} that F_{spring} must overcome for pull-out. The analysis shows that pull-in occurs at $V = 10.5$ V indeed and that V must be reduced to 9.65 V for pull-out.

The time needed for a single actuation is called a single mechanical delay, t_{mech} [13]. t_{mech} is also derived from Equation 2.1, and is given by the operating voltage V_{DD} and material parameters. The parameters α , β , and γ are set by Q .

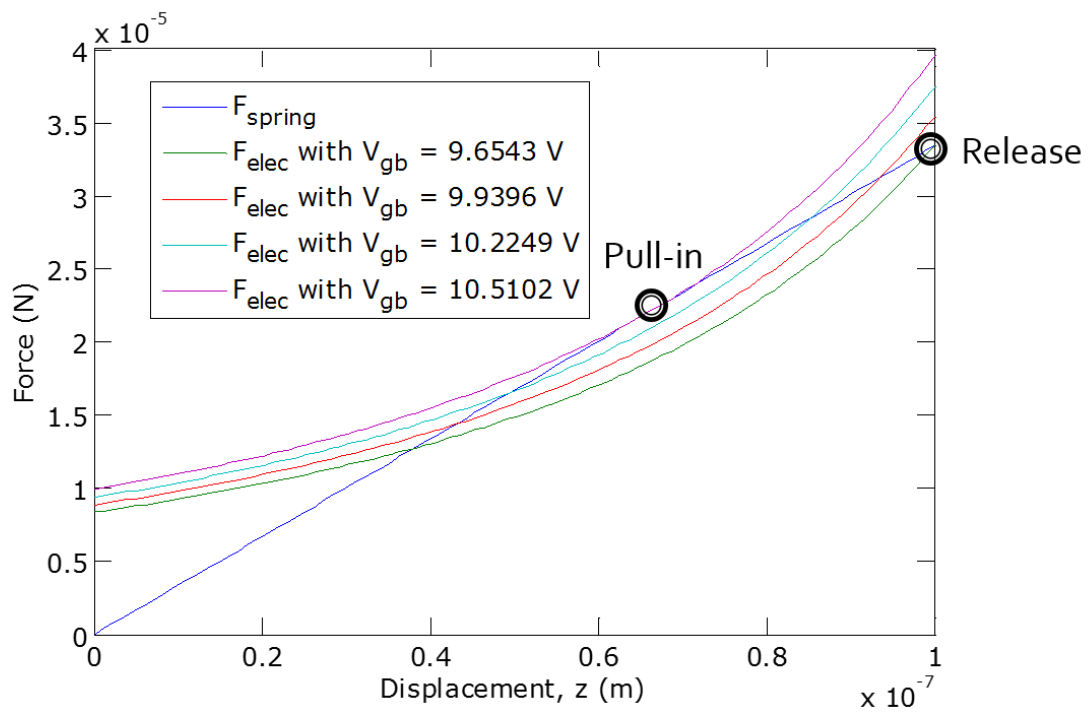


Figure 2.2. Static analysis of the 4T NEM relay whose dimensions are in Table 2.2. The electrostatic force F_{elec} is plotted for different voltages V against the spring force F_{spring} for all displacements of the movable gate, where $z = 0$ is the *off*-state. This analysis is with zero surface adhesion force.

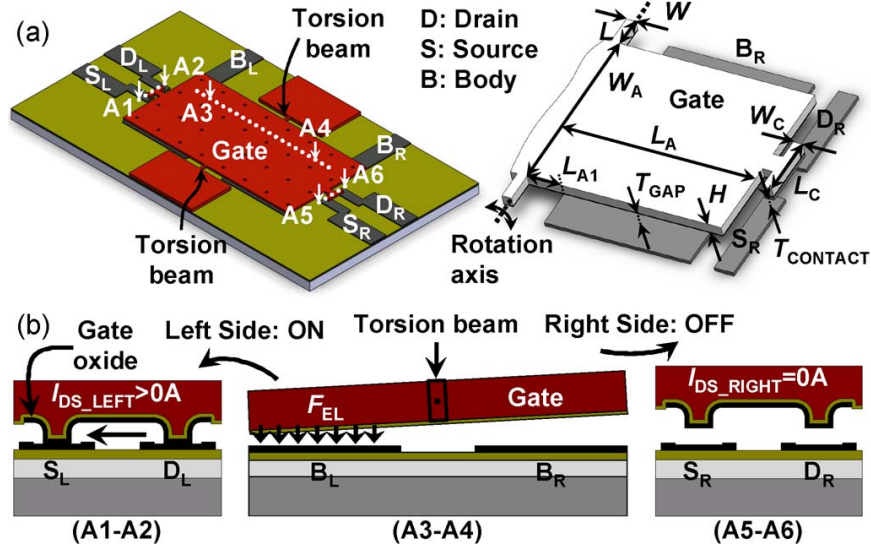


Figure 2.3. Seesaw NEM relay [8].

$$t_{mech} \cong \alpha \sqrt{\frac{m_{eff}}{k_{eff}}} \left(\frac{gd}{g_0}\right)^\gamma \left(\frac{V_{DD}}{V_{pi}} - \chi\right)^{-\beta} \quad (2.4)$$

Typically t_{mech} is longer than the electrical delay for signal propagation along a wire, t_{elec} . This factor matters for the purpose of circuit design.

Additional device model details for the 4T NEM relay, including the expression for k_{eff} and derivation of t_{mech} , are given for an actual process and predictive scaled process in [13]. The reliability of the contacts is also projected to improve with scaling [14].

In addition to crab-leg NEM relays, another type of NEM relay is the seesaw. Fig. 2.3 illustrates the seesaw NEM relay which is demonstrated to work in [8]. It features two sets of drain, source, and body terminals and a single gate terminal. The body terminals are biased at opposite voltages with one at V_{DD} and the other at V_{SS} which is typically ground. This way either the left side is actuated and the right side is deactivated, or vice versa. This enables perfectly complementary switching because up to one of the two pairs of channels is conducting at any given time [15, 16].

The switching energy for a generalized NEM relay is given in Equation 2.5, where C_{total} is the total amount of capacitance on an output.

$$E_{sw} = \frac{1}{2} C_{total} V_{DD}^2 \quad (2.5)$$

Device	4T	6T	Seesaw
Symbol			
Logic	$ V_{GB} > V_{pi} \rightarrow$ D connected to S <hr/> $ V_{GB} < V_{po} \rightarrow$ D not connected to S	$ V_{GB} > V_{pi} \rightarrow$ D_L connected to S_L D_R connected to S_R <hr/> $ V_{GB} < V_{po} \rightarrow$ D_L not connected to S_L D_R not connected to S_R	$ V_{GB_L} > V_{GB_R} $ and $ V_{GB_L} > V_{pi} \rightarrow$ D_L connected to S_L D_R not connected to S_R <hr/> $ V_{GB_R} > V_{GB_L} $ and $ V_{GB_R} > V_{pi} \rightarrow$ D_L not connected to S_L D_R connected to S_R

Table 2.3. Circuit symbols and logical descriptions for different flavors of NEM relays.

■ 2.2 Logical Description of NEM Relays

Each type of NEM relay can be described more simply in terms of their *on* state and *off* state [17]. The circuit symbol and logical description for NEM relays used in circuits are shown in Table 2.3.

The logical descriptions in Table 2.3 show that 6T and seesaw NEM relays offer twice the logical capability for the same area as a 4T NEM relay. Also, seesaw NEM relays are ideal for perfectly complementary logic, because the two source terminals can be connected as one output and then the gate will select either the left drain or right drain terminal to pass through to the output.

■ 2.3 Circuit Design with NEM Relays

NEM relays can be used to build digital logic as with CMOS, to the extent where NEM relays can be used as drop-in replacements for transistors when biased appropriately. The nMOS and pMOS equivalents for NEM relays are shown in Fig. 2.4.

The inverter and buffer are made using the appropriate drop-in replacements for the CMOS inverter. Since NEM relays can pull the output down to V_{SS} just as well as up to V_{DD} , both inverting and noninverting logic are possible with NEM relays, and the CMOS inverter design can be used for a buffer as well. The inverter and buffer are designed with 4T, 6T, and seesaw NEM relays in Fig. 2.5. The designs also show how half the number of devices is required for circuits with 6T and seesaw NEM relays compared to circuits with 4T NEM relays.

Although CMOS-style designs can be directly translated into NEM relay circuits,

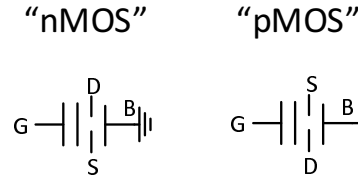


Figure 2.4. Equivalent nMOS and pMOS drop-in replacements with appropriately biased NEM relays.

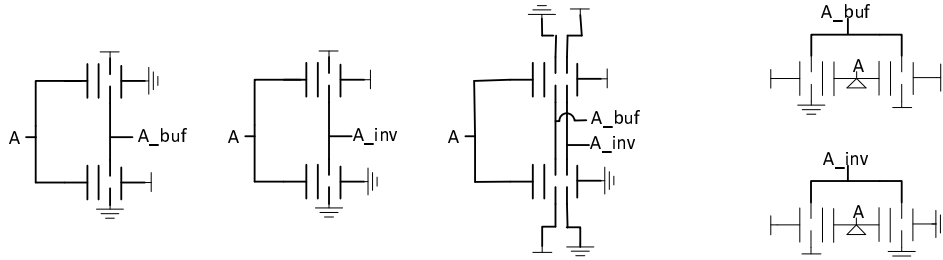


Figure 2.5. Inverter and buffer designed with NEM relays. The input is A, the inverter output is A_inv, and the buffer output is A_buf. 4T, 6T, and seesaw NEM relay designs are shown.

it is wiser to design NEM relay circuits differently due to one critical disparity between CMOS and NEM relays. That is, NEM relays have a mechanical delay t_{mech} that is significantly longer than the electrical delay t_{elec} . This means that NEM relay circuits are better designed when all actuations take place simultaneously. Although CMOS circuits are designed to have few transistors in series due to the quadratic increase in Elmore delay per transistor in series, NEM relay circuits can tolerate long stacks of relays in series because the total delay is still limited by the mechanical delay. The phenomenon where t_{elec} is increased by increasing the NEM relay stack length until $t_{elec} \geq t_{mech}$ is illustrated in Fig. 2.6 [5].

In essence, NEM relays circuits are best designed with pass transistor techniques which stack together many NEM relays in series. This means that NEM relay gates are connected to circuit inputs only if the circuit is to operate within a single mechanical delay, t_{mech} . In cases where a pass-transistor design requires far too much area compared to another design for the same circuit, it may be better to choose the other design and suffer multiple t_{mech} delays. However, more often than not, the number of NEM relays required is lower than the number of transistors required for the equivalent CMOS circuit. This is illustrated in the circuit in Fig. 2.7 which is built with CMOS and NEM relays [1].

The fan-in and fan-out of logic gates is less important with NEM relays than it was with CMOS. This is because the latency of a NEM relay depends much less critically on fan-in and fan-out due to the dominance of mechanical delay. While CMOS circuits limit the fan-in and fan-out of a circuit to four typically and introduce additional gate

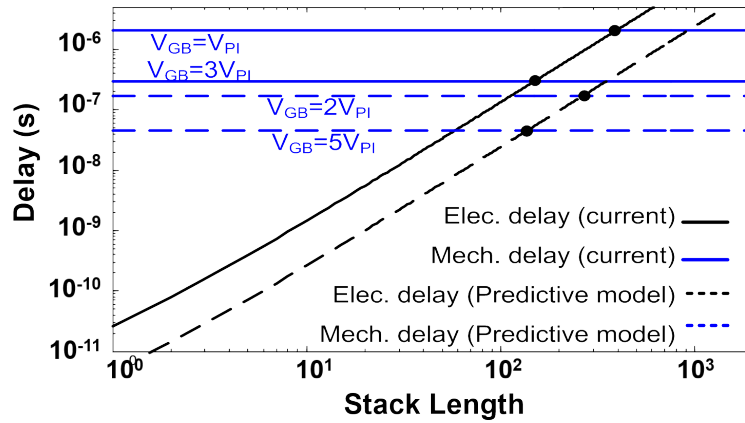


Figure 2.6. Electrical vs. mechanical delay for a long stack of NEM relays in current ($1 \mu\text{m}$) and scaled (90 nm) NEM relay process technology, courtesy of H. Fariborzi [5].

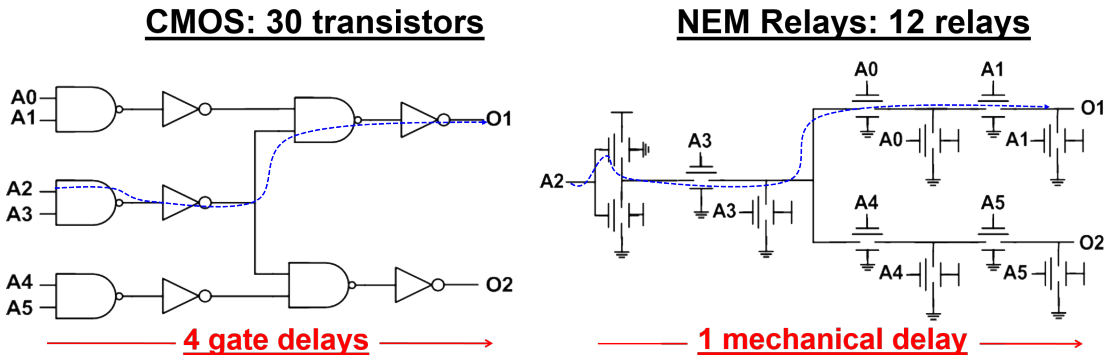


Figure 2.7. The same circuit is built with CMOS and NEM relays. Less NEM relays are required than transistors. Schematics are courtesy of F. Chen [1].

stages if there are more inputs, NEM relays circuits do not have such a limitation. This makes it possible to use fewer NEM relays with a greater proportion of inputs on gates or bodies compared to the number of transistors of the equivalent circuit in CMOS.

The feasibility of putting the output of one or more NEM relays on the gate of the next stage of NEM relays in a circuit is called composability. The composability of NEM relays circuits is experimentally verified in Fig. 2.8, where a NEM relay NOT gate outputs with full swing to be a valid input to another NEM relay circuit [2]. The hysteresis gap $V_{pi} - V_{po}$ also sets the minimum swing needed on V_{DS} to have NEM relay circuit composability. The swing on V_{GB} can be much larger in magnitude because the gate and body do not carry current, unlike the drain and source.

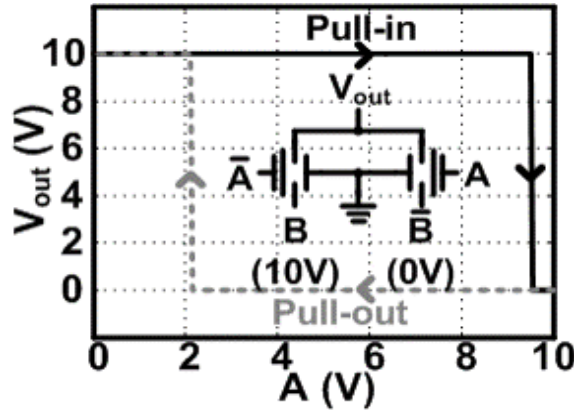


Figure 2.8. Schematic and measured voltage transfer curve of a NEM relay NOT or XOR gate showing hysteresis and full rail swing at the output. Data from [2].

■ 2.4 Process Technology and Scaling

The process technologies considered in this work are described with model parameters tabulated in [11]. This includes the 2nd generation 4T NEM relay process, scaled 6T NEM relay process, predictive 90nm equivalent 4T NEM relay process, and predictive 90nm equivalent 6T NEM relay process. The model for scaling crab-leg NEM relay devices is described in [13].

The need to develop circuit design infrastructure and novel NEM relay circuit designs described in [18] is addressed with solutions described in Sections 5.3 and 6.3.

As the process technology advances, the same circuit designs of VLSI systems may be implemented using the new process nodes. This warrants the exploration of floating-point unit (FPU) design with NEM relays presented in this work.

Experiments on NEM Relay Chips

THE experimental demonstration of NEM relay circuit designs is carried out by a collaborative team that handles the design, fabrication, and testing of NEM relays chips. The team has students and researchers under faculty at MIT (Prof. V. Stojanović), the University of California at Los Angeles (Prof. D. Marković), and the University of California at Berkeley (Prof. T.-J. King Liu and Prof. E. Alon). The team developed the complete computer-aided design (CAD) infrastructure described in Sections 5.3 and 6.3 to produce the layout of test chips that are then fabricated. The NEM relays process technology is CMOS-compatible, because the NEM relays are fabricated in a back-end-of-line (BEOL), low-temperature process.

