μSystems Research Group

School of Electrical and Electronic Engineering

NON-INVASIVE POWER GATING TECHNIQUES FOR BURSTY COMPUTATION WORKLOADS USING MICRO-ELECTRO-MECHANICAL RELAYS

Haider M. Alrudainy

Technical Report Series

NCL-EEE-MICRO-TR-2017-207

July 2017

DECLARATION

I hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award. Where other sources of information have been used, they have been acknowledged.

Newcastle upon Tyne July 2017

Haider Alrudainy

CERTIFICATE OF APPROVAL

I confirm that, to the best of my knowledge, this thesis is from the student's own work and effort, and all other sources of information used have been acknowledged. This thesis has been submitted with my approval.

ALEX YAKOVLEV

To the soul of my wonderful Mother, and my lovely family.
— Haider

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisors Prof. Alex Yakovlev, Dr. Andrey Mokhov for their support and guidance through my PhD journey. They have always been a source of motivation and my inspirational model as a researcher.

I am grateful to the higher committee of education development in Iraq (HCED) for funding my PhD study their scholarship programme and also would like to express my sincere gratitude to the Iraqi Cultural Attache in London.

As part of my work, It is required establishing application of MEM relays in real applications, such as many-core systems. During this time, I worked with the Newcastle team of PRiME research project (which is a multi-million programme grant funded by EPSRC, UK). Thus, I would like to acknowledge the support I received from the PRiME project team.

I would like also to express my gratefulness and appreciation to my colleagues and friends in the School of Electrical and Electronic Engineering, especially those in Microelectronic Systems research (MSD) group, at Newcastle University for their guidance and assistance through my study. Especially, I appreciate the support of my wonderful colleagues Alaa Al-shkarchi, Yasir Abdullah, Dr. Nizar Dahir, Dr. Ra'ed Aldujaily, Dr. Hussein Leftah, and Ali Majeed for their productive suggestions, fruitful discussions, and subjective criticism.

I am also grateful to Dr Matthew Spencer for his advice in modifying and build the switch model simulator to include the more physical feature into simulating the MEMS.

I would like to offer my special regards to all the staff of the school of Electrical and Electronic Engineering in Newcastle university, especially Dr. Rishad Shafik, and Dr Fei Xia.

last but not least, I would like to thank my wonderful family for their continuous support and motivation throughout my PhD journey.

ABSTRACT

Electrostatically-actuated Micro-Electro-Mechanical/Nano-Electro- Mechanical (MEM/NEM) relays are promising devices overcoming the energy-efficiency limitations of CMOS transistors. Many exploratory research projects are currently under way investigating the mechanical, electrical and logical characteristics of MEM/NEM relays. One particular issue that this work addresses is the need for a scalable and accurate physical model of the MEM/NEM switches that can be plugged into the standard EDA software.

The existing models are accurate and detailed but they suffer from the convergence problem. This problem requires finding ad-hoc workarounds and significantly impacts the designer's productivity. In this thesis we propose a new simplified Verilog-AMS model. To test scalability of the proposed model we cross-checked it against our analysis of a range of benchmark circuits. Results show that, compared to standard models, the proposed model is sufficiently accurate with an average of 6% error and can handle larger designs without divergence.

This thesis also investigates the modelling, designing and optimization of various MEM/NEM switches using 3D Finite Element Analysis (FEA) performed by the COMSOL multiphysics simulation tool. An extensive parametric sweep simulation is performed to study the energy-latency trade-offs of MEM/NEM relays. To accurately simulate MEMS/NEMS-based digital circuits, a Verilog-AMS model is proposed based on the evaluated parameters obtained from the multiphysics simulation tool. This allows an accurate calibration of the MEM/NEM relays with a significant reduction in simulation speed compared to that of 3D FEA exercised on COMSOL tool.

The effectiveness of two power gating approaches in asynchronous micropipelines is also investigated using MEM/NEM switches and sleep transistors in reducing idle power dissipation with a particular target throughput. Sleep transistors are traditionally used to power gate idle circuits, however, these transistors have fundamental limitations in their effectiveness. Alternatively, MEM/NEM relays with zero leakage current can achieve greater energy savings under a certain data rate and design architecture. An asynchronous FIR filter 4 phase bundled data handshake protocol is presented. Implementation is accomplished in 90nm technology node and simulation exercised at various data rates and design complexities. It was demonstrated that our proposed approach offers 69% energy improvements at a data rate 1KHz compared to 39% of the previous work.

The current trends for greater heterogeneity in future Systems-on-Chip (SoC) do not only concern their functionality but also their timing

and power aspects. The increasing diversity of timing and power supply conditions, and associated concurrently operating modes, within an SoC calls for more efficient power delivery networks (PDN) for battery operated devices. This is especially important for systems with mixed duty cycling, where some parts are required to work regularly with low-throughput while other parts are activated spontaneously, i.e. in bursts. To improve their reaction time vs energy efficiency, this work proposes to incorporate a power-switching network based on MEM relays to switch the SoC power-performance state (PPS) into an active mode while eliminating the leakage current when it is idle. Results show that even with today's large and high pull-in voltages, a MEM-relay-based power switching network (PSN) can achieve a 1000x savings in energy compared to its CMOS counterpart for low duty cycle. A simple case of optimising an on-chip charge pump required to switch-on the relay has been investigated and its energy-latency overhead has been evaluated.

Heterogeneous many-core systems are increasingly being employed in modern embedded platforms for high throughput at low energy cost considerations. These applications typically exhibit bursty workloads that provide opportunities to minimize system energy. CMOS-based power gating circuitry, typically consisting of sleep transistors, is used as an effective technique for idle energy reduction in such applications. However, these transistors contribute high leakage current when driving large capacitive loads, making effective energy minimization challenging.

This thesis proposes a novel MEMS-based idle energy control approach. Core to this approach is an integrated sleep mode management based on the performance-energy states and bursty workloads indicated by the performance counters. A number of PARSEC benchmark applications are used as case studies of bursty workloads, including CPU- and memory- intensive ones. These applications are exercised on an Exynos 5422 heterogeneous many-core platform, engineered with a performance counter facilities, showing 55.5% energy savings compared with an on-demand governor. Furthermore, an extensive trade-off analysis demonstrates the comparative advantages of the MEMS-based controller, including zero-leakage current and non-invasive implementations suitable for commercial off-the-shelf systems.

PUBLICATIONS

Journal and magazines publications:

1. **Haider M. Alrudainy**; A. Aalsaud; R. Shafik; A. Mokhov; F. Xia; A. Yakovlev, *Understanding Idle Energy Minimization for Bursty Computation Workloads in Heterogeneous Many-Core Systems using Micro-Electro-Mechanical Relays*, (to be submitted), IEEE Transactions on Circuits and Systems II, TCAS-II

2. **Haider M. Alrudainy**; A. Mokhov; A. Yakovlev, *A Survey of Emerging Applications Utilizing Miro-Electro-Mechanical Relays for Energy and Reliability Efficient Digital Circuits*, (submitted), IEEE Circuits and Systems Magazine

Conference publications:

1. **Haider M. Alrudainy**; A. Mokhov; A. Yakovlev , *A Scalable Physical Model for Nano-Electro-Mechanical Relays*, Power and Timing Modeling, Optimization and Simulation (PATMOS), 29 Sept.-1 Oct. 2014, doi: 10.1109/PATMOS.2014.6951889, ISBN:978-1-4799-5412-4

2. **Haider M. Alrudainy**; A. Mokhov; N. Dahir; A. Yakovlev, *MEMS-Based Power Delivery Control for Bursty Applications*, Circuits and Systems (ISCAS), 2016 IEEE International Symposium on, 22-25 May 2016, pp 790-793, March 2015, DOI: 10.1109/ISCAS.2016.7527359, ISBN: 978-1-4799-5341-7

3. **Haider M. Alrudainy**; A. Aalsaud; R. Shafik; A. Mokhov; F. Xia; A. Yakovlev, *MEMS-Based Runtime Idle Energy Minimization for Bursty Workloads in Heterogeneous Many-Core Systems*, (Submitted), Design, Automation Conference (DAC), 18-22 June 2017

4. **Haider M. Alrudainy**; A. Mokhov; F. Xia; A. Yakovlev,*Ultra-low energy data driven computing using asynchronous micropipelines and nano-electro-mechanical relays*, in ISVLSI, pp. 1-6, July 2017 accepted.