The first test chip, CLICKR1, helped the team determine better device dimensions as well as other improvements for the next test chip, CLICKR2. CLICKR2 circuits were built using the enhanced 4T NEM relay device shown in Table 2.1, and experiments on the test chip made way for some key circuit demonstrations. Each successive test chip featured a new generation of NEM relay device used for circuits. Table 3.1 shows the scaling of the device cell for each test chip, including the overall device dimensions of

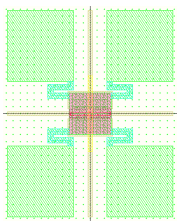
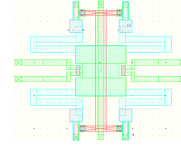
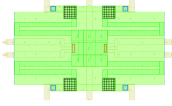
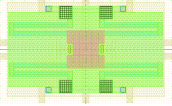
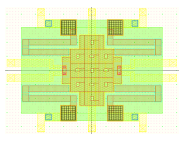
CLICKR2	CLICKR3	CLICKR4	CLICKR5	CLICKR6
				
Device: 120 μm x 150 μm Gate overlap: 30 μm x 30 μm	Device: 20 μm x 20 μm Gate overlap: 7 μm x 7 μm	Device: 63 μm x 36 μm Gate overlap: 15 μm x 15 μm	Device: 64 μm x 36 μm Gate overlap: 15 μm x 15 μm	Device: 50 μm x 36 μm Gate overlap: 16 μm x 25 μm

Table 3.1. Device cell used for circuits on each test chip. Device images are taken from CAD environment and are not to scale relative to each other. CLICKR3 is fabricated at Sematech and all others are fabricated at UC Berkeley.

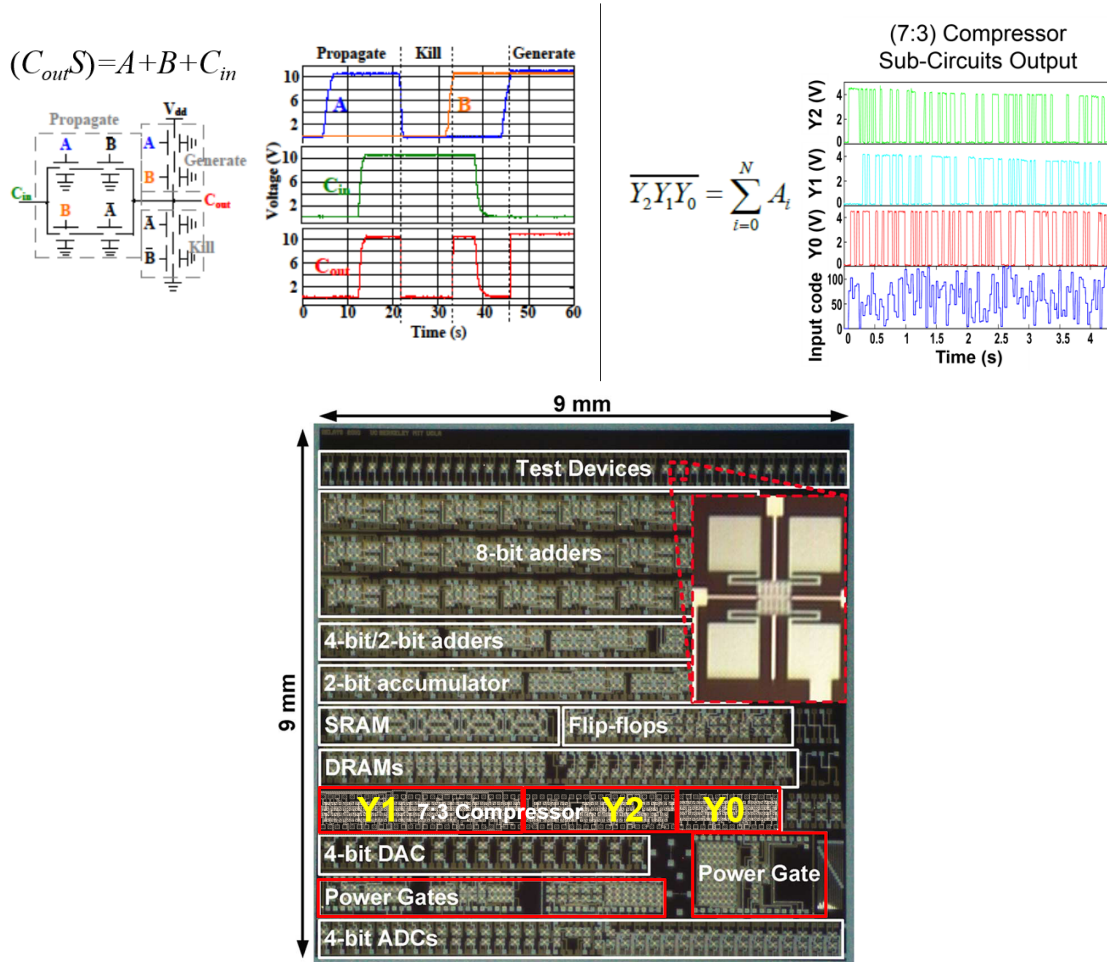


Figure 3.1. On the top left is the full adder output courtesy of F. Chen, on the top right is the (7:3) compressor output courtesy of H. Fariborzi, and on the bottom is a micrograph of the CLICKR2 test chip containing the circuits.

the movable structure plus anchors and the approximate dimensions of the actuation overlap area under the gate.

The full adder and (7:3) compressor for a multiplier were demonstrated on CLICKR2 and are reported in [2, 5]. The demonstration results are reproduced in Fig. 3.1. Other circuits from this test chip, including a latch and DRAM are reported as well [19].

■ 3.1 Scaled NEM Relays

The energy-efficiency benefits of NEM relay circuits are realized by scaling down the device dimensions [13]. The scaling theory in [13] applies to the test chips that have been successively scaled in each generation.

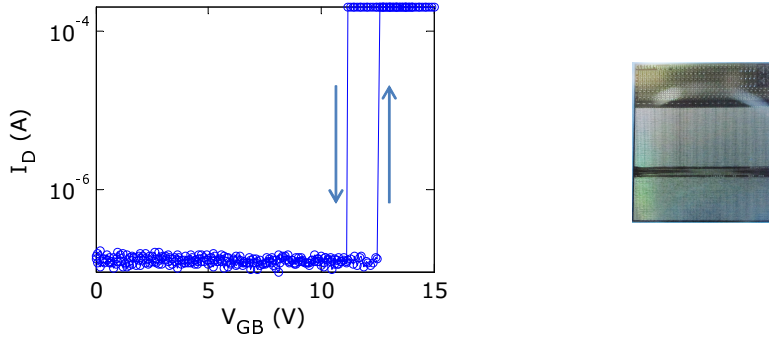
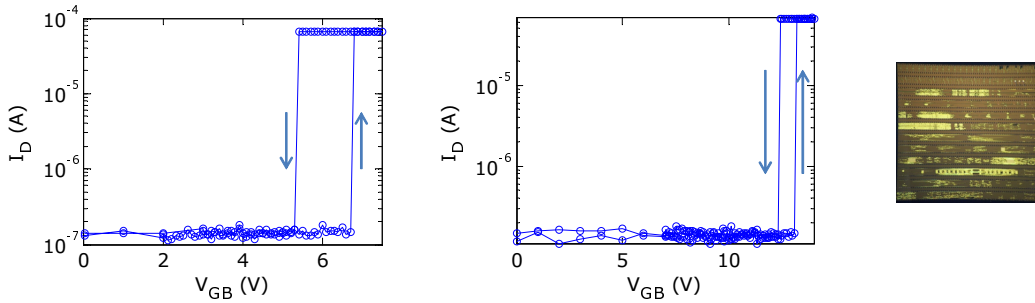


Figure 3.2. From left to right are a high V_{pi} NEM relay and the CLICKR4 test chip containing those devices.



	L	V_{pi}	V_{po}	$R_{external}$	V_{DD}	$R_{DS,relay}$
Low V_{pi} device	$5.952 \mu\text{m}$	7 V	5 V	50 k Ω	4 V	7 k Ω
High V_{pi} device	$1.96 \mu\text{m}$	13.1 V	12.4 V	50 k Ω	4 V	7 k Ω

Figure 3.3. From left to right: the transfer characteristic for a low V_{pi} NEM relay with long L , that for a high V_{pi} NEM relay with short L , and a micrograph of the CLICKR3 test chip containing those devices. Measured parameters for those devices are tabulated.

One way to check the energy-efficiency benefits of scaled NEM relays is to compare their measured V_{pi} with that of larger generation NEM relays. CLICKR4 has larger device dimensions and a measured V_{pi} is 12.5 V shown in Fig. 3.2. CLICKR3, which was fabricated after CLICKR4, has a scaled down NEM relay and its measured V_{pi} is 7 V shown in Fig. 3.3. This is among the devices with the lowest measured pull-in voltages.

The design of the low V_{pi} device in Fig. 3.3 is the design for the device used for all circuits on CLICKR3. The high V_{pi} device in Fig. 3.3 is a unique device with short L for experimenting purposes. It is better to fabricate long L devices because, as Equation 2.2 shows, $V_{pi} \propto L^{-\frac{3}{2}}$ and is very sensitive to L if it is short, and as such the yield is indeed better for long L devices. Since the long L devices are used in the circuits, circuits such as the (7:3) compressor were working in the scaled process [11].

■ 3.2 Rudimentary Circuits and Tests

Circuits are tested by first characterizing individual devices in the circuit, as permitted by the available pads, and then running gradually more complex circuit tests until performing the test of the full circuit.

External resistors are needed especially on the nets connected to NEM relay drains and sources, because they limit the current and prevent breaking by undesired heating (contact welding) or oxide breakdown. External resistors may also be put on the nets connected to gates and bodies to check if there is any parasitic leakage into the gate or body nets, which ideally have no current but may be shorted to other nets due to process issues. The external resistor values should be chosen based on the maximum current permitted in the channel.

The older testing board had crossbar switches that let the user write a script on a computer to connect each pin either to a supply voltage, to ground, or to another pin. This allowed one to build circuits with two to five relays using the padded out individual devices. However, since the crossbar switches on the testing board failed often, a new testing board was developed that cannot connect pins to each other but could drive or read a larger number of circuit pins without failing. The testing boards have spots to place surface mount external resistors, though it is often better to choose axial-lead resistors that would limit the current to an amount appropriate for the process technology, and place those resistors in series with the power supplies.

While device dimensions were scaled down between CLICKR2 and CLICKR4, the circuit complexity also grew. CLICKR4 is a test chip with a Picoblaze clone microcontroller circuit built with NEM relays. With the scaled down device size, most parts of the microcontroller fit on the chip, except the instruction memory.

CLICKR4 also had individual devices for device and rudimentary circuits testing. The transfer characteristics were measured and verified on these devices, and there was some variation in V_{pi} from device to device. The NEM relays on CLICKR4 had a tungsten (W) channel as in earlier test chips. It was discovered that the native oxide that formed at the contacts was relatively thick enough to require a breakdown voltage to permit conduction through the channel. The measured oxide breakdown V_{DS} for W NEM relays was typically 6 V.

W NEM relays could be oxide broken usually when there were up to two in series. If there were more in series it would require a higher voltage that may destroy devices instead. The microcontroller design was changed to provide access to internal nets via oxide breaker devices that could be used to oxide break all NEM relays essential to the microcontroller functionality. This design was spun on CLICKR5.

In addition to wafers with W NEM relays, additional wafers were built with ruthenium (Ru) NEM relays. Ru was chosen because its oxide RuO_2 is conductive. As it turned out, Ru NEM relays did not need to be oxide broken, perhaps relieving the need for oxide breaker devices. However, Ru NEM relays were more likely to break at currents above 100 μA , requiring low V_{DS} . CLICKR6 is a re-spin of CLICKR5 with additional rudimentary circuits to test for better device characterization. The microcontroller on

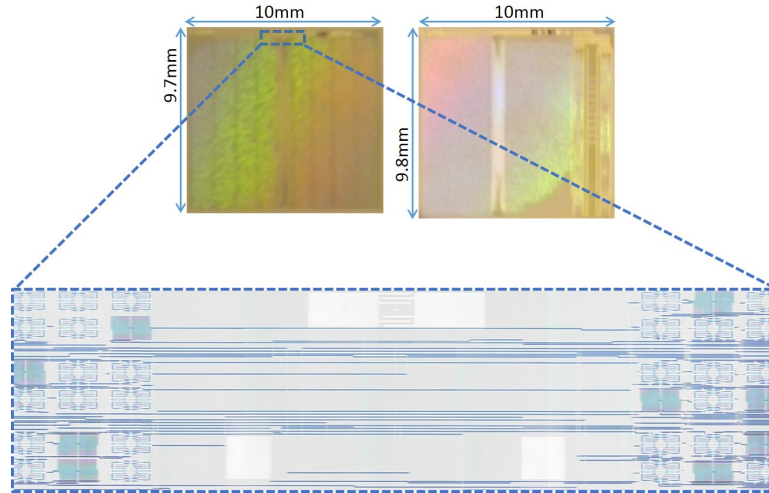


Figure 3.4. Microcontroller test chips, CLICKR5 (left) and CLICKR6 (right).

CLICKR5 and CLICKR6 are shown in Fig. 3.4.

An example of the oxide breaker devices introduced in CLICKR5 is shown on the top of Fig. 3.5, where breaker devices provide access to oxide break all devices in a flip-flop. CLICKR6 also has a corresponding padded-out flip-flop circuit shown on the bottom of Fig. 3.5, and this circuit was tested. A buffer in the flip-flop master stage is verified to work partially, in that it pulls down V_{DDSD} to the node MID when CLK2 is high and D is pulsed. The pull up path did not work because the device with V_{DD} on the body did not actuate. No oxide break V_{DS} was necessary because the Ru NEM relays do not require oxide breaking.

It makes sense to turn on the feedback device last, meaning it should be kept deactuated until other parts of the circuit work.

Tests on more recent NEM relays chips made with ruthenium show that oxide breaking can be avoided. External resistors of $100\text{ k}\Omega$ limit currents to within $100\text{ }\mu\text{A}$.

Although only small-scale circuits were demonstrated on the later test chips featuring much larger scale circuits, there is still much hope for future generations of process technology scaling and circuit integration. The Ru NEM relays do not require oxide breaking and operate at $V_{DS} \leq 3\text{ V}$ from the start, and have been observed to actuate multiple times and have reasonable yield. The experiments described here helped uncover process technology improvements, which are all the more necessary to realize larger scale systems. Experience has proven many examples of working combinational logic in NEM relays, motivating the design of the most complex arithmetic unit in a computational system, the floating-point unit, with NEM relays.

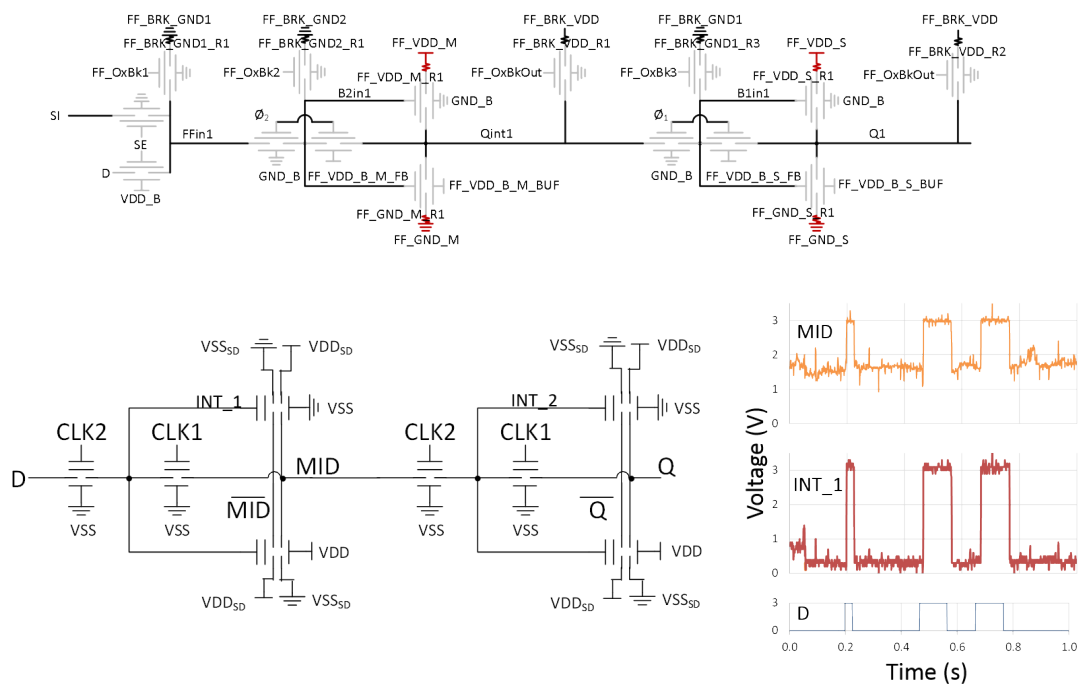


Figure 3.5. Flip-flop with oxide breaker devices on CLICKR5, courtesy of C. C. Wang (top). Partial buffer test on a CLICKR6 flip-flop (bottom). In the CLICKR6 buffer test, CLK2 is high and the pulse on D is followed at the output MID.

Floating-Point Numbers and the Floating-Point Unit (FPU)

COMPUTERS represent numbers by assigning a meaning to the bits in a binary word. Integers are often represented as two's complement or sign-magnitude numbers. When the length of a word cannot accommodate the range of real numbers that it needs to represent, certain bits in the word can be used instead for an exponent that defines the location of the binary point in a sign-magnitude number, similar to how in scientific notation the exponent defines the location of the decimal point. Dedicating bits for the exponent thereby increases the range of real numbers that a word can represent.

A floating-point number is a representation of a real number in which the binary point is not fixed [20]. Computations with floating-point numbers are done with a floating-point unit (FPU).

■ 4.1 Floating-Point Numbers

IEEE-754 is the standard for representing floating-point numbers with binary words. The value represented by a floating-point number binary word containing the sign, exponent, and fraction is $(-1)^{\text{sign}} \times (1 + \text{fraction}) \times 2^{\text{exponent} - \text{bias}}$. The specific assignment of the bits in single-precision (32-bit) and double-precision (64-bit) IEEE-754 floating-point numbers is listed in Table 4.1 [20]. The exponent is biased, meaning that an offset called the bias is subtracted from the positive exponent. The bias is also listed in Table 4.1. The significand, $1 + \text{fraction}$, adds 1 to the fraction because the significand is designed to begin with a 1. That is, the significand has a 1 in the implied bit position just above the most significant bit of the fraction [20].

There are exceptions to this representation [20]. In the case of representing the number 0 as a floating-point number, the exponent and fraction are both all 0s. Infinity and not-a-number (NaN) are defined with an exponent that is all 1s, and with a fraction of all 0s for infinity or otherwise for NaN. Another exception is when the exponent is all 0s but the fraction is nonzero, in which case the number is considered denormalized and the significand has no implied 1 above the most significant bit of the fraction, in order to permit the representation of numbers smaller than the smallest normalized

Binary word type	Bits	bias	sign	exponent	fraction
Single-precision floating-pt.	32	127	Bit 31	Bits 30 to 23	Bits 22 to 0
Double-precision floating-pt.	64	1023	Bit 63	Bits 62 to 52	Bits 51 to 0

Table 4.1. IEEE-754 standard for single-precision and double-precision floating-point numbers.

number, $(1 \times 2^{0-\text{bias}})$. When these exceptions do not apply, floating-point numbers are normalized because of the implied 1 in the bit position above the most significant bit of the fraction.

The improved ability to represent real numbers with floating-point numbers leads to the design of the advanced hardware required to process numbers represented as such.

■ 4.2 FPU Design

The floating-point unit (FPU) is a processor that performs computations on floating-point numbers. The operations typically supported on floating-point numbers are addition, subtraction, multiplication, and division, though additional operations may be available.

While FPU designs vary greatly, they share many common elements. The operands and results of floating-point calculations are stored in registers, whose data may be loaded from or stored to a separate memory. An FPU control unit is needed to provide the correctly timed control of the registers depending on when they are being read or written to. The FPU control unit also enables the arithmetic operation blocks based on the operation to be performed on the operands. The FPU control unit may have an internal instruction memory and data memory, or its instructions and data may be hard-wired or taken from outside the FPU. Fig. 4.1 shows this generic design of an FPU.

FPU designs are often pipelined to use all available blocks at the same time and increase throughput. However, the dominance of the mechanical delay in NEM relay circuit latency makes it more important to focus on single-pipeline-stage FPUs for design with NEM relays.

While modern computers use both single-precision and double-precision FPUs, the important design trade-offs can be revealed by simply focusing on the single-precision FPU for the purpose of design with NEM relays.

Common operations may be combined together to increase overall throughput. Fig. 4.1 shows a generic FPU and Fig. 4.2 shows an FPU with a combined multiply and add.

FPU designs with CMOS have evolved based on separating pipeline stages based on electrical delay of different paths that run in parallel, and this has enabled increases in throughput [21]. FPU designs with NEM relays also demand a look at what can be run in parallel to reach results faster, perhaps even in a pipelined fashion.

A closer look at the types of combinational logic used for all types of FPU operations

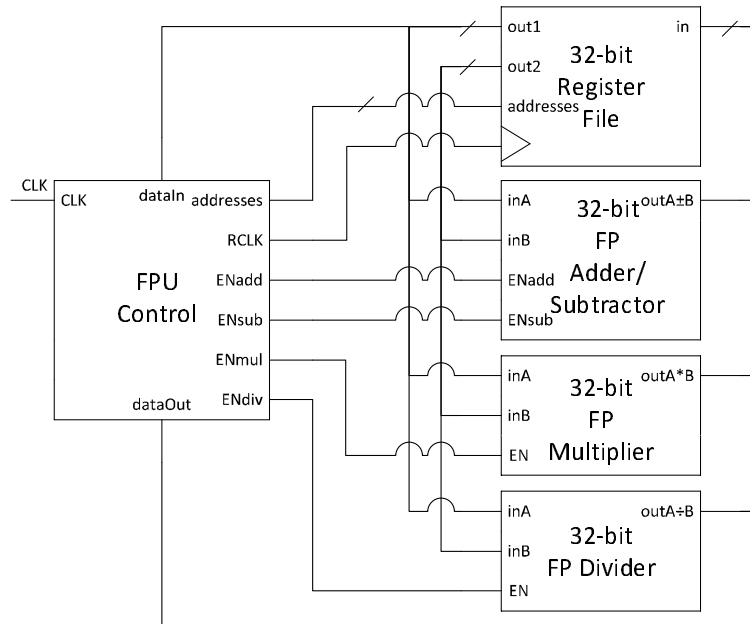


Figure 4.1. Generic single-precision FPU design.

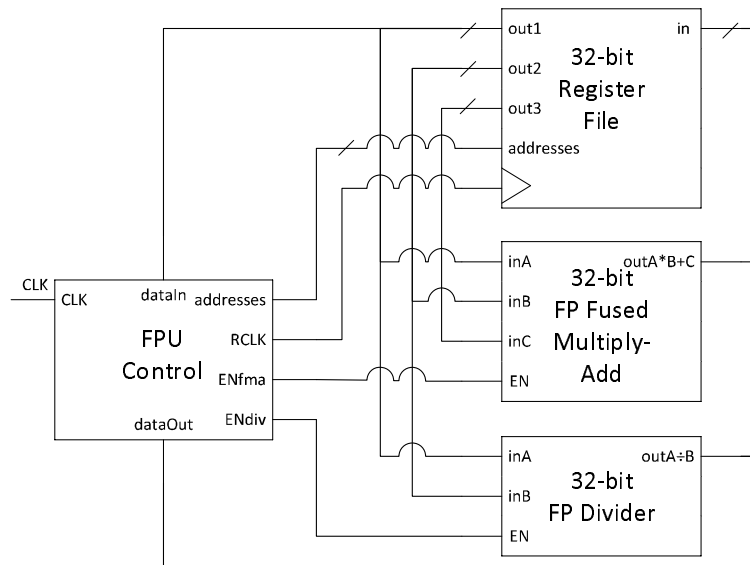


Figure 4.2. Single-precision FPU design with a fused multiply-add.

is needed to see what NEM relay logic could be run in parallel. First, the floating-point add instruction logic contains logic for: adding, subtracting, multiplexing, shifting, and rounding binary integers, and, critically, normalizing floating-point numbers [22]. The normalization of floating-point numbers can be done slowly in iterations as in [20] but is generally best done in a single iteration with a leading-zero detector or anticipator [22]. Rounding requires an incrementer, which is a block that adds one plus the input. Second, the floating-point multiply instruction has the same types of logic as for the floating-point add instruction but also has a binary number multiplier. The multiplier may be combinational, or in the case of a pipelined FPU it is often better to have a pipelined multiplier. Third, the floating-point division instruction is considerably more complex but includes the types of logic of the floating-point multiply instruction as well as heavy use of adders, lookup tables (LUTs), and multiplexers [23]. Thus, there is much in common between the types of combinational blocks used for each floating-point operation.

Although a divider is not designed with NEM relays in this work, the blocks needed to build a NEM relays divider are available. One possible implementation would be SRT division, named after Sweeney, Robertson, and Tocher [23]. Such an implementation could encode the binary decision diagram with multiplexers. The look-up tables (LUTs) can be translated into multiplexer trees and optimized as described in Chapter 5.

Some of the combinational blocks needed for floating-point operations have already been designed with NEM relays, while others are designed in this work. Among the major NEM relay circuits that have been designed and demonstrated are the adder, subtractor, multiplier, and various logic gates [2]. From this, it is observed that shifters, leading-zero detectors, and leading-zero anticipators are the new blocks that are designed in this work to realize an FPU designed with NEM relays.

■ 4.3 Representing FPU Design with the Fused Multiply-Add Architecture

Given the variety of FPU architectures available, it is best to select one FPU architecture to design with NEM relays that can be compared apples-to-apples to a CMOS implementation. This approach provides a well-defined, concrete implementation of an FPU with NEM relays, including the new blocks that must be designed to realize any FPU designed with NEM relays.

The ideal FPU architecture needs to represent the main types of operations performed with floating-point numbers. The multiply-accumulate (MAC) operation combines a multiplication and an addition into a single step:

$$out = A \times B + C \quad (4.1)$$

The MAC operation can be applied to floating-point numbers. In this form, it represents the most common floating-point operation and offers higher accuracy and performance because no rounding is done between separate multiply and add steps [24].

The fused multiply-add (FMA) is a multiply-accumulate design that has low latency

because it aligns the addend significand of C in parallel with multiplication of the significands of A and B [24]. After the addend is added to the multiplier result, the result is normalized and rounded. This means that the FMA has parallel operations from the start, and ends in a more serial normalization step. The FMA requires a large leading-zero detector or anticipator and a large shifter for normalization. Further FMA architectural details are outlined in [25].

The cascade multiply-add (CMA) performs the same operation as the FMA, but the CMA combines multiplier partial product terms before alignment. Compared to the FMA, the leading zero detector or anticipator and the shifter used for normalization in the CMA are smaller because the CMA chooses which input to shift beforehand. An analysis in CMOS shows that the energy per operation of an FMA and that of a CMA in the same process are essentially the same, showing that various FPU circuit designs for the same operation have only small differences in energy consumption [24].

The FMA and CMA are the most common architectures for the MAC operation, and an analysis of the most energy-efficient designs for them in CMOS is available [24]. The FMA has been popular ever since it was used in an IBM PowerPC microprocessor [26]. The highly parallel architecture of the FMA, the common use of the FMA in FPUs, and the availability of a detailed energy analysis of the FMA designed with CMOS provide motivation to design the FMA with NEM relays to represent the design of an FPU with NEM relays.

The Fused Multiply-Add (FMA) Designed with NEM Relays

HAVING chosen the fused multiply-add (FMA) architecture to represent the design of the FPU with NEM relays, what follows is the design of the circuits to compose the architecture. This chapter identifies how to optimize the design of NEM relay circuit blocks to obtain high throughput while balancing the trade-offs between circuit area, power consumption, and speed. In the process, several novel component circuit designs are introduced.

■ 5.1 Shortcomings of CMOS-Style Circuit Design

The typical choice of CMOS as the process technology for a circuit has driven design decisions for the circuit as well. However, the CMOS design paradigm does not work well for circuits with NEM relays. The pass-transistor circuit design paradigm for NEM relays, described in Section 2.3, should be used to optimize circuits in the FMA.

Among the key drawbacks of applying the CMOS circuit design paradigm to designing NEM relay circuits in the FMA are that the circuit delay and circuit area become excessively large. The longer circuit delay stems from the many gate stages in CMOS-style circuit designs, where each stage introduces a mechanical delay t_{mech} into the critical path from input to output. The larger circuit area stems from having to use more NEM relays than are needed due to CMOS limitations on the number of inputs to a device and to a single gate stage of a circuit. The greater area footprint and delay time of such circuits also leads to higher energy consumption, adding yet another argument for applying the NEM relay circuit design paradigm to vastly reduce the circuit delay, area, and energy per operation of the FMA designed with NEM relays.

To illustrate how crucial it is to follow the NEM relays design paradigm to optimize the circuit design of the FMA, Fig. 5.1 shows the total number of mechanical delays in the critical path of the FMA for when: (1) the design is translated directly from optimal CMOS to NEM relays, (2) the design in (1) also has adders optimized for NEM relays, (3) the design in (2) also has decoders and multiplexers optimized for NEM relays, (4) the design in (3) also has the multiplier optimized for NEM relays, and (5) the FMA design has all optimizations for NEM relays reported in this work.

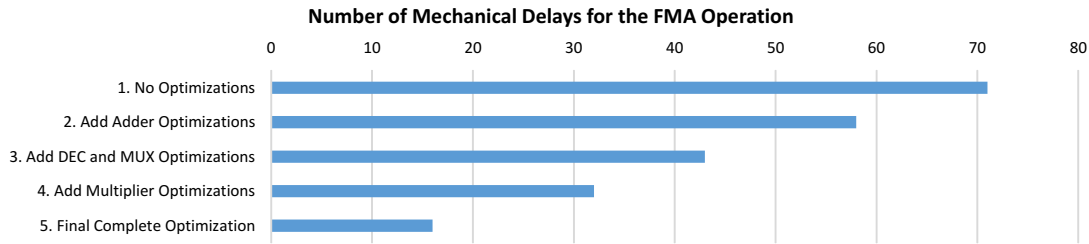


Figure 5.1. Reducing the number of mechanical delays for the fused multiply-add (FMA). Adder (ADD) and multiplier (MUL) optimizations are recommended by [2, 5]. Additional decoder (DEC) and multiplexer (MUX) optimizations are the beginning of further optimizations introduced in this work.

The optimal CMOS circuits that are translated to NEM relays without optimization have a common maximum fan-in of four per gate stage, which is increased when optimizing for NEM relays. Adders are implemented as Sklansky adders in optimal CMOS, but they are better implemented as Manchester carry chain adders to be optimal for NEM relays [1]. Multiplexers designed for CMOS have three gate stages, increasing by one every time the number of inputs increases by a power of four due to the limited fan-in of four in CMOS [27].

Circuits optimized for NEM relays are described in earlier literature. Adders and multipliers have already been optimized for NEM relays in [2] and [5], respectively, and these works outline the importance of applying the pass-transistor design paradigm to optimizing other circuits. The optimized designs of other circuits are fully described in this work, including decoders, special cases of multiplexers, and, especially, the leading zero detector (LZD) and leading zero anticipator (LZA) blocks. The optimal design of single- t_{mech} NEM relay logic gates and adders are shown in Fig. 5.2.

■ 5.2 Implementing the FMA Architecture in Hardware

Even with the architectural description of the FMA available in [24], it must be brought to a behavioral hardware description for a hardware implementation. The behavioral description would then go through a standard synthesis, place, and route flow to be turned into a chip layout that can be fabricated. Fortunately, the team that performed the energy analysis on the FMA also provided the tool used to turn the system-level description into a behavioral hardware description [24].

The tool, FPGen, is an FPU generator built to explore the design space of FPU architectures, providing insights on the energy efficiency and throughput for a number of parameterizable designs [28]. This FPU generator is based on the Genesis2 framework, a tool that enables system designers to generate the hardware description for a set of given parameters and desired trade-off optimizations [29].

The behavioral Verilog of the FMA shown in this work was generated by FPGen, with appropriate parameters set to have a pipeline depth of 0 for a combinational

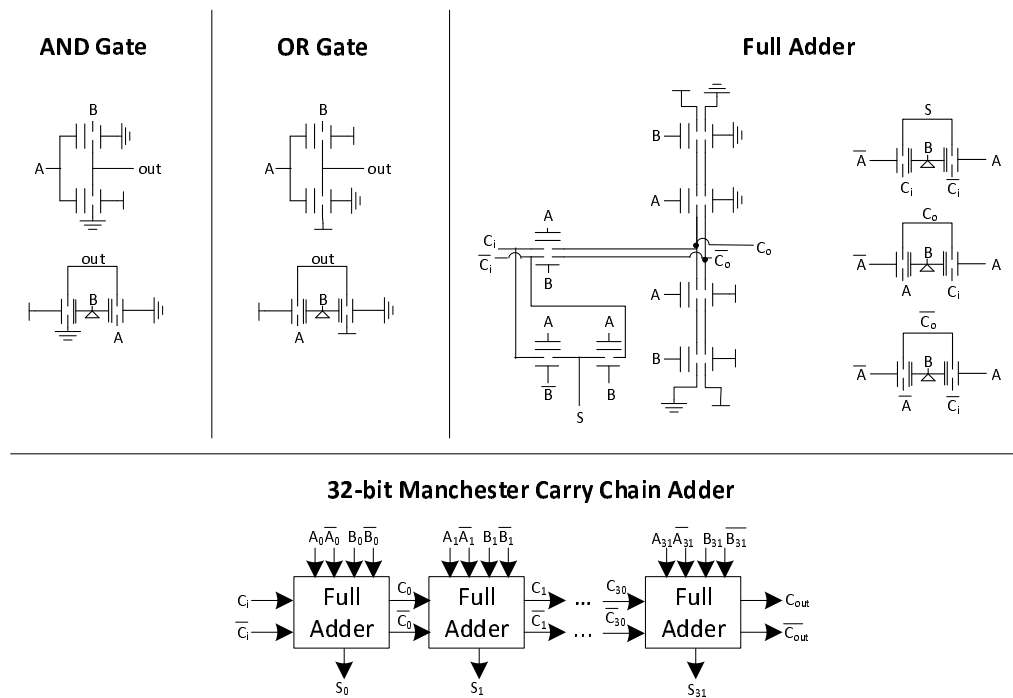


Figure 5.2. AND gate, OR gate, full adder, and 32-bit adder designed with 6T and seesaw NEM relays. All circuits produce their result within a single mechanical delay.