5. **Haider M. Alrudainy**; R. Shafik; A. Mokhov; A. Yakovlev,*Lifetime Reliability Characterization of Nano/Micro -Electro-Mechanical Switches Used in Power Gating of Digital Integrated Circuits*, in DFT, pp. 1-6, October 2017 submitted.

Workshop and forum publications:

1. **Haider M. Alrudainy**, A. Mokhov; N. Dahir; A. Yakovlev, *A Scalable Physical Model for Nano-Electro-Mechanical Relays Based Finite Element Analysis (FEA) Parameters Verification*, in the proceedings of DAC, The 51st Annual Design Automation Conference, June 1-5, 2014, San Francisco, CA, USA, Work in Progress poster.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

ACRONYMS

ZDRTO zero delay ripple turn on

DL delay line

MEMS Micro-Electro-Mechanical Switch

MEM Micro-Electro-Mechanical

NEMS Nano-Electro-Mechanical Switch

NEM Nano-Electro-Mechanical

PSC power switch controller

DSP digital signal processing

PDN power delivery network

VLSI very large scale integrate circuit

SoC system-on-Chip

CC combinational circuit

PCN power conversion network

VRM voltage regulator module

RFIDs radio-frequency identifications

EDA electronic design automation

PPS power performance state

FEA finite element analysis

ICs integrated circuits

FB function block

FBs function blocks

CMOS complementary metal-oxide-semiconductor

MEM/NEM Micro-Electro-Mechanical/Nano-Electro-Mechanical

MEMS/NEMS Micro-Electro-Mechanical Switch/Nano-Electro-Mechanical Switch

DSP	digital signal processing
DVFS	dynamic voltage and frequency scaling
DVS	dynamic voltage scaling
PSN	power switch network
ALE	Arbitrary Lagrangian-Eulerian
Q	quality factor
ICs	integrated circuits
IOTs	Internet of Things
WSNs	Wireless Sensor Networks
PDN	power delivery network
SoC	system-on-chip
PVT	process-voltage-temperature
VLSI	very large scale integration

Part I

Thesis Chapters

1

INTRODUCTION

1.1 MOTIVATION

During the last four decades, scaling of the complementary metal-oxide-semiconductor (CMOS) has been the impetus in progressing of microelectronics industry enabling significant improvement in terms of performance, integration, and cost. This is attributed to the fact that higher performance and lower cost are two primary goals for many digital integrated circuits (ICs) [127]. Dynamic energy, historically, has been dominating the energy consumption of digital circuits and has been the central focus of energy reduction research for many years. However, continuing technology scaling has significantly caused an increase in leakage energy consumption, as shown in Fig. 1.1, demanding a variety of leakage reduction techniques to be developed. The mainstream of future electronics design in some emerging applications is to shift from performance-driven goals to energy constraints, and as the leakage energy continues to grow exponentially, this presents a fundamental challenge in achieving these targets.

From a system level point of view, dynamic and leakage energy reduction is a major design goal of modern embedded systems. Dynamic energy can be lowered either by reducing the work required to accomplish a task, or by adopting dynamic voltage and frequency scaling (DVFS) to run the task at lower frequency and voltage. Leakage energy, however, is not related to the circuit activity, but is more closely linked to the circuit's technology, operating temperature, and chip layout. Therefore, the only way to reduce it is by removing the supply voltage during idle periods.

According to a recent study presented in [46, 167], a large proportion of a chip has to be shut-off so that it can still operate within its given power budget due to the "dark silicon" phenomenon. This "dark silicon" issue becomes more obvious at 8nm, and 22nm technology node, where roughly 50%, and 21%, respectively, of the chip may have to be power gated at any given time while. Furthermore, according to the International Technology Roadmap for Semiconductor (ITRS), by the year 2020, the predicted chip power consumption for modern SoCs will have risen by a factor of roughly $\times 10$ compared to 2012 [21]. This leads to mandatory power gating of significant parts of the future many-core platforms.

One of the most prominent opportunities to mitigate leakage energy is during standby periods, when no computation occurs. The microprocessor industry has widely adopted power gating techniques

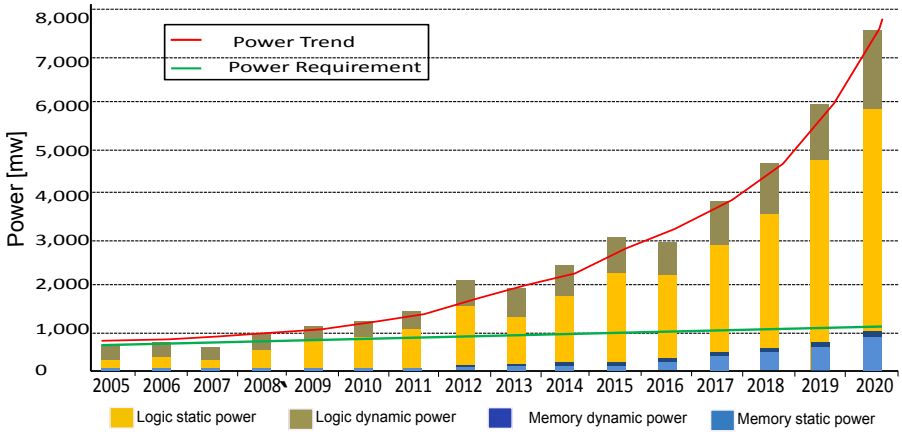


Figure 1.1: IC power trends until 2020: actual vs. specified (Source: International Technology Roadmap for Semiconductors (http://www.itrs.net/).

coupled with software-controlled sleep modes in such applications exhibiting substantial idle periods [26]. With this technique, transistors are used to disconnect the power from unused portions of a microprocessor, which reduces leakage power. This approach is attractive because it mitigates leakage without requiring any modification to the logic or operation of the power-gated circuitry. However, sleep transistors themselves contribute high leakage current, especially when driving a large capacitive load. This thesis, therefore, proposes a new power gating techniques based on the Micro-Electro-Mechanical Switch/Nano-Electro-Mechanical Switch (MEMS/NEMS) to completely eliminate leakage energy dissipation in applications whose workloads are bursty in nature.

The first chapter provides an overview of low energy designs in digital circuits and gives preliminary details of the subsequent chapters in the thesis. The major components of energy dissipation in digital circuits are discussed in Section 1.1.1. A synoptic outline of the established dynamic energy techniques is described in Section 1.1.2. Section 1.1.3 summarizes the impact of technology scaling on the leakage energy dissipation and Section 1.1.4 discusses some effective techniques employed for energy-efficient digital circuits which are relevant to the work reported in this thesis. Section 1.1.5 provides examples of energy-constrained applications where low-throughput microprocessors are used. Furthermore, a case of energy-efficient microprocessor architecture is discussed where two modes of operations are postulated as follows: firstly, high throughput coupled with burst activities; and secondly, rationally low throughput coupled with regular activities. Finally, the contributions of each subsequent chapter are outlined in Section 1.2.

1.1.1 *Energy in digital circuits*

In order to design energy-efficient digital CMOS circuits, it is essential to understand the main sources of energy consumption. Typically, these can be mainly divided into dynamic energy and leakage energy [140]. Dynamic energy dissipation is caused by switching activity and occurs whenever the digital circuit is doing useful work. Leakage energy, whereas, is consumed whenever the digital circuits remain connected to the supply voltage regardless of whether or not a useful computation is being executed. This means that when the digital circuits are performing a useful work i.e. ***active mode*** of operation, total energy is presented by both dynamic and leakage energy. However, when the digital circuits are not performing useful computation i.e. ***Idle mode*** of operation, the leakage energy is the only contributor of energy consumption. The total energy dissipation of digital circuits, therefore, is given by:

$$E_{total} = E_{dyn} + E_{leak} \quad . \tag{1.1}$$

The total energy dissipation in digital CMOS circuits over a number of cycles N_{cycle} with clock period t_{period} can be expressed as follows:

$$E_{total} = E_{dyn} + E_{leak} = N_{cycle}.V_{dd}^2.C_{eff} + N_{cycle}.V_{dd}.I_{leak}.t_{period} \quad , \tag{1.2}$$

where C_{eff} represents the summed average capacitance given by the product of the switching probability α and total load capacitance C_L [140]. Assuming the clock frequency is f, the total switching power of digital CMOS circuits can be then given as:

$$P_{total} = P_{dyn} + P_{leak} = V_{dd}^2.C_{eff}.f + V_{dd}.I_{leak} \quad . \tag{1.3}$$

It can be seen from Eqs. 1.2 and 1.3 that dynamic power is proportional to the product of the square supply voltage V_{dd} and clock frequency, whereas dynamic energy is only proportional to the square of supply voltage V_{dd}. Consequently, although reducing clock frequency can reduce dynamic power, the required dynamic energy to complete the task remains the same. A lower operating frequency, however, would result in an increase in total energy consumption because of the increasing clock period in Eq.1.2. As a result, more energy is consumed as leakage, leading to lower energy-efficiency.