FMA and bitwidths set to that of the IEEE-754 single-precision floating-point number standard. The justification for these parameter choices is in Section 4.2. The FMA architecture, composed of circuits designed with CMOS, is shown in Fig. 5.3.

The generated design was verified in a behavioral Verilog simulation with several test cases and corner cases. VCS is the RTL simulator chosen for that verification.

A higher level view of the circuit blocks in Fig. 5.3 reveals their architectural functionality. In the first section “Multiply $A \times B$, Prepare Exponent, and Align C,” the significands $a[22:0]$ and $b[22:0]$ are multiplied in the multiplier on the left. In parallel with this multiplication is the alignment of the addend significand $c[22:0]$, which is right shifted by an amount determined by the anticipated result exponent $a[30:23] + b[30:23] + \text{offset} - c[30:23]$.

In the second section “Add and Normalize,” the aligned significand $c[22:0]$ is added to the remaining partial products of the multiplier to get `result_preshift`. The leading-zero anticipator (LZA) has anticipation logic that takes the intermediate terms from the added multiplier partial products and the “Final C Logic” which operates on the aligned significand $c[22:0]$. This anticipation logic provides the binary number whose leading zero position is determined by the leading-zero detector (LZD). The LZD tells the left shifter how much to shift `result_preshift`, thereby performing normalization. Then the final sign, significand, and exponent are determined.

In the final section “Select Output,” an incremented version of the exponent and significand are available, and a multiplexer chooses whether the final result is (1) left untouched, (2) if the incremented result is chosen instead for rounding, or (3) if a special constant such as infinity or not-a-number (NaN) is the appropriate result. Hence the “Select Output” section performs rounding. Put together, the circuits composing the FMA follow the architecture described in [25].

Considering the sheer number of circuit blocks in the FMA, NEM relay circuit blocks cannot be written purely manually. The infrastructure exists to turn a behavioral hardware description into a structural netlist of a NEM relay circuit, and this is discussed next.

■ 5.3 NEM Relays CAD Infrastructure for Synthesis

A set of computer-aided design (CAD) infrastructure has been maintained by the collaborative team that taped out the test chips shown in Chapter 3. The infrastructure includes the electronic design automation (EDA) to synthesize structural netlists of NEM relay circuits from a behavioral hardware description, place and route the structural netlists onto a chip layout with pins, simulate NEM relay circuits in an analog environment based on physical Verilog-A models or in a digital environment based on functional Verilog models, perform design rule checks (DRC) on a layout, perform a layout versus schematic (LVS) check, and more. This section focuses more on synthesis, while Section 6.3 covers other relevant aspects.

The NEM relay synthesis flow was developed by Kevin Dwan [30] and Cheng C.

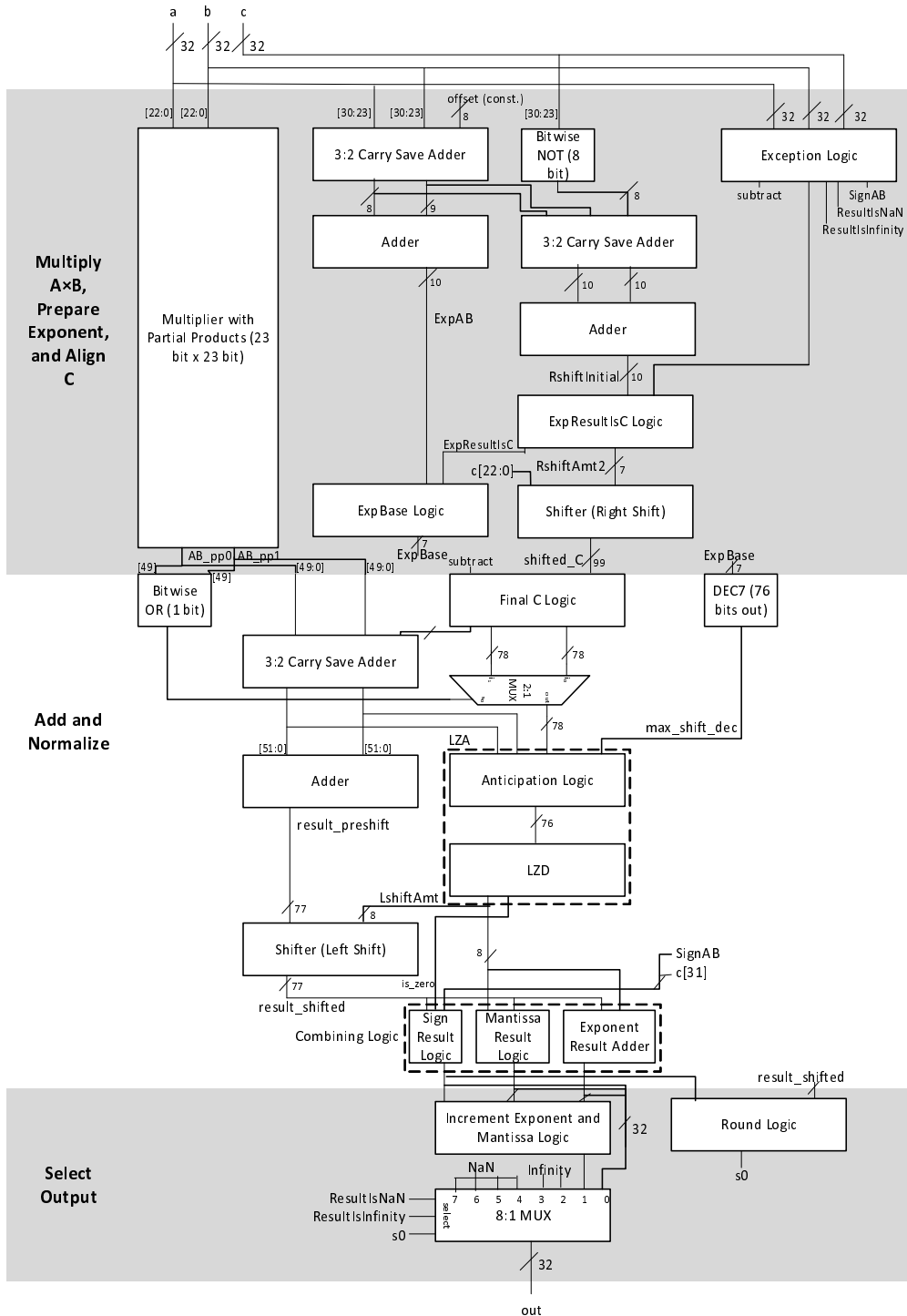


Figure 5.3. Fused multiply-add (FMA) designed with CMOS.

Wang at the University of California at Los Angeles. It involves:

1. Writing behavioral Verilog, which is usually a register-transfer level (RTL) hardware design abstraction.
2. Translating that into a gate-level netlist with a commercial tool, Cadence RTL Compiler, using a Liberty file that provides the Verilog primitives for logical functions with NEM relays.
3. Using the custom MATLAB script written by Cheng C. Wang that builds the full and then pruned binary decision diagram (BDD) and generates a NEM relays structural Verilog netlist based on that. 6T or seesaw NEM relays may be chosen for the circuit. The circuit made maintains a single mechanical delay operation.

Many cells that run within a single mechanical delay were synthesized by the aforementioned synthesis tool. The tool is designed to synthesize circuits that take only one mechanical delay because there is only one gate stage by design. The synthesis script may introduce dummy devices into the input rail to output path to ensure equal electrical delay between any input and output, though these devices may be safely removed.

One important consideration made by the synthesis tool is that it only connects inputs to the gates of NEM relays, it only connects outputs to the sources of NEM relays, and it only connects input rails to the drains of NEM relays and power rails to the bodies of NEM relays. This way, the logical path from input to output always goes from drain to source, and the state of whether a NEM relay is open or short circuit is determined by the gate and body input signals of a NEM relay.

Since the synthesis tool constructs a BDD to represent circuit functionality, the time taken complete synthesis depends exponentially on the number of inputs. This means that the tool will finish in a reasonable amount of time only when there are a small number of inputs. This makes it worthwhile to identify replicated blocks to synthesize only what is replicated, or otherwise manually design circuits when there is a clear pass-transistor design for it. On a side note, another synthesis tool for NEM relays circuits was released very recently and overcomes some prior issues [17].

After synthesis, place and route are performed. A process design kit (PDK) is made for each test chip's process technology, and it includes the complete logical, physical, layout, and schematic descriptions for NEM relay unit cells as well as the layer maps, LEF, and DRC/LVS decks. For place and route, the LEF and Liberty file are needed for layout information and timing information, respectively. The flow developed works with Cadence Encounter to perform place and route with a given process and export GDS for the final circuit layout. Although the commercial tools are timing-driven for CMOS designs, their intermediate results are intercepted to design for NEM relays.

Among the circuits designed for the FMA with NEM relays, some were synthesized while others were manually written as structural Verilog netlists due to the availability of a design or due to impractical synthesis times for circuits with a large number of

No. of input bits to pick from	No. of 2:1 MUX	No. of dual 2:1 MUX
2	1	1
4	3	2
8	7	4
16	15	8
32	31	16
n	$n - 1$	$\lceil \frac{n}{2} \rceil$

Table 5.1. Multiplexer sizes.

inputs. All cells were taken through the place and route flow described in Chapter 6. The manually optimized circuits for the FMA are described in the following sections.

■ 5.4 Decoders and Multiplexers with NEM Relays

Decoders and multiplexers are heavily used in the FMA, such as the decoders in the Exception Logic; these circuits are the first to be optimized for design with NEM relays.

A decoder has one output port for every combination of binary values on its input ports. Only one of those ports is active, and that port is chosen by the input code.

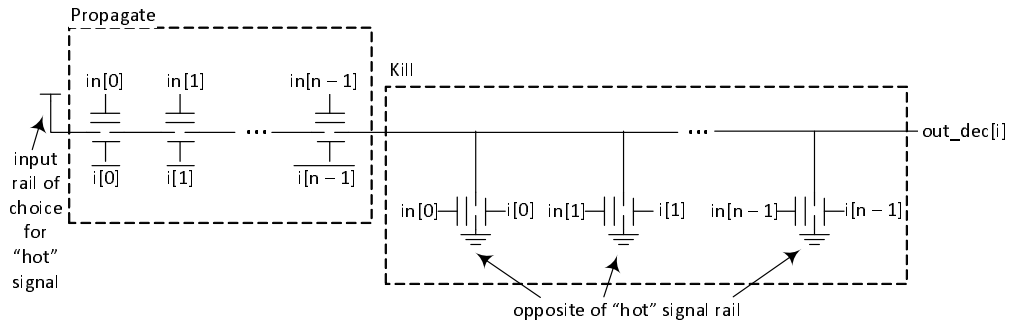
Decoders are designed with NEM relays as illustrated in Fig. 5.4. The port considered active receives the input rail of choice for the “hot” signal, which is typically V_{DD} but may be V_{SS} instead. The propagate path passes through the “hot” signal only if the input matches the input code unique to that path. If the input does not match the input code, a kill signal is sent with the opposite of the “hot” signal rail, which is typically V_{SS} . For a decoder that has n output bits, the input is at least $\log_2 n$ bits wide. For each output bit, there is a single decoder output with NEM relay body voltages defined by the input code that makes the output active.

This style of decoder circuit uses the minimum number of NEM relays needed to obtain a result within a single mechanical delay.

Multiplexers are designed with NEM relays as shown in Fig. 5.5. The seesaw NEM relay on the right of this figure has its sources connected together and different inputs on its drains, making it essentially a 2:1 multiplexer because the signal on the gate chooses which input to pass through to the output on the sources. The 2:1 multiplexer made of 6T NEM relays is actually a dual multiplexer, because it uses the same two devices to multiplex two sets of inputs using the same input on the gates.

Larger multiplexers are designed by cascading 2:1 multiplexers as shown in Fig. 5.6. This forms a tree-like structure. Table 5.1 gives the sizes of larger multiplexers. The number of 2:1 multiplexers corresponds to the seesaw NEM relay count and the number of dual 2:1 multiplexers times two is the 6T NEM relay count. The FMA contains many multiplexers in its behavioral hardware description, making it critical to know how many cascaded 2:1 multiplexers are needed for in the FMA circuit blocks designed with NEM relays.

Single Decoder Output for Input Code i



n-bit Output Decoder

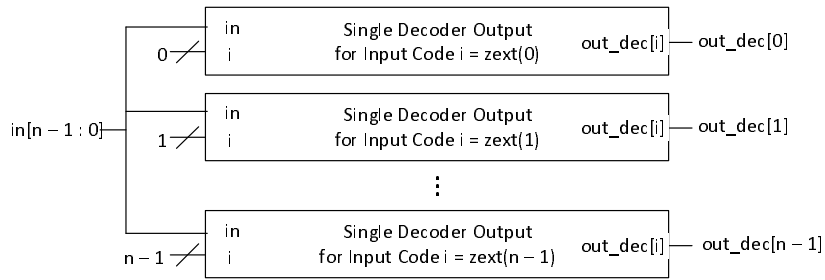


Figure 5.4. Generic optimal design of a decoder with NEM relays.

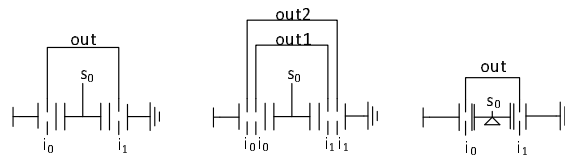


Figure 5.5. 2:1 multiplexers designed with 4T, 6T, and seesaw NEM relays. The 6T NEM relay implementation is a dual 2:1 multiplexer.

Frequently, many of the inputs to a multiplexer may be known to be constant values. In this case, it is best to prune NEM relay leaves from the multiplexer tree where the leaf would always output the same constant value to the stage above it. It turns out that the FMA has several circuit blocks where a large multiplexer is used and it is known ahead of time that certain inputs are constant. An example is the 8:1 MUX which selects the output of the FMA, where a number of the inputs are tied to a constant NaN or infinity signal. In such cases, not only should the 2:1 MUX sub-blocks whose output is already known be removed, but these sub-blocks also need to be removed so that they are not included in a switching power estimate because their activity factors are zero. The 2:1 MUX sub-blocks are referred to as tree leaves.

Let c be the number of adjacent constant inputs to multiplexers starting from the most significant or least significant input bit of a large multiplexer. It is required that $c < n$ in an n :1 multiplexer.

The number of 2:1 multiplexers that can be removed in a seesaw NEM relay n :1 multiplexer design is:

$$2^{\lfloor \frac{c}{2} \rfloor} - 1 \quad (5.1)$$

The number of dual 2:1 multiplexers that can be removed in a 6T NEM relay n :1 multiplexer design is:

$$\sum_{i=2}^{\lceil \log_2 c \rceil} \left\lfloor \frac{c}{2^i} \right\rfloor \quad (5.2)$$

Although these removals do not apply to a general multiplexer, they are handy for the special cases in the FMA when certain input signals are known to be unchanging.

■ 5.5 Shifters with NEM Relays

Shifter circuits are implemented with multiplexers, as shown in Fig. 5.8. Such shifters are called barrel shifters. This makes it crucial to apply the same type of optimization introduced for multiplexers to shifters. The large datapath width in the FMA also requires large shifters, meaning that shifters in the FMA consume a large number of devices nominally unless they are optimized.

The left and right shifter circuits, implemented with barrel shifters, need only perform logical shifts. Thus no additional circuits are devised to perform arithmetic shifts.

As was done with multiplexers in Section 5.4, barrel shifters can also have certain multiplexer leaves pruned. Again, this is necessary to get an accurate power estimate because such multiplexer leaves whose outputs are constant have an activity factor of zero.

The shifters in the FMA often have constant 0 bits in the input binary number that is shifted, starting from the most or least significant bit. In the case of shifters whose input bitwidth is between powers of two, that means the binary multiplexer

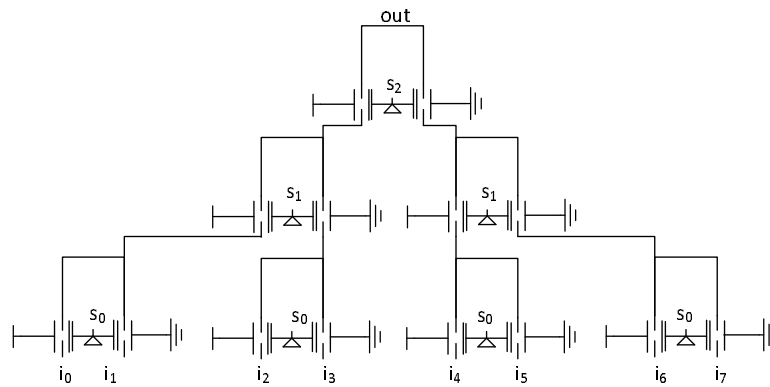
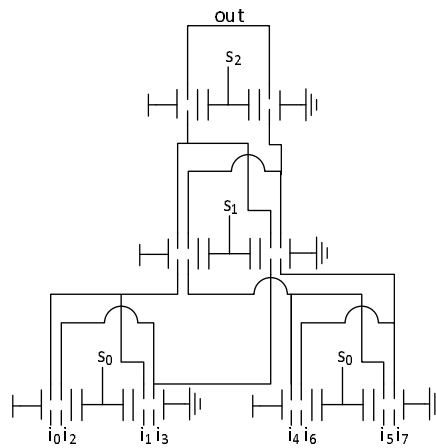
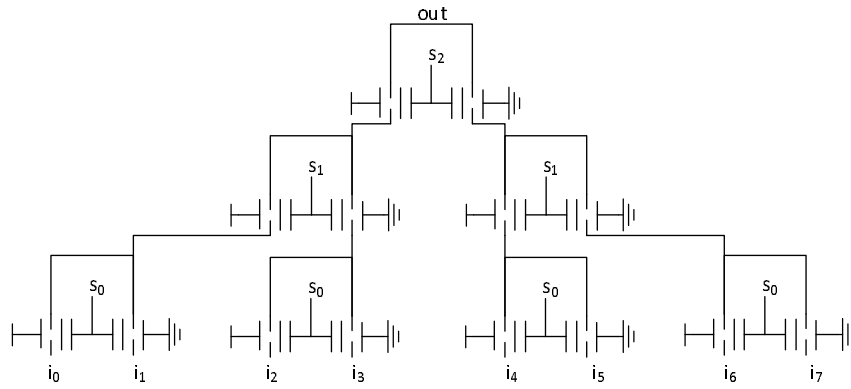


Figure 5.6. 8:1 multiplexers designed with 4T, 6T, and seesaw NEM relays. An 8:1 multiplexer comprises 2:1 multiplexers.

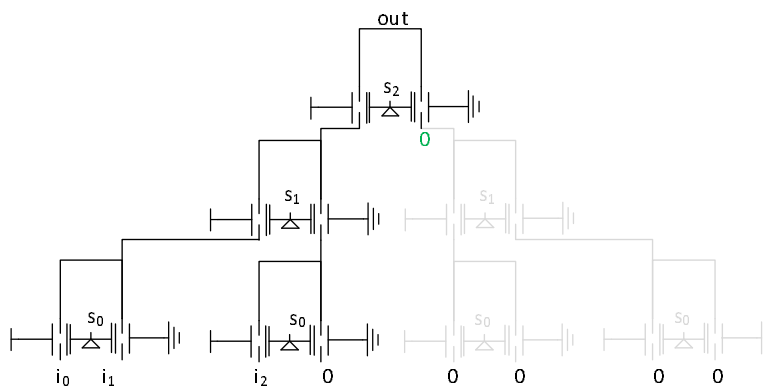
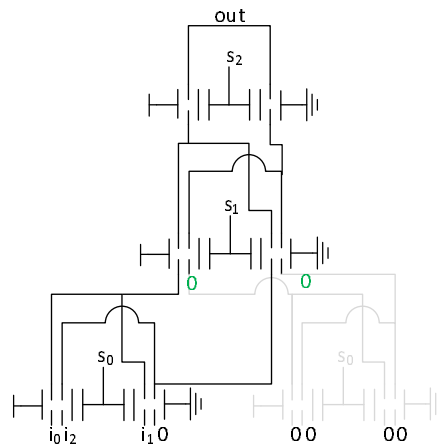
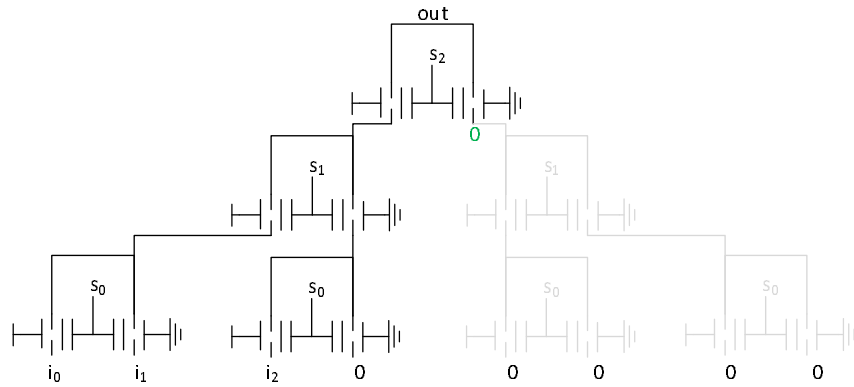


Figure 5.7. The 8:1 multiplexer of Fig. 5.6 may have a constant signal (e.g. 0 in this figure) on some inputs. For this case, the multiplexer tree is pruned to use the least number of NEM relays.

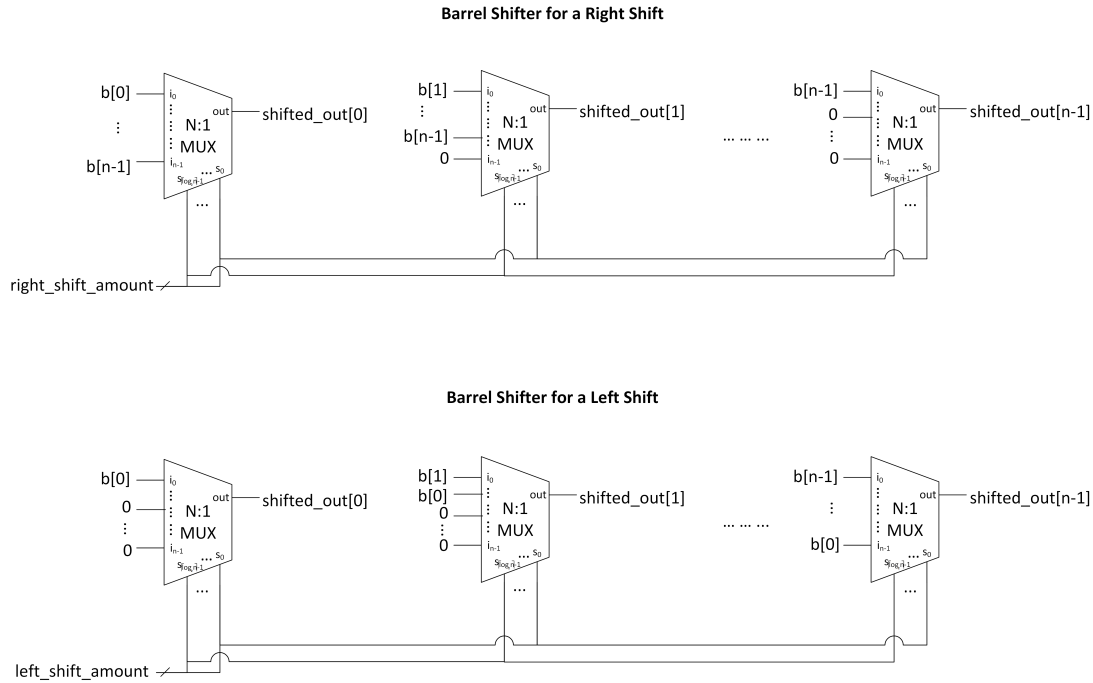


Figure 5.8. Barrel shifters used for right shift (top) and left shift (bottom) operations.

tree is bound to have leaves which are set as 0. This means that, in addition to the multiplexer inputs that are set at constant 0 for the filler bits of shifter outputs, there are also inputs that are also set to constant 0 due to the extra tree level needed for a shifter input bitwidth that is not an exact power of two.

The following formulation shows the number of 2:1 multiplexers with activity factors of zero, due to constant inputs, that can be removed from such shifters. Define y as the difference between the shifter binary multiplexer tree width and shifter input bitwidth. If there is an extra tree level, then $y > 0$ and the additional 2:1 multiplexer leaves may be removed.

$$y = 2^{\lceil \log_2 n \rceil} - 2^{\log_2 n} = 2^{\lceil \log_2 n \rceil} - n \quad (5.3)$$

Both types of leaves in the multiplexer trees with constant inputs are removed, without assuming that those sets of constant inputs are adjacent to each other. It turns out that the same number of leaves can be pruned for left shifters and right shifters.

The number of 2:1 multiplexers that can be removed from a seesaw NEM relay barrel shifter is:

	99-bit shifter	77-bit shifter
# Dual 2:1 MUX per output before pruning	64	64
# Dual 2:1 MUX total in shifter before pruning	6336	4928
# Dual 2:1 MUX removed for extra tree levels	154	536
# Dual 2:1 MUX removed for filler bits	2138	2138
# Dual 2:1 MUX total in optimized shifter	4044	2254

Table 5.2. Reduction of the number of dual 2:1 multiplexers that compose the 6T NEM relay barrel shifters in the FMA, based on inputs known to be constant. Each dual 2:1 MUX has 2 6T NEM relays.

$$\sum_{i=1}^y \left(2^{\lfloor \frac{i}{2} \rfloor} - 1 \right) + \sum_{i=1}^{n-1} \left(2^{\lfloor \frac{i}{2} \rfloor} - 1 \right) \quad (5.4)$$

The number of dual 2:1 multiplexers that can be removed from a 6T NEM relay barrel shifter is:

$$\sum_{i=1}^y \left(\sum_{j=2}^{\lceil \log_2 i \rceil} \left\lfloor \frac{i}{2^j} \right\rfloor \right) + \sum_{i=1}^{n-1} \left(\sum_{j=2}^{\lceil \log_2 i \rceil} \left\lfloor \frac{i}{2^j} \right\rfloor \right) \quad (5.5)$$

Table 5.2 shows the reductions that were applied to the 99-input barrel shifter and 77-input barrel shifter in the FMA design.

■ 5.6 The Leading Zero Detector (LZD) Optimized for NEM Relays

The leading zero detector (LZD) is in the critical path of the FMA, making it crucial to reduce its number of mechanical delays. An LZD takes a binary number on the input \mathbf{b} and provides an output \mathbf{p} which encodes the position of the first 1 bit from the most significant bit of \mathbf{b} in binary. Only if \mathbf{b} is all 0s will the output valid bit \mathbf{v} be 0.

The behavioral Verilog for the 76-bit LZD used in the FMA, when synthesized for CMOS, yields 13 gate stages, which translates to 13 mechanical delays. The alternative to this CMOS-style circuit design is to identify ways to reduce gate stages to significantly reduce the number of mechanical delays.

One difficulty with creating a single- t_{mech} LZD is that the 76-bit LZD has 76 inputs, which is far more than what can be synthesized in a reasonable amount of time with NEM relays. Different size LZDs were synthesized starting with the smallest 2-bit LZD, and synthesis results are reported in Fig. 5.9.

It turns out that a custom RTL description of the LZDs yielded a smaller, more efficient synthesized design compared to that synthesized with the available RTL description in FPGen. The custom behavioral description is called the Tree Design in Fig. 5.9 and the Tree Design is illustrated in Fig. 5.10.

LZDs may be combined to form larger LZDs [31]. Fig. 5.11 shows how an n -bit LZD is created with two $\frac{n}{2}$ -bit LZDs and additional logic. This means that the largest single-

Modular Design	Using 2-bit LZDs	Using 4-bit LZDs	Using 8-bit LZDs	Using 16-bit LZDs
2-bit LZD	1	N/A	N/A	N/A
4-bit LZD	2	1	N/A	N/A
8-bit LZD	3	2	1	N/A
16-bit LZD	4	3	2	1
32-bit LZD	5	4	3	2
64-bit LZD	6	5	4	3
76-bit LZD	7	6	5	4
128-bit LZD	7	6	5	4

Table 5.3. Number of mechanical delays, t_{mech} , when using single t_{mech} LZDs and combining them to form larger LZDs with the multiplexing technique presented in [31].

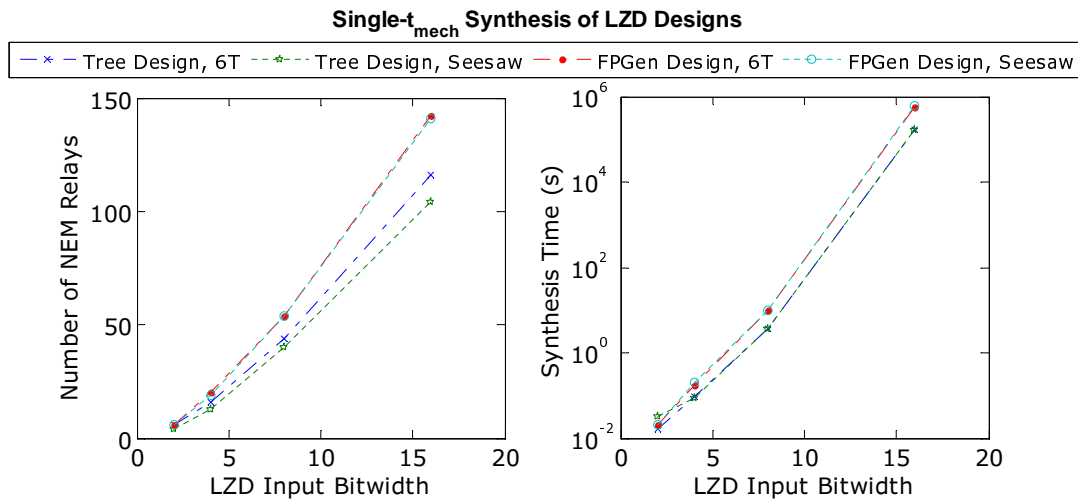


Figure 5.9. Complexity of LZDs. These are synthesized LZDs before pruning unnecessary relays.

t_{mech} LZD that can be synthesized can be used to build a larger LZD of the desired size using the design strategy in Fig. 5.11. Table 5.3 shows the number of mechanical delays for LZDs constructed by combining various smaller size single- t_{mech} synthesized LZDs. Although the multiplexer and OR gate in Fig. 5.11 could be pass-transistor designs, the NOT gate and select bit of the multiplexers must have an input on a gate [32]. This explains the single t_{mech} , labeled in Fig. 5.11, which is accrued when two LZDs are combined for a larger LZD.

LZDs made through synthesis also have devices that are not used that can be pruned from the tree schematic. For the various single- t_{mech} LZDs synthesized into seesaw NEM relay circuits, the number of relays in the circuit before and after pruning is shown in Table 5.4.

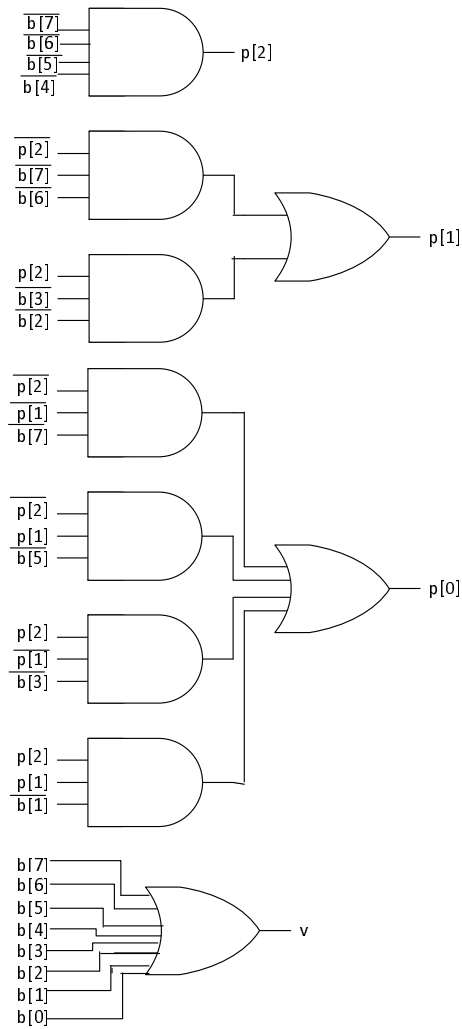


Figure 5.10. Tree design used as the RTL for area-efficient LZDs. This particular design is for an LZD with 8 inputs.

	2-bit LZD	4-bit LZD	8-bit LZD	16-bit LZD
Before pruning	4	13	40	111
After pruning	3	8	22	58

Table 5.4. Number of seesaw NEM relays before and after pruning unnecessary seesaw NEM relays from the synthesized leading zero detectors (LZDs). The schematics for the pruned LZDs are in Figs. 5.12, 5.13, 5.15, and 5.16.