Traditionally, scaling down the supply voltage (V_{dd}) can drastically lower the energy used per operation due to the quadratic dependence of dynamic energy consumption on V_{dd}, as can be seen in Eq. 1.2. However, reducing the supply voltage (V_{dd}) increases propagation de-

lay time in digital circuits, which in turn causes a decrease in operating frequency. Consequently, further reduction in supply voltage, passing the $V_{optimum}$, can result in an increase in total energy consumption due to the dominance of leakage energy. This is attributed to the exponential increment of propagation delay time for $V_{dd} < V_{optimum}$ [146]. The propagation delay time of a gate can be roughly estimated as [4]:

$$T_{propagation} \propto \frac{C_L.V_{dd}}{(V_{dd} - V_{th})^n} \quad , \tag{1.4}$$

where n is a technology-dependent parameter used to model short channel effects and ordinarily has a value ranging between 1-2, and V_{th} represents the threshold voltage of the transistors.

1.1.2 *Dynamic energy minimization*

Dynamic energy has dominated the total energy dissipation of digital CMOS circuits for many years. Examining Eq. 1.2, it can be seen that dynamic energy can be lowered by reducing either the effective load capacitance C_{eff}, or the operating voltage (V_{dd}). However, lowering V_{dd} increases the propagation delay, Eq. 1.4, and hence methods to reduce dynamic energy consumption attempt to affect these variables without affecting any performance degradation. In this section, a brief summary of some extensively adopted dynamic energy reduction techniques is presented.

1.1.2.1 *Clock gating*

The dynamic energy dissipation of a processor is mostly dominated by the clock tree [143]. It has been noted that almost 32% of the dynamic power in the Alpha 21264 microprocessor is consumed because of the global clock network [53]. This relatively sizeable dynamic power dissipation is due to the large capacitive load of the clock tree as well as high activity. Furthermore, although a non-trivial proportion of registers may hold the same logic state over several clock cycles, the internal switching of the gate from the toggling of the clock adds to the total dynamic power dissipation. Consequently, this observation led to the development of the clock gating technique, which was presented to prevent the switching of parts of the clock tree where registers state did not demand to be updated. The simplicity of design and ease of implementation of clock gating allows it to be used in all types of sequential digital circuits. It has been reported that 33% and 50% savings in dynamic energy can be achieved by implementing clock gating techniques in microprocessors [52] and FPGAs [70] respectively.

1.1.2.2 *Dynamic Voltage and Frequency Scaling (DVFS)*

Dynamic Voltage and Frequency Scaling (DVFS) targets both the operating clock frequency and the supply voltage (V_{dd}). This technique can be used when a system is not required to operate at its maximum frequency due to workload demands [51]. Therefore, the supply voltage can be reduced to significantly decrease dynamic energy and improve energy-efficiency, Eq. 1.2, resulting in lowering the operating frequency due to the increment of propagation delay, Eq. 1.4. It has been reported that the energy-efficiency of a processor can be improved by up to $\times 10$ when employing DVFS [51]. The advantage of DVFS is the capability to switch between low performance, low energy and high performance, high energy states relying on the current workload. As a result, its efficiency at minimizing dynamic energy consumption has prompted many runtime control algorithms in a variety of different systems[3, 15, 172].

An alternative to DVFS is to employ multi-supply/multi-voltage operation, where different sections of the system-on-chip (SoC) may require various levels of performance state. By utilizing multiple supplies, voltage islands can be formed to partition subsections of the SoC that are of the critical path, enabling dynamic energy savings [145, 168]. This means that the chosen path has to cope with rises in delay time correlated with the lowered supply voltage without hindering overall system performance. This method is most prevalent in SoCs that demand caches to be run as quickly as possible whereas CPUs and the rest of the SoC can operate at a decreased voltage while still satisfying the timing constraints of the overall system.

It is also possible to design systems with dynamic reconfiguration, which allows the choice between high-performance and energy-efficient components in runtime depending on particular application [110], or alternatively trade energy for parallelism [169].

Others have proposed constructing a digital circuit that can dynamically tune V_{th} and V_{dd} to track the optimum operating voltages accordingly as the activity factor of the circuit is changed [130]. However, the actual implementation of such a system is extremely challenging because tracking the activity factor in an energy-efficient way is a difficult task. Furthermore, employing body-biasing regions and multiple power rails introduces significant area overheads. Moreover, the body-biasing technique is significantly less effective in newer CMOS technologies due to higher doping concentrations [134].

Asynchronous circuits can provide even higher flexibility in terms of the operating voltage and effective frequency. For example, a new methodology of building adaptive asynchronous controller to support a wide range of power and timing modes of operation which can be chosen during run-time has recently been proposed in [109].

1.1.3 *Scaling Implications on Power Dissipation*

Over the last four decades, the semiconductor industry has followed a trend in CMOS scaling where the geometric dimensions of devices have been reduced by about 30% every 2-3 years [20, 61]. The foremost purpose behind this scaling trend is to decrease the cost of ICs fabrication [67]. With a 30% reduction in process size, area decreases by 50% (0.7×0.7=0.49) allows the doubling of number of CMOS transistors in the same silicon wafer die. Furthermore, a 30% reduction in device geometry drives to a 30% decrease in gate delay, resulting in an improvement of $\frac{1}{0.7}$ in digital circuits performance [20]. Scaling the device dimensions and reducing supply voltage by 30% leads to an improvement in dynamic power consumption by 50% with each new CMOS generation, according to the "Dennard Scaling Law" [41]. This can be shown by substituting the decreased supply voltage and geometry into:

$$\begin{aligned} P_{dyn}(new) &= (0.7 \times C_{eff}(old)) \times (\frac{1}{0.7} \times f(old)) \times (0.7 \times V_{dd}(old))^2 \\ &\approx 0.5 \times P_{dyn}(old) \quad . \end{aligned} \tag{1.5}$$

The scaling of CMOS transistors has thus resulted in faster, smaller devices, and lower dynamic power dissipation. However, with each new generation of CMOS technology, the sub-threshold leakage current is increased, and this happens because of the demand to decrease the threshold voltage (V_{th}) as a component of technology scaling [42, 140].

The continuing CMOS scaling trends predicted by the ITRS shows that the aggressive scaling of transistors to improve performance, reduce cost, and increase integration will continue to have a positive effect on dynamic power dissipation but at the expense of increased leakage power dissipation [141]. Accordingly, whilst leakage power consumption is already a problem now, it will become an even more pressing concern for future designs [89].

1.1.4 *Leakage energy reduction*

Given the increasing predominance of leakage energy consumption in digital CMOS circuits, a plethora of studies are being conducted into decreasing it within these circuits. There are various techniques for overcoming leakage energy dissipation within a microprocessor as reported in the following Sections. Power gating is considered to be the most practical and effective technique to combat leakage [4, 90, 152] and is studied in Chapters 4, 5, and 6.

1.1.4.1 *Power gating*

Power gating is a leakage minimization method that has obtained increasing attention over the last decade and is commonly used in many circuits, and at the micro architecture level [4, 90, 152]. Furthermore, it has recently been adopted in many system level designs such as Intel Core i7 microprocessor [26], and Nvidia Tegra microprocessors which is based on the ARM Cortex-A9. The substantial aim of power gating is to allow two modes of operation: an ***active mode*** in which a digital CMOS circuit can continue computing as normal, and a ***sleep mode*** in which the digital circuit can enter an idle state. Commonly, power gating facilitates the low leakage ***sleep mode*** by cutting off the supply voltage to the digital circuits and therefore can be termed as "shut-down power gating".