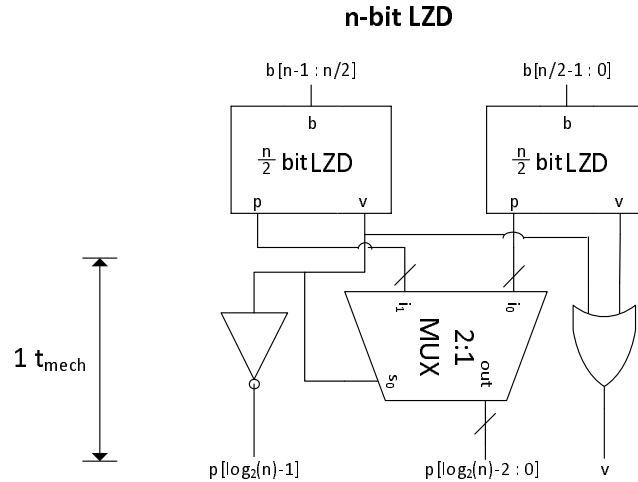


Figure 5.11. Generic modular design of an LZD with n input bits using two LZDs with half the number of input bits [31].

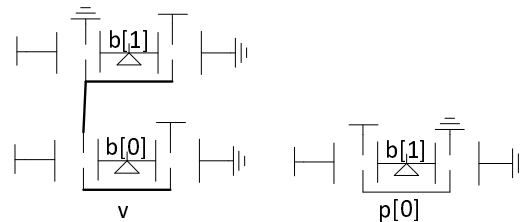


Figure 5.12. LZD with 2 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices.

The schematics for the single- t_{mech} LZDs with 6T and seesaw NEM relays are shown in Figs. 5.12, 5.13, 5.14, 5.15, and 5.16. The schematics with seesaw NEM relays are optimized to use the lowest number of devices, and the complementary nature of the devices gives insight on the circuit function and general design of LZDs. First, it is observed that the output v is always the result of an OR operation on all bits of the LZD input, b . This OR logic tree is separate from the logic for the p outputs. Second, the logic tree for $p[0]$ includes all b bits and has alternating paths to the input supply rails, skipping NEM relays with even or odd bits of b on the gate. Third, the most significant bit of p depends on only the upper half of bits in b . Fourth, the middle bits of p also have alternating paths like $p[0]$ but they skip a different number of NEM relays. By seeing these patterns, only certain middle bits of output p need to be synthesized for even higher-bit LZDs because the other output logic trees are already understood.

The LZD circuits designed and shown in this work can also be used as pass-transistor implementations of the same circuit in CMOS.

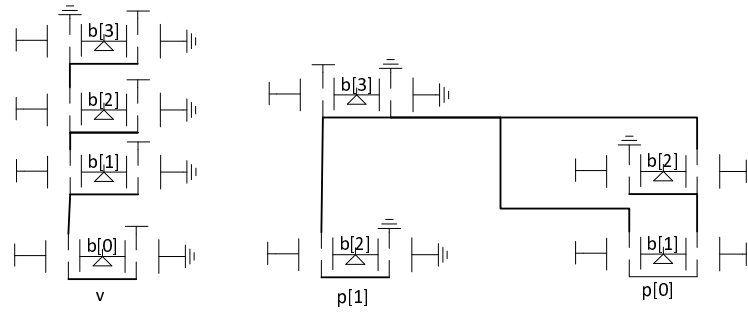


Figure 5.13. LZD with 4 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices.

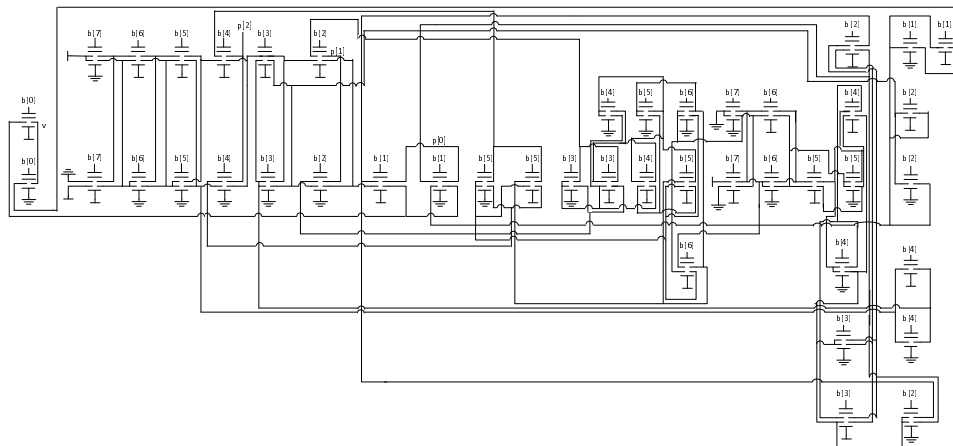


Figure 5.14. LZD with 8 input bits designed with 6T NEM relays. The circuit runs in one mechanical delay. This circuit design is synthesized and dummy NEM relays introduced are not removed from this schematic, making it less intuitive.

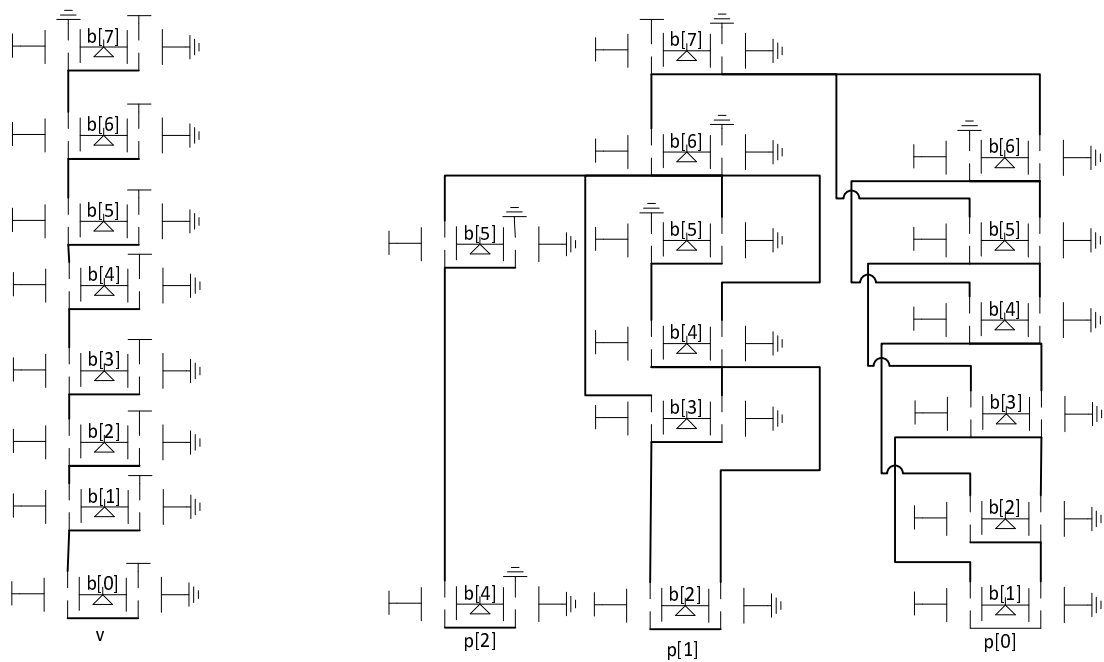


Figure 5.15. LZD with 8 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices, and it provides the intuition for higher-bit single- t_{mech} LZD designs.

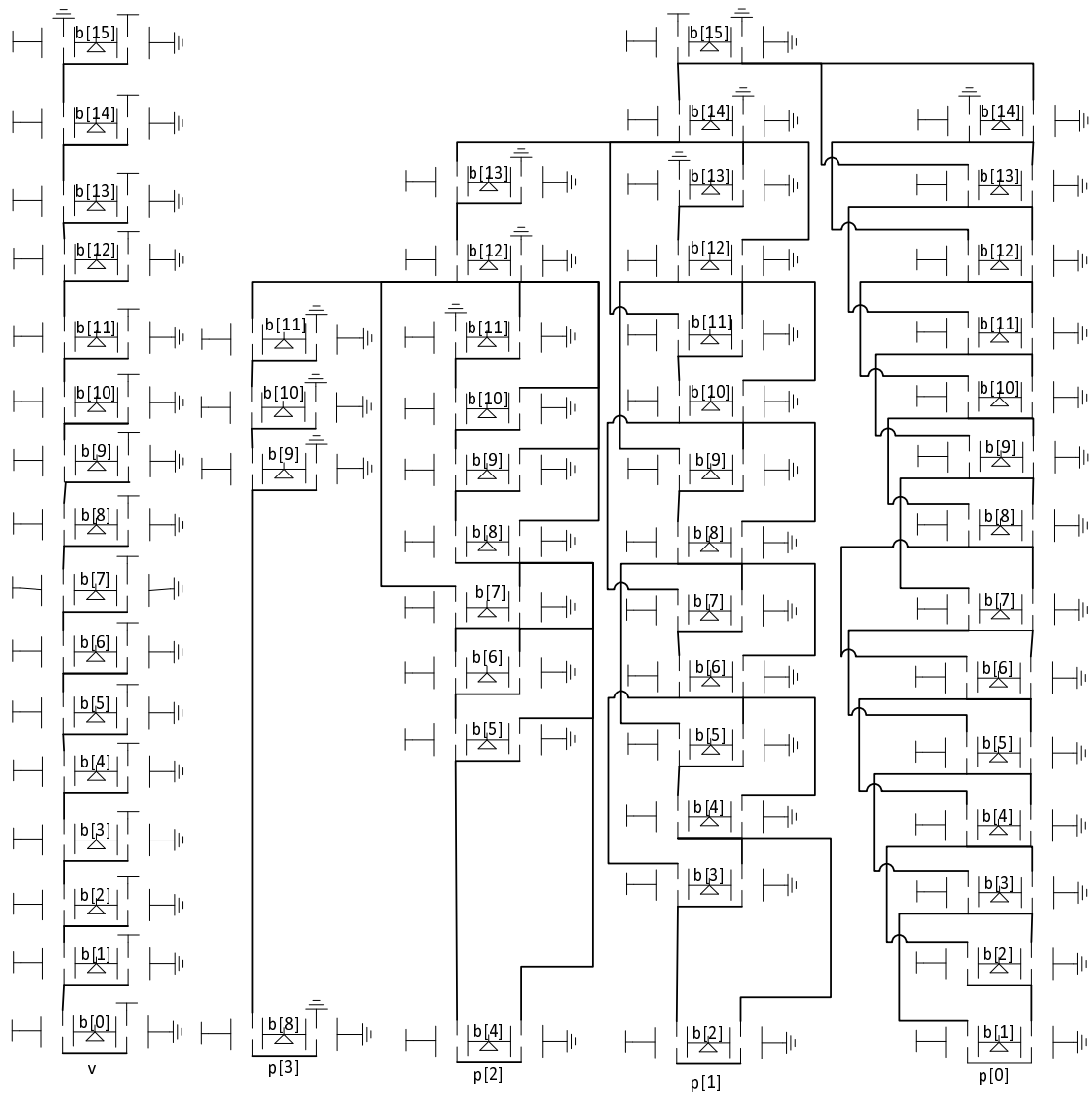


Figure 5.16. LSD with 16 input bits designed with seesaw NEM relays. The circuit runs in one mechanical delay. This design is optimized to use the lowest number of devices, and it provides significant intuition for higher-bit single- t_{mech} LSD designs.

■ 5.7 The Leading Zero Anticipator (LZA) Optimized for NEM Relays

The FMA generated by FPGen uses a leading zero anticipator (LZA) for the FMA. An LZA contains Anticipation Logic and an LZD. The Anticipation Logic developed in this work is based on FPGen RTL but is custom built with NEM relays pass-transistor style logic [24]. The Anticipation Logic operates on the propagate, generate, and kill terms of the adder that feeds into it and the `Final C` values that also feed into it. It is observed that the Anticipation Logic is bitwise, meaning that the same logic for a bit i can be applied to all bits. Boolean logic symbols are used to describe the Anticipation Logic block of the LZA for each bit i . The Anticipation Logic takes propagate (P), generate (G), and kill (K) signals used to get the sum of A and B, and combines that information with the decoded `ExpBase` signal, known as `max_shift_dec` or M , to produce an edge vector whose leading zero is detected by the LZD following the Anticipation Logic.

$$P_i = A_i \oplus B_i \quad (5.6)$$

$$G_i = A_i B_i \quad (5.7)$$

$$K_i = \overline{A_i B_i} \quad (5.8)$$

The unit bitwise logical operation of the Anticipation Logic is:

$$V_i = (P_{i-2} \oplus G_{i-1} \oplus G_i) + M_i \quad (5.9)$$

This bitwise logical description is synthesized as a single- t_{mech} 6T NEM relays circuit and replicated for each bit going into the LZA for the purpose of FMA power estimation.

■ 5.8 Circuit Topology Analysis for Area, Energy, and Delay Trade-offs

Although there is a relatively clear-cut process to design of the adder, decoder, multiplexer, shifter, and other logical circuit blocks in the FMA, some knobs remain to pick a design based on the desired trade-offs. Knobs of interest are the multiplier and leading zero detector, which come in several designs, and the optimal choice to make for each depends on the desired trade-off.

An adder of any bit-width is designed with a Manchester carry chain, which provides a result within a single mechanical delay [2]. This design is already optimal, and thus it is left as is.

Table 5.5 lists the number of 6T NEM relays for each single- t_{mech} synthesized LZD. The process for determining the energy per operation is described in Chapter 6, but since the 16-bit LZD is the chosen design here onwards, the E_{op} for the other contending single- t_{mech} LZD circuits is shown in Table 5.5.

Considering circuit topology alone, the energy consumption is directly proportional to the number of NEM relays in the circuit. Large LZDs that are designed for a single t_{mech} tend to have more NEM relays than the same-input-size LZD made using the technique in Fig. 5.11 made of two LZDs of half the input bitwidth. Specifically the

LZD Circuit	6T NEM Relays	E_{op} (aJ)
2-bit LZD (FPGen Design)	6	32.1
2-bit LZD (Tree Design)	6	18.4
4-bit LZD (FPGen Design)	20	72.2
4-bit LZD (Tree Design)	16	38.7
8-bit LZD (FPGen Design)	54	151.2
8-bit LZD (Tree Design)	44	50.1
16-bit LZD (FPGen Design)	142	129.6
16-bit LZD (Tree Design)	116	44.3

Table 5.5. Single- t_{mech} LZDs designed with 6T NEM relays, any of which may be used in a NEM relays FPU depending on the desired trade-off. These designs were not pruned. E_{op} is based on the analysis in Chapter 6, and is shown here because only the 16-bit LZD (Tree Design) is used in the final optimized FMA in Table 6.3.

former design has 1.4 times the number of relays of the latter design, not including the MUX, NOT, and OR gate in the latter design. This suggests that the higher throughput enabled by the former design is traded off with the lower area and energy of the latter design. This trend applies to the LZDs that are synthesized as well as to the manually optimized LZDs that have unnecessary NEM relays pruned off in Table 5.4. In this table of optimal LZD sizes, the minimum number of NEM relays also increases more than twofold for an LZD that has twice as many inputs.

The design trade-offs for the LZD are reminiscent of the design trade-offs uncovered for the multiplier. A multiplier with large compressors is used in the FMA design, providing the shortest number of mechanical delays out of designs available. This choice comes with the penalty of additional NEM relays in the circuit compared to a multiplier built of smaller (3:2) compressors, which has 60% of the number of NEM relays compared to the design with large compressors [5]. Hence, the design trade-off between a higher delay, smaller-area multiplier with (3:2) compressors and a shorter delay, large-area multiplier with large compressors, is much like the design trade-off between higher-delay, smaller area LZDs and lower-delay, large-area LZDs.

The LZD could be designed to yield a result within a single mechanical delay, and the framework to do so has been laid out in this work. However, doing so saves only three t_{mech} in the whole FMA, which is only a 19% improvement over the FMA design presented here.

6T and seesaw NEM relays are used to design the FMA circuit. One topological note is that, for the same LZD design, the ratio of the number of 6T NEM relays for the LZD to the number of seesaw NEM relays for the LZD starts at 1.5 for the 2-input LZD, and settles at 1.1 as the LZDs get larger. This means that 6T NEM relay circuits becomes just as area efficient as seesaw NEM relay circuits when the BDD for the function is large. Reducing the number of NEM relays means that the amount of switching power is also reduced, unless the relays removed already had a zero activity factor.

The completed design of the FMA with NEM relays is shown in Fig. 5.17.