Fig. 1.2 describes a conceptual overview of the execution time and power dissipation of the CPU-intensive *ferret* application, where 4 LITTLE cores are fully operated with only one out of the four big cores kept active. This application is executed on the Exynos 5422 big.LITTLE octa-core heterogeneous platform. As the figure shows, the idle power of the big cores is experimentally measured as a function of operating frequency, as explained further in Chapter 6. In this experiment, three tasks separated by a period of time are observed. Firstly, no application is launched where only 0.2 watts are consumed by the system as indicated in "Task1". Secondly, the application is launched without executing, supply voltage and clock are remain operational, where the idle power consumed is a function of operating frequency and ranges from 1, 0.4, and 0.2 watt at 2000, 1400, and 200 MHz respectively. Finally, the application is instantiated and executed on core 7, which is the A15 big core type, as indicated in "Task3". In this case, the total power consumption ranges from 3.5, 1.9, and 0.3 watt at 2000, 1400, and 200 MHz respectively. Therefore, power gating capitalizes on these idle periods to cut-off the supply voltage to the idle processor, and thereby leakage power dissipation in a heterogeneous many-core systems is reduced. When the idle period begins, a sleep signal is generated by the power gating controller which enables the power gating circuitry to cut off the supply voltage and place the processor into a low leakage state referred to as the *sleep mode* of operation.

Fig. 1.3 illustrates a conceptual view of how power gating technique can be implemented in synchronous digital circuits. Normally, a power switch network (PSN), also referred to as sleep switches, is set in series with a computational logic block and provide power from the supply voltage to the entire block. Typically, these switches are based on CMOS transistors which comprise either PMOS power transistors referred to as *header* or NMOS power transistors referred to as *footer*. In this thesis, however, an emerging devices referred to as MEMS/NEMS have been adopted to target applications with low duty cycles mode

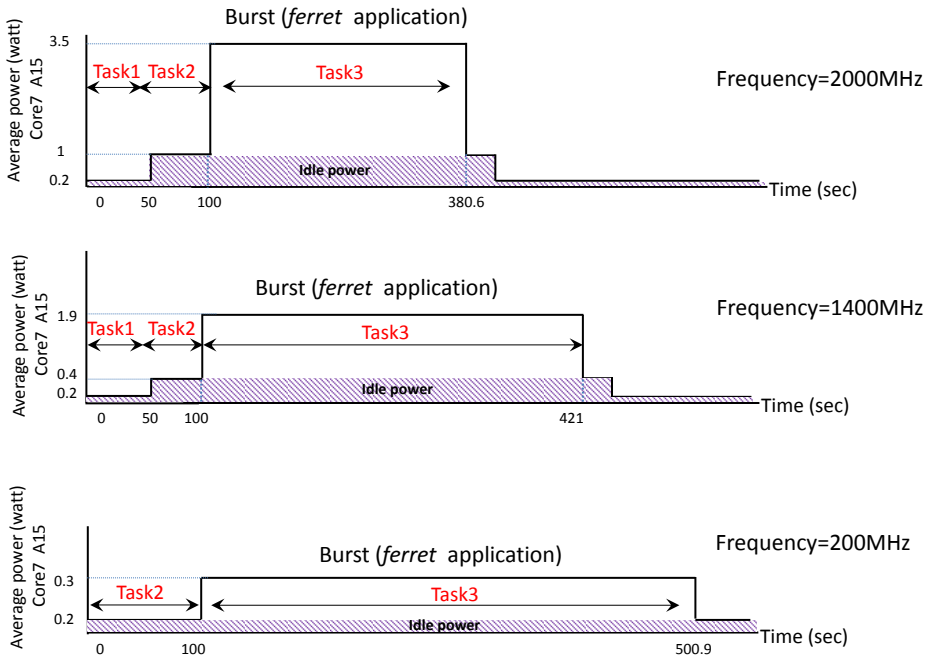


Figure 1.2: Idle power dissipation of Exynos 5422 big.LITTLE octa-core heterogeneous platform performing *ferret* application.

of operation as well as bursty workload behaviour. This is due to the high leakage energy dissipation of CMOS power transistors in such applications, especially when driving high capacitive loads. This in turn necessitates an increase in the width of these transistors to meet the on-state current requirement (I_{on}), thereby increasing the leakage current (I_{off}) in the idle state.

The source terminal of power switches in the PSN is connected to the supply voltage (V_{dd}) and the drain electrode becomes the effective power supply rail to the power gated block and is referred to as the virtual V_{dd}. The state of these switches is controlled by a control signal generated by the power gating controller, as can be seen in Fig. 1.3. By shutting off the switches in the PSN, the virtual V_{dd} discharges leading to leakage current (I_{off}) being decreased to that of the power gating switches [87]. This method of power gating is named as *coarse grain* power gating and the power gated block is often referred to as a *power domain* (PD).

To capitalise on CMOS scaling, as mentioned in Section 1.1.3, low-threshold transistors are being employed in the PD block to maintain high performance whereas high-threshold power switches are being used in the PSN to power gate the PD block. Adopting high- and low-threshold CMOS transistors in digital design is referred to as Multi-Threshold CMOS (MTCMOS)[87, 113]. Previous work has demonstrated that up to 25× reductions in leakage power can be achieved by employing this technique in the ARM926EJ CPU [87].

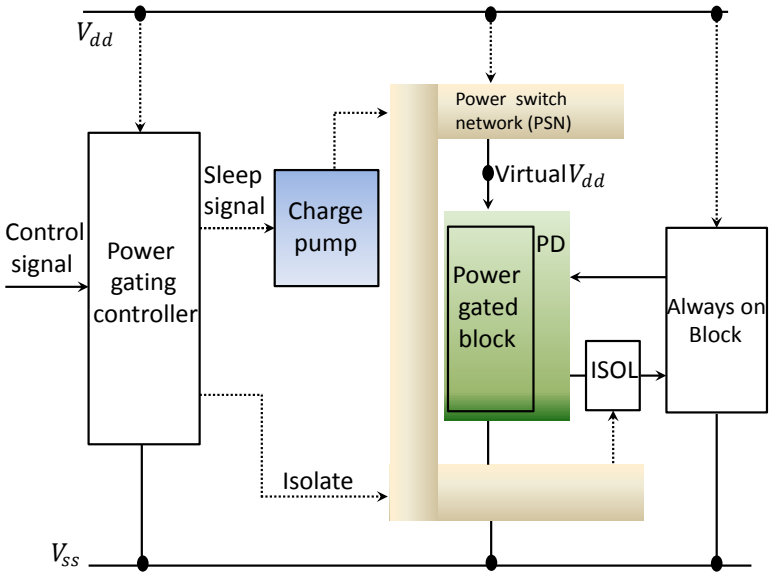


Figure 1.3: Modified architecture of coarse power gating technique based on [87].

However, a high-threshold power transistor has a low driving current (I_{on}), thereby necessitating an increase in the number of parallel power switches in the PSN which in turn leads to an increase of the overall power dissipation in the power gating circuitry.

In practice, the inclusion of a PSN introduces a small IR drop, as these switches can typically be modelled as resistors when the digital circuit is in the active mode of operation [84]. It can be observed from Eq. 1.4, that when substituting virtual V_{dd} with $(V_{dd} - V_{drop})$, where V_{drop} is the voltage drop across the PSN, the propagation delay of the digital gates in the PD block is increased accordingly, thereby resulting in performance degradation. To that end, if the PSN is comprised of CMOS transistors, the effective on-state resistance of these transistors can be reduced by increasing their combined width. Furthermore, employing many power gating transistors in parallel can be used to limit the IR drop.