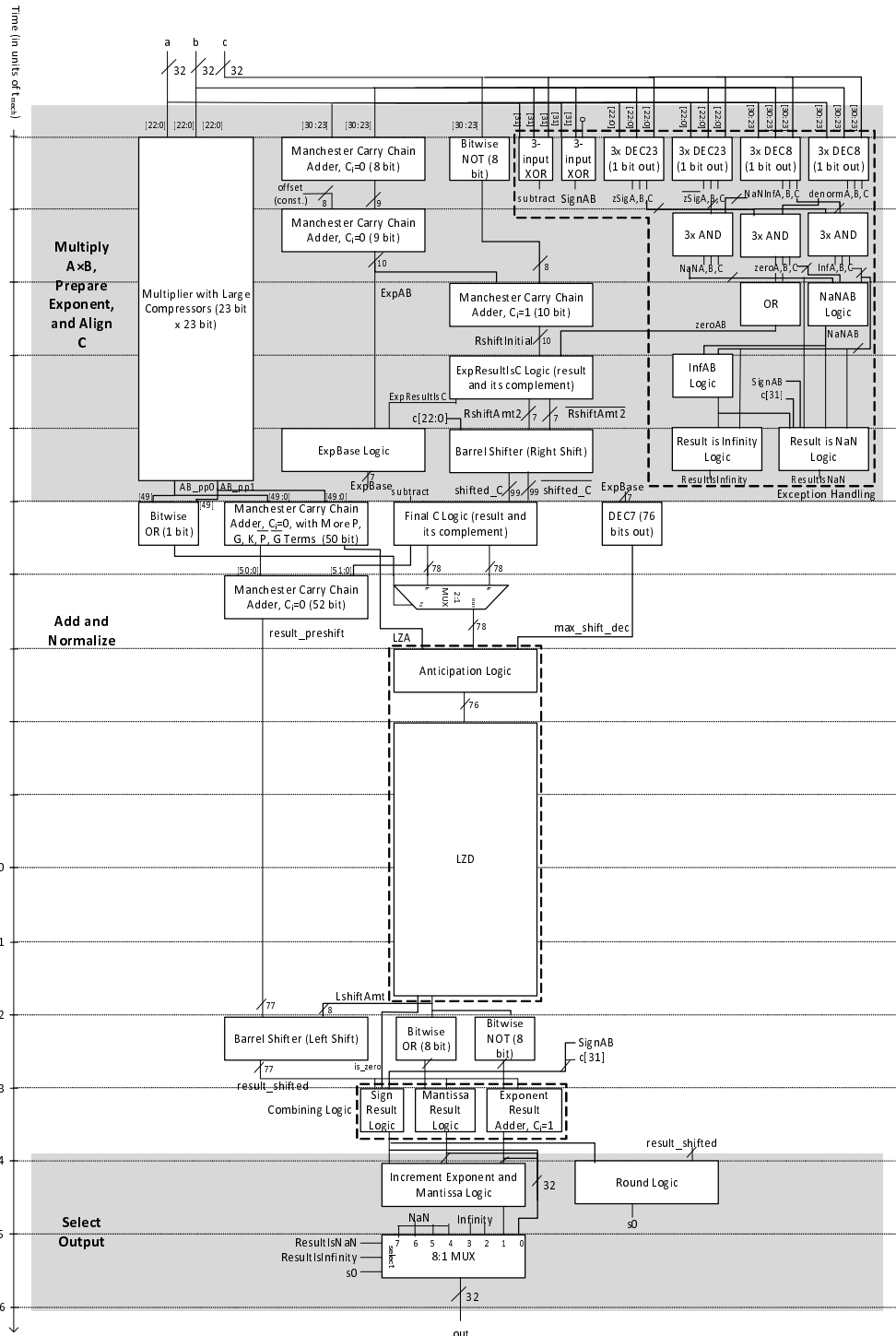


Figure 5.17. Fused multiply-add (FMA) designed with NEM relays.

Energy and Performance Analysis and Comparison with CMOS

THE energy per operation, performance, and area trade-offs for the design of the FMA with NEM relays are considered in this chapter. A scalable technique to analyze these variables in large systems of NEM relay circuits is presented.

■ 6.1 Basis for Analysis

The power consumption of NEM relay circuits is determined exclusively based on their dynamic energy, $E_{dynamic}$. This is the same as switching energy, E_{sw} , and is given for a single circuit node by the voltage swing and the capacitance driven.

$$E_{sw} = \frac{1}{2}CV_{DD}^2 \quad (6.1)$$

With the additional knowledge of the activity factor α on the node and the frequency at which the inputs are toggling, f , one determines the dynamic power of this switching event:

$$P_{dynamic} = \alpha f E_{sw} \quad (6.2)$$

$$\frac{P_{dynamic}}{f} = \alpha E_{sw} \quad (6.3)$$

The total energy per operation, E_{op} , is taken by combining the switching energy information for each node i :

$$E_{op} = \sum_i \alpha_i E_{sw} \quad (6.4)$$

$$P_{dynamic,total} = \sum_i \alpha_i f E_{sw} \quad (6.5)$$

$$E_{op} = \frac{P_{dynamic,total}}{f} \quad (6.6)$$

Parameter	Present Scaled 6T Relay	Predictive 90nm 6T Relay
A_{ov} (μm^2)	45	1.54
g_0 (nm)	100	10
g_d (nm)	50	5
R_{on} ($\text{k}\Omega$)	2-3.5	2-3
C_{gb} (fF)	7.2 (on), 3.8 (off)	2.3 (on), 1.46 (off)
C_{gc} (fF)	16	0.4
$C_{gs/d}$ (fF)	0.32 (on), 0.17 (off)	0.3 (on), 0.18 (off)
V_{pi}, V_{po} (V)	8, 6	0.04, 0.03
t_{mech} (μs)	0.2 ($V_{DD} \cong 2V_{pi}$), 1 ($V_{DD} \cong 1.2V_{pi}$)	0.02 ($V_{DD} \cong 2V_{pi}$), 0.08 ($V_{DD} \cong 1.2V_{pi}$)

Table 6.1. Device model parameters for current and predicted NEM relay process technology. Data from [11].

Note that the designer may choose to separate input rails to the drains, V_{DDSD} and V_{SSSD} , from the input rails to the gates and bodies, V_{DD} and V_{SS} . In this case, the aforementioned energy per operation analysis holds except now the switching energy becomes:

$$E_{sw} = \frac{1}{2}C(V_{DD} - V_{SS})(V_{DDSD} - V_{SSSD}) \quad (6.7)$$

The analysis in this work does not separate the drain input rails from the gate and body input rails, so Equation 6.1 applies throughout this work. However, Equation 6.7 offers an opportunity to further lower energy per operation, because $(V_{DDSD} - V_{SSSD})$ is limited by the hysteresis gap $(V_{pi} - V_{po})$.

■ 6.2 Process Technology

The present state-of-the-art working process technology is the Scaled 6T process. To understand the benefits of energy-efficient computing when the process is scaled further, the 90nm process technology node equivalent to that in CMOS is predicted. The device model parameters for both of these processes are reproduced in Table 6.1 and completely described in [11].

The capacitance seen at the gate of a NEM relay is given by the NEM relay capacitances between the gate and other effective grounds. This same capacitance can be used to approximate the capacitance seen by the output of a NEM relay circuit, C_{out} , which is assumed to drive the input of another NEM relay.

$$C_{out} = C_{gb} + C_{gc} + 2C_{gs/d} \quad (6.8)$$

■ 6.3 NEM Relays CAD Infrastructure for Verification and Analysis

Once structural Verilog netlists for NEM relay circuits are made following the discussion in Section 5.3, they can be functionally verified in a logical simulation, verified in a physical simulation, or placed and routed to be fabricated on a test chip.

A scan simulation flow is developed to simulate structural Verilog netlists of scan chains with the Verilog-A device models. The Verilog-A model was written by Fred Chen and Hei Kam [1], and was significantly improved by Matthew Spencer. This captures the analog device behavior to predict scan chain behavior. The flow developed automates the choice of either Verilog or SPICE netlist simulation with HSPICE or Spectre with the appropriate Verilog-A model for any duty cycle or input pattern. Structural Verilog netlists are automatically converted to SPICE netlists if needed. This flow was used to verify the scan chain on a microcontroller on a taped out NEM relays chip.

This scan simulated flow is repurposed for the general purpose of simulating any circuit with the MEMS Verilog-A device model, stimulating all possible input combinations. This flow is used for getting the total power consumption of various FMA circuit blocks designed with NEM relays.

The place and route process may be used to estimate the power consumption of the circuit. Based on values in the Liberty file, the place and route flow in Cadence Encounter can also give an estimate of the leakage power, switching power, and internal power. Since there is no leakage power with NEM relays, and internal power is just a way to track power back to certain cells, only switching power is of interest. Switching power is calculated by Equation 6.2. The place and route flow was heavily modified to get a switching power estimate for every circuit block in the FMA. Interconnects are described in Section 6.6. The place and route switching power estimation technique can be scaled to the overall circuit complexity of the FMA, and the process technology for the estimation can be scaled as well, making power estimation preferred over physical Verilog-A-based simulations.

■ 6.4 Energy per Operation for Each Subcircuit

A functional Verilog simulation is necessary to perform a switching power estimate. In this simulation, every possible input combination is stimulated in order from all 0 to all 1, stepping at a frequency f . The functional simulations use a logical model of the 6T NEM relay device, which is represented as a tri-state buffer that is enabled when the gate signal is opposite the body signal. The toggle activity and waveforms on all nodes is recorded in a VCD file, which is then used by the power estimation tool to get α and f . The switching power estimation reports include the transition density, $T = \alpha f$. The frequency chosen for all circuit functional simulations is $f = 100$ kHz. Note that E_{op} does not depend on f because f cancels out; nevertheless, f is chosen as a reasonable NEM relay circuit frequency. Based on the transition density reported for each functional Verilog simulation, α was calculated and verified to be correct for

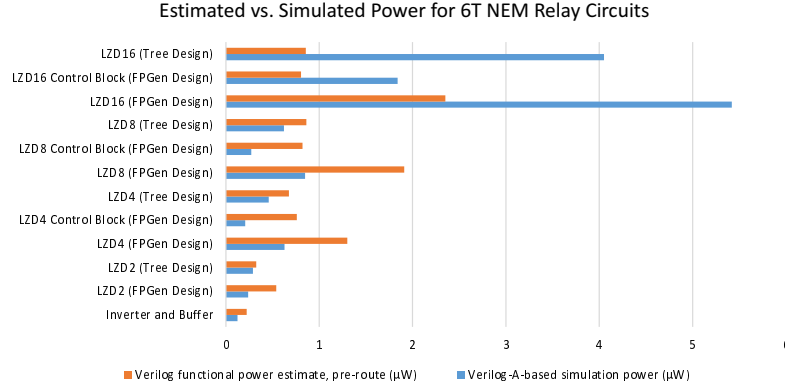


Figure 6.1. Comparison of the switching power estimation and the circuit power consumption from a physical-model-based simulation for the same circuits designed in the present scaled 6T NEM relay process. Both the functional Verilog simulation and Verilog-A-based simulation use $f = 100$ kHz. This shows that switching power estimation is a reasonable approximation in most cases.

circuits that could be manually verified. The functional Verilog simulations also verified the correct operation of the circuits designed.

Circuit switching power estimates also require device pin capacitances. The pin capacitances chosen for the Liberty file are $C_s = C_{out}$, $C_g = C_{out}$, $C_b = 0$, and $C_d = C_{gs/d}$, where C_{out} is given by Equation 6.8. C_s matters the most because the sources are the output nets and C_s is used to determine the energy when an output switches. The underlying assumption is that each NEM relay output drives another NEM relay. This is a good approximation of the widely varying output capacitances of each NEM relay. To validate this claim, switching power estimates for various circuits were compared against the power consumption reported in a physical simulation of the same circuits based on the 6T NEM relay Verilog-A model.

The Verilog-A-based physical simulation is run based on the flow created to stimulate all input combinations at a given frequency. The power consumption of the simulated circuit, where all input combinations are stimulated in the simulation within a total time t_{sim} , is:

$$P_{total} = \frac{\int_0^{t_{sim}} I_{DD} dt}{t_{sim}} V_{DD} \quad (6.9)$$

Interconnect power is not included in the physical simulation because it only simulates devices, and glitch power is included in the physical simulation.

A comparison of the pre-route switching power estimate with Equation 6.5 and physically simulated circuit power consumption given by Equation 6.9 is shown for various circuits in Fig. 6.1. The Liberty file for the power estimation and Verilog-A model for the physical simulation are for the present scaled 6T relay process in Table 6.1. The same $f = 100$ kHz is used in both the functional Verilog simulation and

the Verilog-A-based physical simulation. Results show a close match except when the transition density is very low, allowing physical glitch power to dominate as in the case of the 16-bit LZD designs.

■ 6.5 FMA Analysis and Comparison between CMOS and NEM Relays

The critical path in the final optimized FMA design leads to a total of 16 mechanical delays for the FMA operation. In the chosen predictive 90nm 6T NEM relays technology node, the mechanical delay set by the overdrive voltage is $0.08 \mu\text{s}$ for $V_{DD}=1.2V_{pi}$. The total time to complete one FMA operation is on the order of $1 \mu\text{s}$.

The FMA energy vs. performance trade-off is shown in Fig. 6.2 for CMOS and NEM relays. The energy per floating-point (FP) operation for the FMA designed optimally with state-of-the-art CMOS is from [24]. Up to a factor of 15 reduction in energy per operation can be realized by switching the process and design paradigm from CMOS to NEM relays. The FMA throughput is slower with NEM relays, but this can be made up by using parallelism.

■ 6.6 Interconnect Model Including Scaling

An interconnect model is devised for the 90nm predictive 6T NEM relay process for the purpose of more accurate circuit power estimation. Based on suggestions in [33], tungsten (W) metal is chosen as the first routing layer connected to the NEM relay electrodes and aluminum (Al) metal is chosen as the second routing layer. The W routing layer thickness is 50 nm and the Al routing layer thickness is 125 nm. Additionally, the minimum width is 120 nm and minimum pitch is 188 nm for both the W routing layer and Al routing layer.

Process data in the library exchange format (LEF) for the process technology of the scaled 6T relay in [11] is scaled down to the process technology of the 90nm equivalent 6T relay. These two process nodes are in Table 6.1. This information is combined with an interconnect capacitance table (ICT) that is based on the aforementioned routing layer dimensions for the 90nm process. The capacitance information is used with the routed circuit layouts for a more accurate switching power estimate. Although the interconnect dimensions used are not optimal, they are reasonable and are scaled from the present process. Table 6.2 shows how the interconnect dimensions are scaled in the LEF and ICT files from the scaled 6T to the 90nm equivalent 6T process node.

In CMOS processes, the load capacitance has not reduced much in the past few generations of transistor scaling. This is largely because it is limited by the interconnect capacitance. NEM relays processes offer the option to further reduce interconnect capacitance while compensating by increased resistance. This, in turn, leads to lower switching power in scaled NEM relay processes with scaled interconnects.

Fig. 6.3 shows the LZD2 circuit after it goes through the place and route flow for power estimation. The placed and routed circuit enables post-route power estimations including the effect of interconnect capacitances. Appropriate scaling factors are ap-

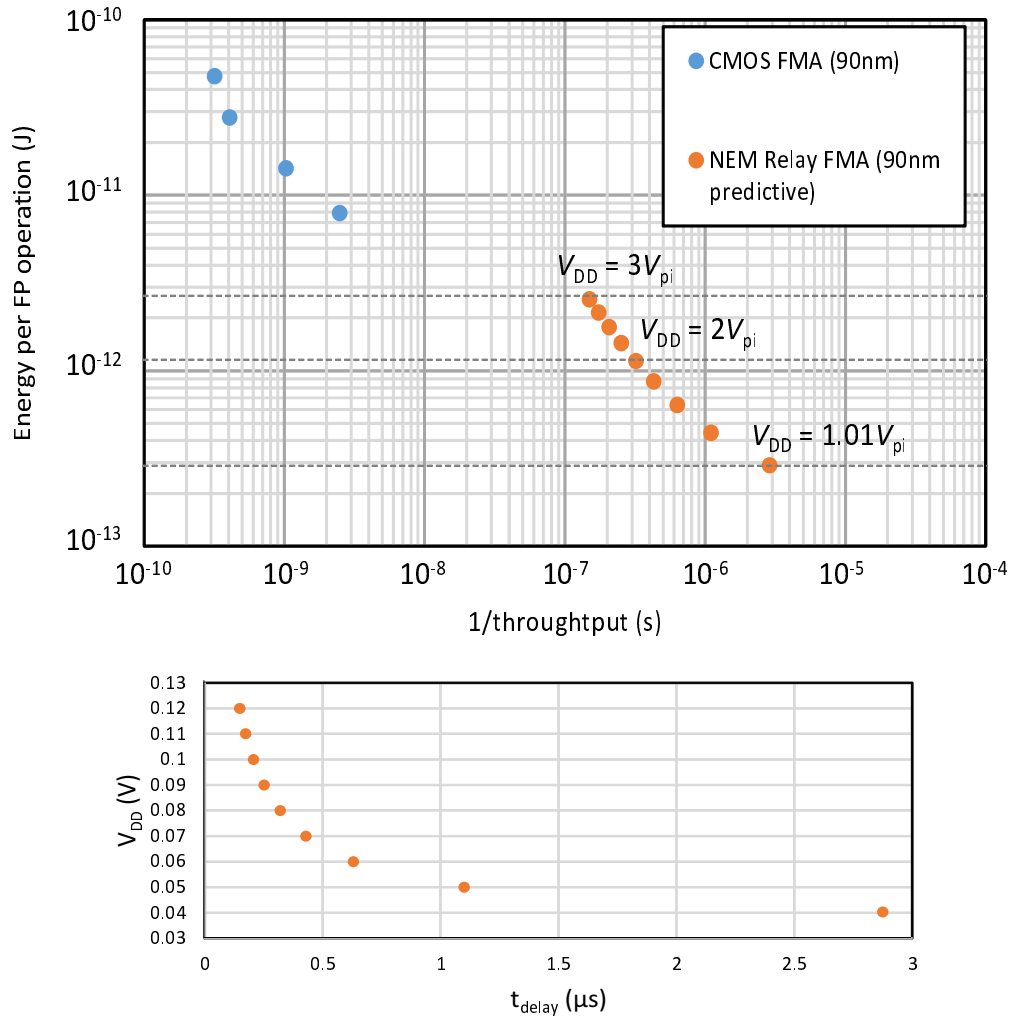


Figure 6.2. Energy and performance trade-off for equivalent fused multiply-add circuits made in the 90nm CMOS technology node [24] and predictive 90nm equivalent NEM relays technology node (top). The total latency of the NEM relay FMA depends on V_{DD} (bottom).