One challenge of implementing power gating in digital circuits is that the output of the PD blocks may ramp off very slowly, thereby causing large short circuit (crossbar) currents in blocks which are always powered on. To prevent these short circuit currents, isolation cells (ISOL) are typically employed between the output of the PD blocks and the input of the always on blocks. This is important because the signals float when the power is disconnected from the PD blocks and can cause short circuit currents in any always on blocks, thereby inducing functional problems and high energy consumption [87]. Isolation can be achieved by employing special isolation gates at the output of the PD blocks, including either "AND style" or "OR style" gates [87]. The assertion of the control signals to the isolation gates

and power gated switches necessitates careful timing to guarantee the valid functionality of the power gating technique. Typically, these signals are generated from a controller state machine which is referred to as *power gating controller*, and is shown in Fig. 1.3.

The present study proposes the employment of zero-leakage MEMS devices in the PSN to eliminate the leakage current by separating drain-source terminals via an air gap in the off-state. One of the challenges of implementing MEMS in any digital design is the requirement of a high actuation pull-in voltage (V_{pi}). To that end, the architecture of the power gating technique used in the previous study has been modified accordingly by incorporating a bootstrap charge pump block which is explained further in Chapter 4.

1.1.5 *Applications*

In the past, performance has been the impetus for the development of many applications due to the demand for interactive devices including tablet computers and smart phones, while, energy-efficiency has been kept as a desirable but not essential ultimate goal. However, there are a plethora of embedded applications where performance is not the primary target and instead energy-efficiency is the main constraint.

In this thesis a system level architecture is proposed for energy-constrained portable applications, as discussed in detail in Chapter 5. It is postulated that a computer system can be built in a similar fashion to biological systems where two types of operation, regular and burst, are incorporated such that a constantly active part has to be relatively slow and all the fast processing has to be done in specialised (peripheral) units.

Regular activities occur most of the time, and are intended to serve the needs of the entire system. These activities are typically determined by the dynamics of the system and overall structure. Intuitively, for large size systems regular activities are fast and for small systems they are slow. Typically, regular activities are synchronized effectively to the power supply as well as to the energy distribution inside the system, in an operation which is regulated itself in an energy efficient way. In fact, the supply itself is an activity which is regulated by such a central management core in the system. That is to say, it is better if something that is supposed to work all the time should be gated regularly rather than irregularly, and with a minimum of energy consumption during each power gating. The speed of change in each cycle is generally related with charging/discharging of energy buffers, and it is well known that if the speed of change is slow, as so called "adiabatic cycling", then the total amount of heat consumption during any charge/discharge cycles is significantly minimal [40, 171].

Bursty activities are not usually those regularly initiated by normal periodic cycles, but rather in response to the demand of interaction

with external changes or conditions. Such activities that can be dealt with by special purpose peripheral subsystems, whose allotted energy resources are supplied by the action of the system's core, and whose operation are specified by their local dynamics and structure. Inspired by the biological evolution, the optimality of response to bursts is specified by the closet fit to the duty cycling required to respond to such bursts. The ultimate level of effectiveness in bursty modes will be specified corresponding to the amounts of energy resources allotted to these peripheral subsystems by the main core [171]. Intuitively, peripheral subsystems that operate in a bursty mode are normally locally timed (based on the limit cycles of the peripheral subsystem, that is to say, according to the characteristic inherent in its topology). The speed of charging/discharging of the energy buffers in the peripheral subsystems can be comparatively high because this is how they approach optimum energy utilisation in those bursts of activity [171]. For instance, a burst should be accompanied by minimum leakage energy as well as by the producing of an effective outcome within the allotted interval of power supply.

Power gating technique using zero-leakage MEMS relays has been employed in such a system, to capitalise on leakage energy reduction of burst activities, thus improving the total energy-efficiency. This is attributed to the fact that, unlike CMOS power switches, MEMS based power gating favour targeting complex architecture as well as burst nature of operation. Furthermore, MEMS relays can be preferably implemented in a non-invasive power gating applications due to their limited lifetime compare to that of CMOS counterparts, as explained further in Chapter 2.

As a result, to that end, power gating based on MEMS can also achieve more energy saving than CMOS counterparts in applications that operate at low throughput with a significantly long idle period. Wireless Sensor Networks (WSNs) are one of such type of applications. Typically, WSNs comprise of small sensing nodes that can be used for environment, habitat, and health monitoring. Hundreds of WSNs are placed out in the field to collect, process, and submit data over periods of months to years. Traditionally, the processor used in WSNs demands operating frequency in the range between kHz-MHz. Therefore, to maximize device life time average power dissipation of these devises are desirably set in the order of several μWs for 10s-100s. As an example, the Zebranet application uses a Texas Instruments MSP430 which operates at 32KHz and dissipates approximately 300μW [173]. Furthermore, biomedical devices are another area where high performance is inessential, but energy-efficiency is a key challenge. Moreover, the Internet of Things (IOTs) emerged applications are counted to be analogous to the WSNs necessitating similar low processor performance but one of the key challenges is constrained energy resources [12].

1.2 THESIS ORGANIZATION AND KEY FINDINGS

This thesis is organized into seven chapters, as shown in Fig. 1.4. The major contributions of this thesis is summarized as follows:

Chapter 1 "Introduction": introduces the motivations, objectives and structure of this thesis.

Chapter 2 "Background and Literature Review": This chapter provides a coherent overview of extensively adopted and recently reported state-of-the art MEMS/NEMS relays that can be used for particular kinds of applications. Subsequently, this survey can be the basis for any research intending to take advantages of using these emerging devices and to address their challenges. Furthermore, a device-level comparison of different types of MEMS/NEMS relays is provided. The result of this comparison highlights the future requirements for the characteristics of MEMS/NEMS so that they could be implemented in many emerging applications.

Chapter 3 "A Scalable Physical Model for Nano-Electro-Mechanical Relays ": we propose a switch model simulator based on Verilog-AMS that can be plugged into the standard EDA software. To test the scalability of the proposed model we cross checked it against our analysis of a range of benchmark circuits. This chapter also investigates the modelling, designing and optimization of various MEM/NEM switches using 3D Finite Element Analysis (FEA) performed by the COMSOL multiphysics simulation tool. An extensive parametric sweep simulation is performed to study the energy-latency trade-offs of MEM/NEM relays. To accurately simulate MEMS/NEMS-based digital circuits, a Verilog-AMS model is proposed based on the evaluated parameters obtained from the multiphysics simulation tool. This allows an accurate calibration of the MEM/NEM relays with a significant reduction in simulation speed compared to that of 3D FEA executed on COMSOL tool.

Chapter 4 "MEMS-based power gating of asynchronous micropipelines for ultra low energy driven computing": This study demonstrates the threshold at which our approach can achieve greater energy savings in relation to the design architecture and data rate of the input. Our proposed paradigm offers 69% energy improvements at data rate of 1KHz compared to a 39% achieved in the previous paradigm.

Chapter 5 "MEMS-Based Power Delivery Control for Bursty Applications " we propose to incorporate a power-switching network based on MEM relays to switch the SoC power-performance state (PPS) into an active mode while eliminating the leakage current when it is idle. Results show that even with today's large and high pull-in voltages, a MEM relay-based power switching network (PSN) can achieve a 1000× savings in energy compared to its CMOS counterpart for low duty cycle.

Chapter 6 "MEMS-Based Idle Energy Minimization for Bursty Workloads in Heterogeneous Many-Core Systems " we propose a novel MEMS-based zero-leakage current idle energy controller. Core to our approach is an integrated sleep mode management based on the performance energy states indicated by performance counters. For effective energy minimization we use a systematic optimization of the controller design parameters by adopting finite element analysis (FEA) in multiphysics COMSOL tool. We showed that for bursty workloads exemplified by a PARSEC benchmark application up to 55.5% energy savings can be achieved on an Exynos 5422 big.LITTLE octa-core heterogeneous platform. Further, our approach benefits from non-invasive design of the controller suitable for commercial off-the-shelf and low cost implementations.

Chapter 7 "Conclusions and Future Work" outlines the conclusions of the study and explore the implications of the presented work and draw the horizon for prospective future research.

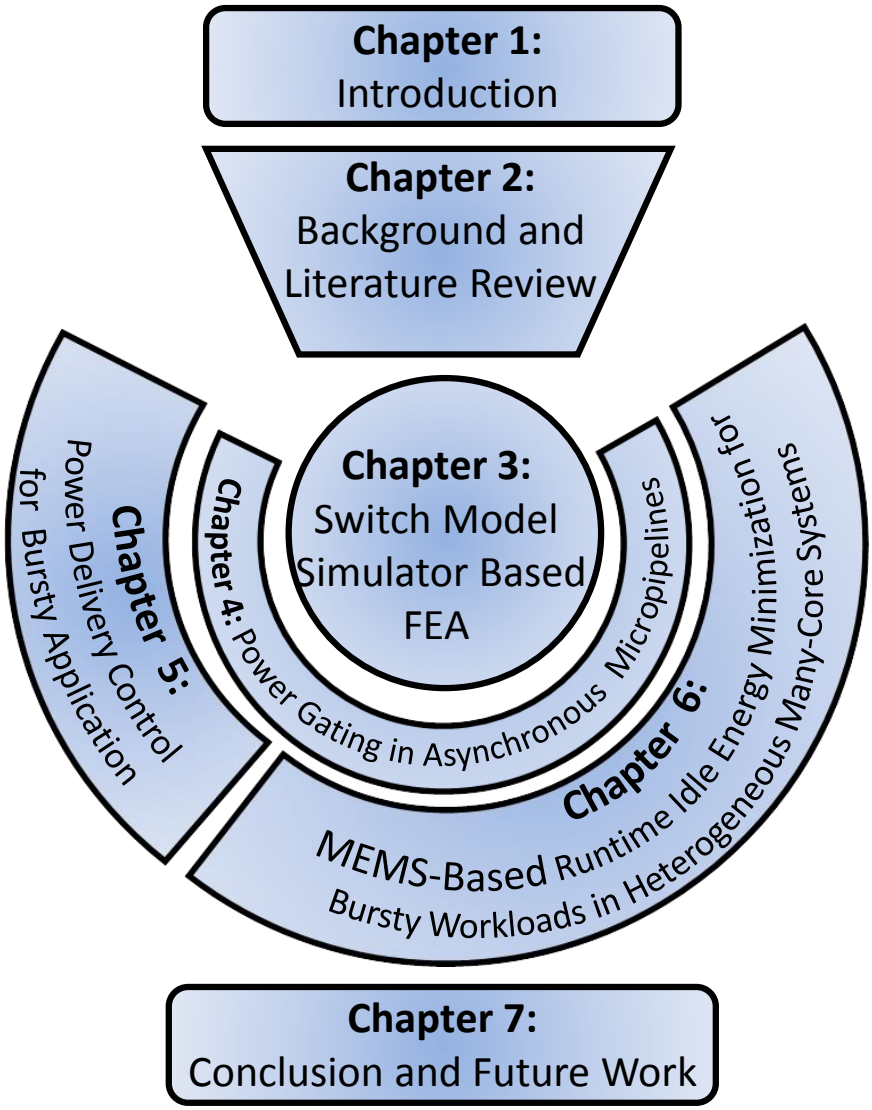


Figure 1.4: Thesis organization.

2

BACKGROUND AND LITERATURE REVIEW

2.1 INTRODUCTION

To overcome the energy-efficiency limitations of CMOS transistors operating at or near sub-threshold voltage, an alternative switching mechanisms are being investigated. These alternative devices include the Tunnelling Field Effect transistor (TFET) [91], Tri-Gate transistor, Nano-electromechanical field effect transistor (NEMFET), ferroelectric FETs, and impact ionization MOS. Unfortunately, many of these CMOS-like transistors achieve a steep sub-threshold swing ($S < 60mV/dec.$) over only a specific range of supply voltage. This leads to a significant shortcoming including either a poor (I_{on}/I_{off}) current ratio or a very low I_{on} current at a low supply voltage.

A feasible way to overcome the limitation of energy-efficiency in CMOS circuits can be deduced from Fig. 2.1. If the slope of sub-threshold regime of the CMOS can be made steeper, CMOS circuits would experience immeasurable low leakage current at the same supply voltage, making it possible for further improvements in energy-efficiency. This would require, then, an alternative device that overcomes the essential CMOS energy-efficiency limit.

To that end, the Micro-Electro-Mechanical/Nano-Electro-Mechanical (MEM/NEM) relay has recently emerged for ultra-low-power digital circuit applications, as summarized in Table 2.4. This is because the relay exhibits perfect abrupt on-off switching behaviour (less than 0.1 mv/decade [115]), since it has nearly an ideal switch characteristic. Furthermore, the Micro-Electro-Mechanical (MEM) relay has immeasurable off-state leakage current and steeper sub threshold slope. Therefore, its operating voltage can be scaled down to approach zero, in principle. Thus, the MEM relay can potentially overcome the fundamental energy efficiency limit of the CMOS technology.

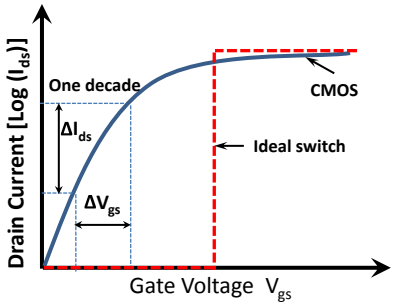


Figure 2.1: Sub-threshold regime of MEM relay and CMOS transistor.

Table 2.1: Survey of electro-mechanical relays actuation scheme. This table compares MEMS actuation characteristics.

MEMS	Piezoelectric	Electrostatic	Magnetic	Thermal	Ref.
Switching speed [µs]	10-500 (✓)	0.01-0.2 (✓)	100-5000 (~)	100-5000 (~)	[118]
Simple fabrication	(✗)	(✓)	(✗)	(✓)	[142]
Pull-in voltage [V]	5-50 (✓)	1-100 (✓)	1-5 (✓)	0.5-5 (✓)	[164]
Bias current [mA]	~zero (✓)	~zero (✓)	20-150 (✗)	0.5-10 (✗)	[174]
Low power	(✓)	(✓)	(✗)	(✗)	[92]
High force	(✓)	(✗)	(✓)	(✓)	[153]
Scalability	(✓)	(✓)	(✗)	(✗)	[36]

2.2 BACKGROUND OF MEM/NEM RELAYS

In this section we briefly introduce MEM/NEM relays which are used to control power gating in our approach. Typically, relays can be classified based on the method of actuation into electrostatic [144], electrothermal [133], magnetostatic [162], and piezoelectric [103]. Each type of actuation scheme has specific advantages and drawbacks as listed in Table 2.1. Moreover, they could also be classified according to the axis of deflection (lateral, vertical), contact interface (ohmic or capacitive), and geometric shape (see-saw beam, cantilever beam, dual bridge, clamped-clamped beam, parallel plate, sidewall perimeter beam), as summarized in Table 2.3.

Based on the method of actuation, each relay has different characteristics including: bias voltage, bias current, on resistance, delay time, current handling, and endurance as illustrated in Table 2.1.

Among these relays, the electrostatically actuated MEM relay has recently received a remarkable attention in digital logic applications due to its low active power consumption, scalability, fast switching, and ease of manufacture using conventional planar processing techniques [115]. As a result, electrostatic actuated MEM relays have been adopted in the present approach to power gate idle circuits. Among various types of MEM relays, a 4-terminal parallel plate MEM relay in [156] is the focus of this study. This is due to its large recorded number of on/off cycles without demonstrating any operating failure, as illustrated in Table 2.2. Additionally, unlike the 3-terminal MEMS, the state of a 4-terminal switch can be determined based on gate-body terminal voltage (V_{gb}), which is independent of the source-drain voltage. Fig. 2.2(a) shows the schematic 3-D view of the 4-terminal suspended gate MEM relay while Fig. 2.2(b) presents the schematic 2-D cross section showing the parasitic capacitance and resistance in the off-state. To minimize parasitic capacitor effects and improve reliability, drastic improvements in the 4-terminal relay design have led to the fabrication