	Present Scaled 6T Relay		Predictive 90nm 6T Relay	
Routing Layer	Metal2 (Al)	Electrode (W)	Metal2 (Al)	Electrode (W)
Width (μm)	0.8	0.8	0.12	0.12
Area (μm^2)	0.64	0.64	0.096	0.096
Pitch (μm)	1.25	1.25	0.1875	0.1875

Table 6.2. Scaling of interconnects with a constant scaling factor of 0.15.



Figure 6.3. A placed and routed LZD2 circuit in the scaled 6T process technology, whose physical data is scaled to the predictive 90nm equivalent 6T NEM relays process technology and used alongside a predictive 90nm equivalent 6T NEM relays process interconnect capacitance table file to estimate post-route switching power.

plied within the place and route flow to scale the route lengths and device sizes to the predictive 90nm equivalent 6T NEM relays process technology. An interconnect capacitance table (ICT) file including the expected wire capacitances in the predictive 90nm equivalent 6T NEM relay process technology is used by the switching power estimation tool for post-route switching power estimates.

With the ICT available in the switching power estimation flow, both pre-route and post-route power estimates were available. The pre-route power estimates were approximately half of the post-route power estimates for each circuit cell. The complete post-route energy per operation breakdown by circuit block in the FMA is in Table 6.3.

The multiplier used in the FMA is a 32-bit multiplier with Booth encoding that takes five mechanical delays to produce a result. The energy per operation for a 16-bit multiplier built in the same 90nm equivalent 6T NEM relay process technology used for all FMA circuits in this work is reported in [11]. To extrapolate the energy for a 32-bit multiplier in that work, an approximation for it is to multiply E_{op} for the 16-bit multiplier by the ratio of the number of relays in the 32-bit multiplier to the number of relays in the 16-bit multiplier. This yields $(9272/3211) \times (24 \text{ fJ}) = 69 \text{ fJ}$ for the multiplier in the FMA. The multiplier energy per operation includes the energy dissipated in interconnects [11], and thus this E_{op} can be compared with the E_{op} of the remaining circuit cells in Table 6.3 which also includes interconnect energy dissipation.

The energy breakdown of the FMA shows that multiplexers are the dominant consumer of energy, followed by the multiplier and adders. Shifters include these multiplexers. The multiplexers and shifters are pruned as described in Chapter 5, ensuring that nodes with no activity are not included. The multiplier energy could be further reduced if (3:2) compressors were used instead of large compressors, though it may be possible that the multiplier would then get into the critical path of the FMA due to its slightly higher number of mechanical delays. Since the energy analysis is done on a circuit cell by cell basis, the activity factors are based on the individual cells and are

Circuit	Units	E_{op}/unit (aJ)	Relays/unit	Total Relays	Total E_{op} (fJ)
Full Adder	119	37.1	7	833	4.410
Full Adder (with additional P, G, K, \bar{P} , \bar{G} terms)	300	37.1	7	2100	11.100
AND2	3	4.8	2	6	0.014
DEC23	4	55.8	46	184	0.223
DEC7	76	19.7	14	1064	1.490
DEC8	6	23.3	16	96	0.140
ExpResultIsClogic	2	8.8	24	48	0.018
InfABLogic	1	11.7	6	6	0.012
NOT (and BUF)	21	12.0	2	42	0.253
LZD16 (Tree Design)	5	44.3	116	580	0.222
ManResultLogic1	27	88.7	12	324	2.390
ManResultLogic2	1	115.5	16	16	0.116
Multiplier (32-bit, 5 t_{mech} , Booth-encoded)	1	69301.8	9272	9272	69.302
Dual 2:1 MUX	6567	29.1	2	13134	191.000
NanABLogic	1	58.2	16	16	0.058
NanZLogic	1	127.0	18	18	0.127
OR2	91	9.5	2	182	0.865
RoundLogic	1	216.3	22	22	0.216
SignResultLogic	1	77.6	12	12	0.078
XOR3	80	15.3	2	160	1.220
Grand Total				28115	283

Table 6.3. Energy per operation breakdown by circuit block in the FMA. The 90nm equivalent 6T NEM relays process is used and $V_{DD} = 1.01V_{pi}$. The energy per operation comes from the post-route switching power estimate for each circuit cell.

not based the complete FMA activity picture. This is a reasonable approximation.

To analyze how much of the E_{op} of each circuit is dissipated in interconnects, the distribution of the ratio of pre-route switching power to post-route switching power is observed. It turns out that this distribution is mostly uniform, since the ratio has an average of 45.4% and a standard deviation of 5.9%. This means that slightly over half of the E_{op} for every circuit cell goes into the energy dissipated in interconnects, in the 90nm equivalent 6T NEM relays process node.

Conclusions

CIRCUIT design with NEM relays is what enables the zero-leakage devices to deliver on reducing the power consumption of computing, beyond CMOS. In the case of the FPU, the NEM relays pass-transistor circuit design paradigm must be applied to realize a NEM relay FPU that is competitive in throughput compared to an optimal CMOS FPU and an order of magnitude better in energy-efficiency. In the process, several novel custom pass-transistor circuits are designed, including several flavors of the LZD. In this manner, it is seen that the NEM relay adder, multiplier, and now the FPU can be more energy-efficient compared to their optimal CMOS counterparts [1, 5].

The FPU designed and analyzed in this work is the largest system designed with NEM relays. The latency of the NEM relay FMA is reduced from 71 mechanical delays in a CMOS-style design to 16 mechanical delays in a NEM relay pass-logic-style design, as shown in Fig. 5.1. The energy analysis in this work shows that the energy-efficiency benefits of NEM relays extend into full VLSI systems.

■ 7.1 Future Directions

Results from Chapter 3 show that many device enhancements are necessary to realize the large and very-large scale circuits designed with NEM relays.

The reliability of NEM relays high relative to other emergent devices, considering the high yield. However, process variations still exist that need to be modeled and stabilized so that the operating voltages and external resistors are appropriate for circuit operation.

The number of mechanical delays of the FMA designed with NEM relays could be reduced even further if LZDs with more than 16 input bits are designed to run in a single mechanical delay. However, this provides diminishing returns because, out of 16 mechanical delays for the FMA, reducing the LZD number of mechanical delays can be reduced down to 13, which is not a significant improvement compared to the vast improvements already made shown in Fig. 5.1.

The maximum datapath bitwidth for the CMA is smaller than that for the FMA, meaning that a smaller shifter is required and that the LZD is smaller, requiring fewer mechanical delays. Hence the design of a CMA with NEM relays is also worthy of

exploration.

Although a single-precision FPU is explored in this work, the design trade-offs revealed also apply to other designs such as that of a double-precision FPU. One potential difference with the double-precision FPU is that the LZD may contribute even more to the critical path of the FPU arithmetic blocks due to the larger bitwidths, highlighting the importance of reducing the number of mechanical delays of the LZD.

Within the domain of circuits, future designs of graphics processor units (GPUs) and central processing units (CPUs) with NEM relays may be of interest. NEM relays may also find application in more elementary circuits such as crossbars or programmable logic arrays (PLAs).

The fabrication of seesaw NEM relays needs to continue to advance while better device models for them also need to be devised.

With the continued scaling of the NEM relays process technology and advancement of the systems integration, the fabric of future computers may one day be NEM relays.

Bibliography

- [1] F. Chen, Hei Kam, D. Markovic, T.-J. King Liu, V. Stojanovic, and E. Alon. Integrated circuit design with nem relays. In *Computer-Aided Design, 2008. ICCAD 2008. IEEE/ACM International Conference on*, pages 750–757, 2008. doi: 10.1109/ICCAD.2008.4681660.
- [2] M. Spencer, F. Chen, C.C. Wang, R. Nathanael, H. Fariborzi, A. Gupta, Hei Kam, V. Pott, J. Jeon, T.-J. King Liu, D. Markovic, E. Alon, and V. Stojanovic. Demonstration of integrated micro-electro-mechanical relay circuits for VLSI applications. *Solid-State Circuits, IEEE Journal of*, 46(1):308–320, 2011. ISSN 0018-9200. doi: 10.1109/JSSC.2010.2074370.
- [3] B.H. Calhoun, A. Wang, and A. Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *Solid-State Circuits, IEEE Journal of*, 40(9):1778–1786, 2005. ISSN 0018-9200. doi: 10.1109/JSSC.2005.852162.
- [4] R. Nathanael, V. Pott, Hei Kam, J. Jeon, and T.-J. King Liu. 4-terminal relay technology for complementary logic. In *Electron Devices Meeting (IEDM), 2009 IEEE International*, pages 1–4, 2009. doi: 10.1109/IEDM.2009.5424383.
- [5] H. Fariborzi, F. Chen, V. Stojanovic, R. Nathanael, J. Jeon, and T.-J. King Liu. Design and demonstration of micro-electro-mechanical relay multipliers. In *Solid State Circuits Conference (A-SSCC), 2011 IEEE Asian*, pages 117–120, 2011. doi: 10.1109/ASSCC.2011.6123618.
- [6] H. Goldstine. *The Computer: from Pascal to von Neumann*. Princeton University Press, Princeton, New Jersey, 1972.
- [7] Hei Kam, V. Pott, R. Nathanael, J. Jeon, E. Alon, and T.-J. King Liu. Design and reliability of a micro-relay technology for zero-standby-power digital logic applications. In *Electron Devices Meeting (IEDM), 2009 IEEE International*, pages 1–4, 2009. doi: 10.1109/IEDM.2009.5424218.
- [8] J. Jeon, V. Pott, Hei Kam, R. Nathanael, E. Alon, and T.-J. King Liu. Seesaw relay logic and memory circuits. *Microelectromechanical Systems, Journal of*, 19(4):1012–1014, 2010. ISSN 1057-7157. doi: 10.1109/JMEMS.2010.2049826.

- [9] I. Chen, L. Hutin, C. Park, R. Lee, R. Nathanael, J. Yaung, J. Jeon, and T.-J. King Liu. Scaled micro-relay structure with low strain gradient for reduced operating voltage. *ECS Transactions*, 45(6):101–106, 2012. doi: 10.1149/1.3700943.
- [10] V. Pott, Hei Kam, R. Nathanael, J. Jeon, E. Alon, and T.-J. King Liu. Mechanical computing redux: Relays for integrated circuit applications. *Proceedings of the IEEE*, 98(12):2076–2094, 2010. ISSN 0018-9219. doi: 10.1109/JPROC.2010.2063411.
- [11] H. Fariborzi. *Design and Demonstration of Integrated Micro-Electro-Mechanical Relay Circuits for VLSI Applications*. PhD thesis, Massachusetts Institute of Technology, 2013.
- [12] R. Nathanael, V. Pott, Hei Kam, J. Jeon, E. Alon, and T.-J. King Liu. Four-terminal-relay body-biasing schemes for complementary logic circuits. *Electron Device Letters, IEEE*, 31(8):890–892, 2010. ISSN 0741-3106. doi: 10.1109/LED.2010.2050133.
- [13] Hei Kam, T.-J. King Liu, V. Stojanovic, D. Markovic, and E. Alon. Design, optimization, and scaling of MEM relays for ultra-low-power digital logic. *Electron Devices, IEEE Transactions on*, 58(1):236–250, 2011. ISSN 0018-9383. doi: 10.1109/TED.2010.2082545.
- [14] Hei Kam, E. Alon, and T.-J. King Liu. A predictive contact reliability model for mem logic switches. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 16.4.1–16.4.4, 2010. doi: 10.1109/IEDM.2010.5703375.
- [15] T.-J. King Liu, J. Jeon, R. Nathanael, Hei Kam, V. Pott, and E. Alon. Prospects for mem logic switch technology. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 18.3.1–18.3.4, 2010. doi: 10.1109/IEDM.2010.5703386.
- [16] J. Jeon, V. Pott, Hei Kam, R. Nathanael, E. Alon, and T.-J. King Liu. Perfectly complementary relay design for digital logic applications. *Electron Device Letters, IEEE*, 31(4):371–373, 2010. ISSN 0741-3106. doi: 10.1109/LED.2009.2039916.
- [17] M. Jiang. BDD-based logic synthesis of MEM relay circuits. Master’s thesis, University of California, Los Angeles, 2013.
- [18] F. Chen. *Energy-efficient Wireless Sensors: Fewer Bits, Moore MEMS*. PhD thesis, Massachusetts Institute of Technology, 2011.
- [19] F. Chen, M. Spencer, R. Nathanael, Chengcheng Wang, H. Fariborzi, A. Gupta, Hei Kam, V. Pott, J. Jeon, T.-J. King Liu, D. Markovic, V. Stojanovic, and E. Alon. Demonstration of integrated micro-electro-mechanical switch circuits for vlsi applications. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, pages 150–151, 2010. doi: 10.1109/ISSCC.2010.5434010.

- [20] D. A. Patterson and J. L. Hennessy. *Computer Organization and Design: The Hardware/Software Interface, 3rd Edition*. Elsevier, 2005.
- [21] M.J. Flynn, K. Nowka, G. Bewick, E. Schwarz, and N. Quach. The SNAP project: towards sub-nanosecond arithmetic. In *Computer Arithmetic, 1995., Proceedings of the 12th Symposium on*, pages 75–82, 1995. doi: 10.1109/ARITH.1995.465374.
- [22] H. Suzuki, H. Morinaka, H. Makino, Y. Nakase, K. Mashiko, and T. Sumi. Leading-zero anticipatory logic for high-speed floating point addition. *Solid-State Circuits, IEEE Journal of*, 31(8):1157–1164, 1996. ISSN 0018-9200. doi: 10.1109/4.508263.
- [23] S.F. Oberman and M. Flynn. Division algorithms and implementations. *Computers, IEEE Transactions on*, 46(8):833–854, 1997. ISSN 0018-9340. doi: 10.1109/12.609274.
- [24] S. Galal and M. Horowitz. Energy-efficient floating-point unit design. *Computers, IEEE Transactions on*, 60(7):913–922, 2011. ISSN 0018-9340. doi: 10.1109/TC.2010.121.
- [25] E. M. Schwarz. Binary floating-point unit design. In V. G. Oklobdzija and R. K. Krishnamurthy, editors, *High-Performance Energy-Efficient Microprocessor Design*, Series on Integrated Circuits and Systems, pages 189–208. Springer US, 2006. ISBN 978-0-387-28594-8. doi: 10.1007/978-0-387-34047-0_8.
- [26] R.M. Jessani and C. H. Olson. The floating-point unit of the powerpc 603e microprocessor. *IBM Journal of Research and Development*, 40(5):559–566, 1996. ISSN 0018-8646. doi: 10.1147/rd.405.0559.
- [27] R. Zimmermann and W. Fichtner. Low-power logic styles: Cmos versus pass-transistor logic. *Solid-State Circuits, IEEE Journal of*, 32(7):1079–1090, 1997. ISSN 0018-9200. doi: 10.1109/4.597298.
- [28] S. Galal, O. Shacham, J.S. Brunhaver, Jing Pu, A. Vassiliev, and M. Horowitz. FPU generator for design space exploration. In *Computer Arithmetic (ARITH), 2013 21st IEEE Symposium on*, pages 25–34, 2013. doi: 10.1109/ARITH.2013.27.
- [29] O. Shacham, S. Galal, S. Sankaranarayanan, M. Wachs, J. Brunhaver, A. Vassiliev, M. Horowitz, A. Danowitz, W. Qadeer, and S. Richardson. Avoiding game over: Bringing design to the next level. In *Design Automation Conference (DAC), 2012 49th ACM/EDAC/IEEE*, pages 623–629, 2012.
- [30] K. Dwan. Logic synthesis of MEM relay circuits. Master’s thesis, University of California, Los Angeles, 2012.
- [31] V.G. Oklobdzija. An implementation algorithm and design of a novel leading zero detector circuit. In *Signals, Systems and Computers, 1992. 1992 Conference*

-
- Record of The Twenty-Sixth Asilomar Conference on*, pages 391–395 vol.1, 1992. doi: 10.1109/ACSSC.1992.269243.
- [32] V.G. Oklobdzija. An algorithmic and novel design of a leading zero detector circuit: comparison with logic synthesis. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 2(1):124–128, 1994. ISSN 1063-8210. doi: 10.1109/92.273153.
- [33] R. Nathanael. *Nano-Electro-Mechanical (NEM) Relay Devices and Technology for Ultra-Low Energy Digital Integrated Circuits*. PhD thesis, University of California, Berkeley, 2012.