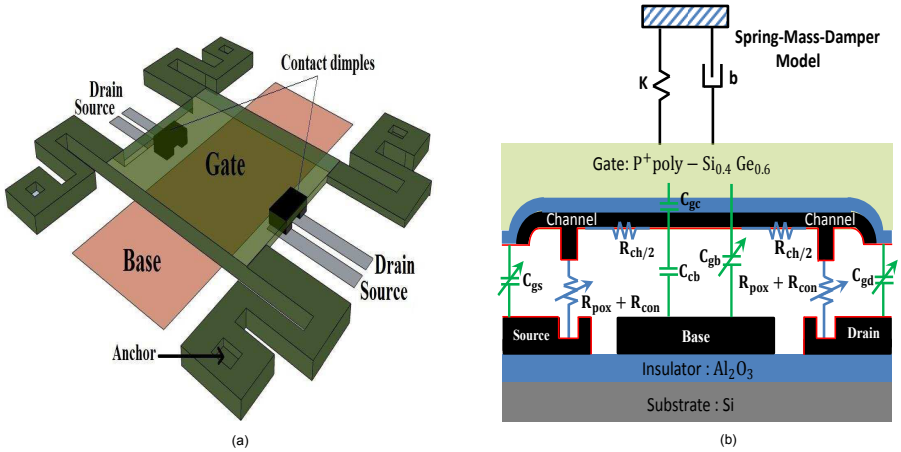
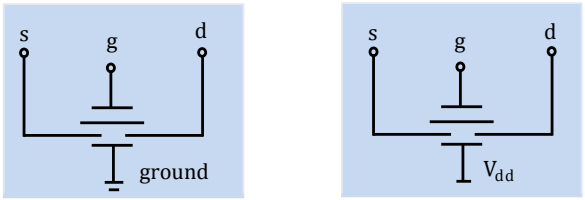


Figure 2.2: Shows the:(a) schematic 3D view of 4-terminal suspended gate MEM relay based on [156]; (b) schematic 2-D cross section view in the off state.

of three generations of devices as reported in [116]. As a consequence, numerous promising implementations of Micro-Electro-Mechanical Switch (MEMS) and Nano-Electro-Mechanical Switch (NEMS) have been proposed recently which exhibit an order of magnitude more power savings than the CMOS in low frequency applications $\leqslant$ 100 MHz [17, 45, 48, 82, 96, 106, 165]. It should be noted that these implementations are demonstrated either using simulation tools or fabrication process, as described in Table 2.4

2.3 OPERATING PRINCIPLES OF ELECTROSTATIC RELAY

The four-terminal MEM relay, as shown in Figs. 2.2 and 2.4, has recently emerged to overcome the shortcomings of the 3-T MEM relay. Adding the fourth terminal means that the relay turns on either by applying a positive V_{gb} (mimicking NMOSFET operation) or by applying a negative V_{gb} (mimicking the operation of the PMOSFET), as shown in Fig. 2.3. Furthermore, connecting many 3-T switches in series causes the gate-to-source voltage to change in an undesirable manner, and hence this may affect the state of the switch [128]. Generally, a 4-terminal relay consists of a gate (g) electrode that carries a metal



(a) N type MEM/NEM relay (b) P type MEM/NEM relay

Figure 2.3: 4-terminal relay device symbol.

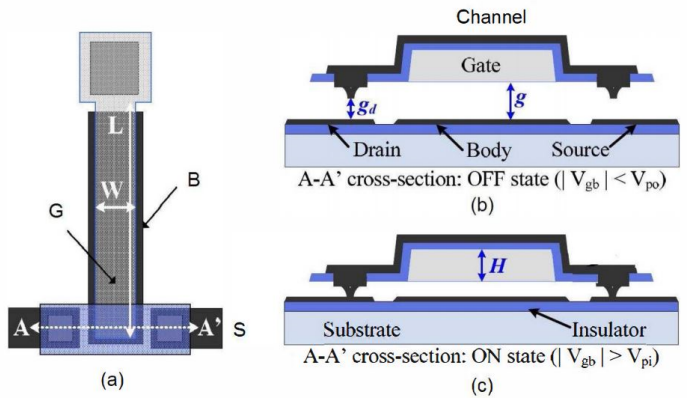


Figure 2.4: Cantilever beam relay as reported in [29]: (a) plan view; (b) cross section view in the off-state; (c) cross-section view in the on-state.

channel separated by a gate oxide [29, 156], in addition to the body (b), drain (d), and source (s) electrode which is located below the gate as shown in Figs. 2.2 and 2.4. To alleviate the impacts of any residual stress that may cause bending the beam out of the plane shown in Fig. 2.4, a suspended gate relay, as shown in Fig. 2.2, has emerged to address this problem. The position of the gate electrode depends on the balance between electrostatic force and spring force. As a result, when the electrostatic force is sufficient enough to overcome the spring force, the channel touches the drain/source electrode, thereby allowing current to flow. Otherwise, an air gap separates the channel from the drain/source electrode when the electrostatic force is less than the spring force.

The body-gate terminal is connected to V_{gb}, the source is connected to V_s, and the drain is connected to the load capacitance C_L. Initially, the gate-to-body electrode is set to zero ($|V_{gb}| = 0V$) and an air gap separates the drain from the source. Due to the air gap, the current flowing between the drain-source terminal is equal to zero so that the MEM relay is in the off-state, as shown in Fig. 2.5.

As V_{gb} is increased, as shown in Fig. 2.6, an electrostatic force is generated between the gate-body terminal. As a result, the gate bends towards the body terminal, and thereby an elastic spring force is generated which counteracts the electrostatic force. Increasing V_{gb}

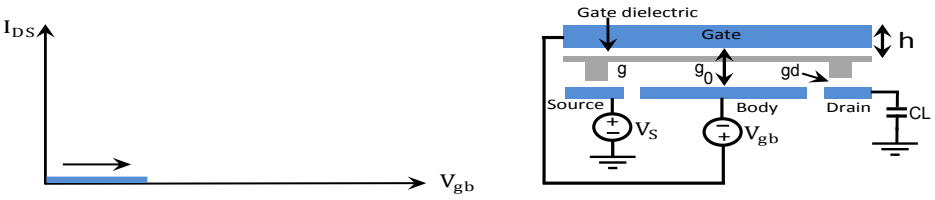


Figure 2.5: Schematic of MEMS before pull-in. This graph shows the MEMS with a gate-body voltage (V_{gb}) applied that is below V_{pi} and the corresponding V_{gs} vs. I_{ds} curve.

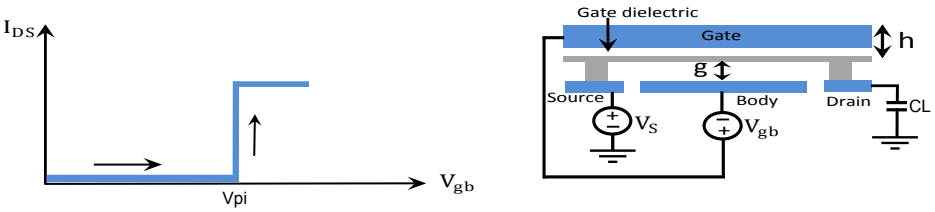


Figure 2.6: Schematic of MEMS after pull-in. This graph shows the MEMS with a gate-body voltage (V_{gb}) applied that is above V_{pi} and the corresponding V_{gs} vs. I_{ds} curve.

leads to an increase in the gates deflection at a higher rate due to the positive feedback which in turn decreases the gap distance and leads to increase electrostatic force. As V_{gb} reaches a level above the pull-in voltage (V_{pi}), a large enough spring force cannot be produced to balance the electrostatic force. Consequently, the gate electrode collapses into the drain-source terminal. At this stage, the gate dimples are in electrical and mechanical contact with the drain-source terminal. As a result, current flows from the source to the drain terminal, and the device is in the on-state, as shown in Fig. 2.6.

To turn off the MEM relay, V_{gb} is lowered, which decreases the electrostatic force between the suspended gate and the body terminal. Due to the reduced gap when the device is pulled-in, the dimples remains in contact with drain-source terminal even when V_{gb} is decreased to V_{pi}, as shown in Fig. 2.7. Eventually, as V_{gb} continues to decline, the spring elastic force becomes larger than the electrostatic force, and thus the dimples pull out of contact. At this stage, current no longer flows from source towards drain terminal, and the MEM relay returns to the off-state. The voltage where this occurs is called the release voltage and is referred to as V_{rl}, as shown in Fig. 2.8. Since V_{rl} occurs at a lower voltage than V_{pi}, hence the MEM relay's IV characteristics demonstrates hysteresis. The hysteresis window increases further with an increase in the adhesion forces between the dimples and drain-source terminals.

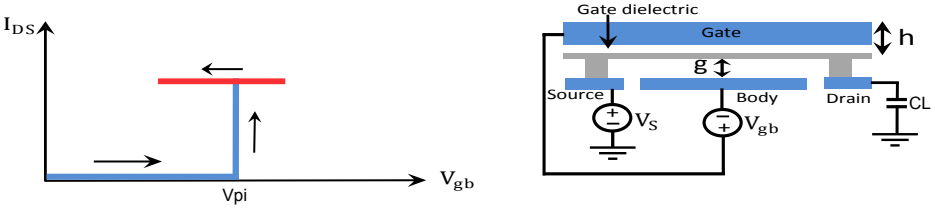


Figure 2.7: Schematic of MEMS before pull-out. This graph shows the MEMS with a gate-body voltage (V_{gb}) applied that is above V_{rl} and the corresponding V_{gs} vs. I_{ds} curve.

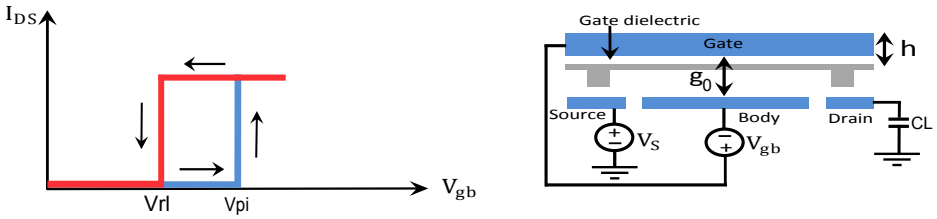


Figure 2.8: Schematic of MEMS after pull-out. This graph shows the MEMS with a gate-body voltage (V_{gb}) applied that is below V_{rl} and the corresponding V_{gs} vs. I_{ds} curve.

2.4 SURVEY OF MEM/NEM RELAYS

Over the past decade, researchers have shown substantially increased interest in utilising MEM/NEM relays as low power logic devices. Table 2.2 gives a summary of the efforts and progress made in developing MEM/NEM relays in terms of materials used, actuation method, V_{pi}/V_{rl}, contact material, contact resistance, longevity, and geometrical shape. Table 2.3 explains the abbreviations used in Table 2.2.

Endurance is of major importance for digital logic applications, since it is important that the device remains functional even beyond its predicted lifetime. This means that endurance is more crucial than speed or power consumption in digital IC applications. For example, it can be assumed that the required endurance for the device to work without failure is ten years. This means that, for digital applications which work at 100 MHz with an average transition ($0 \rightarrow 1$) probability of 0.01, a life time of 3.15×10^{14} on/off cycles will be required. Therefore, it is logical to say that MEMS based digital logic applications requires at least 10^{14} on/off cycles. Among these relays, the MEM/NEM switch published in [101] has the largest recorded number of on/off cycles (2.1×10^9) without experiencing any operating failure, other studies [33, 81, 115, 163] demonstrate a slightly lower number of on/off cycles (1×10^9). It should be noted that these number of cycles are measured when no current is passing through the beam's dimples. Due to stiction issues, a significantly less number of on/off cycles can be obtained when current is passing through the dimples. This means that further investigations are required in terms of materials and geometrical shapes of MEMS in order to improve their endurance.

The on-state resistance of CMOS is non-linear and varies as the transistor operating point changes. Unlike CMOS, the contact resistance of MEM relays is approximately constant and insensitive to the gate slew rates. However, a recent study in [115] shows that R_{on} increases by one order of magnitude after a course of 10^4 on/off cycles. However, this should not significantly degrade circuit performance as the throughput of digital circuits based MEM relays is limited by the mechanical delay rather than the electrical delay. Among the recorded relays shown in Table 2.2, MEMS published in [75, 71, 78, 96, 115]

demonstrate the lowest values of contact resistance found as 12Ω, 20Ω, 312Ω, 800Ω, and 1400Ω respectively.

Optimizing the switching energy consumption of MEM/NEM relays is a crucial factor especially when low power applications will be targeted. To the best of my knowledge, results in [137] demonstrate the lowest achieved switching energy with measured value of 82.59fJ.

2.5 SURVEY OF APPLICATIONS EMPLOYING MEMS/NEMS

MEM/NEM relays are a promising technology that have been employed in a wide range of applications such as the following:

1. ***Medicine:*** MEMS/NEMS-based pressure sensors has been used to monitor blood pressure in various organs.
2. ***Inertial Sensing:*** MEMS/NEMS devices have been used as accelerometers to sense mechanical movement.
3. ***Biotechnology:*** MEMS/NEMS enables the implementation of biochips for detecting hazardous chemicals and high-throughput drug screening and selection.
4. ***Communications:*** MEMS/NEMS technology has been employed in radio frequency (RF) applications, including receivers and transmitters, due to low insertion loss, linearity, low power consumption, and high isolation. The primary requirement of these relays, in order to minimize insertion losses, is to achieve low contact resistance. This can be obtained by adopting high conductivity material such as gold.
5. ***Switches:*** MEM/NEM relays are being widely employed in digital logic applications for low power designs, and due to their high reliability.

In this work, an extensive survey has been conducted of digital logic applications that employing MEM/NEM switches. Empirical results in [80, 116] have shown that MEM/NEM switches can withstand much higher temperature and radiation doses than CMOS counterparts. This means that the MEM relay is more attractive than CMOS in applications in which there are significant variations in radiation and temperature, such as space and military applications. Accordingly, pull-in voltage (V_{pi}) and on-state resistance (R_{on}) of MEMS have been experimentally proven to fluctuate between (8-9.2)V and (1-15)kΩ when the temperature range is between 30-200 C [80, 116]. Meanwhile, it was found that at low temperatures MEM relays work normally. To assess radiation endurance, an MEM relay was irradiated with several doses such as 200 Krad, 2 Mrad, and 20 Mrad. The results showed that V_{pi} is shifted by only 10% after exposure to 20 Mrad. However, lower doses do not cause any variations in the pull-in voltage (V_{pi}).

Table 2.2: Survey of Electro-Mechanical Relays Used for Digital Logic Applications.

Ref.	AM	No.	Geom.	AD	Structural Material	Longevity	V_{pi}/V_{po}	Contact Material	CR (Ω)	E_s (pJ)	Dimension
[78]	E	7	SS	V	Poly-SiGe	400	6.6/0.76	W-W	800	12.6*/6.4*	h=1μm, g_0=200nm, g_d=100nm, L_A=42μm, W_A=40μm
[129]	P	6	CB	V	Si_3N_4, SiO_2, PZT	1×10^6	1.5/NR	Au-Au, Au-Pt, Au-Ru	1K	0.5-3	h=1.21μm, g_d=100nm, L=25μm, W=25μm
[96]	E	6	SW	L	Poly-Si, HfO_2, TiN	NR	31/21.6	TiN	1K	NR	h=660nm, g_0=440nm, L=20μm g_d=340nm
[33]	E	5	DB	V	Si_3N_4	1×10^9	1.5/0.6	W-W	NR	NR	h=110nm, g_0=10nm, L=75μm W=20μm
[125]	E	5	CB	L	Poly-Si	1×10^8	12/6	Pt-Pt	3K	NR	h=1.2μm, g_0=600nm, g_d=500nm, L=16μm , W=600nm
[149]	E	5	CB	L	Poly-Si, TiN	NR	7.9/5	TiN-TiN	23K	NR	g_{gs}=0.6μm, g_{ds}=0.5μm, L=20μm
[38]	E	5	CC	L	Ru	2×10^6	4/NR	Ru-Ru	NR	NR	h=200nm, g_0=75nm, g_d=50nm, L=10μm , W=100nm
[25]	E	5	SW	L	Poly-Si, TiN, Al_2O_3	NR	NR	NR	NR	NR	W_{beam}=1μm, L_{beam}=50μm, g_0=9nm
[50]	E	5	SW	L	SiN_x, HTO, Poly-Si, LTO	NR	20/19.5	LTO-NR	NR	NR	h=2.1μm, g_0=550nm, L=14μm , W=1μm
[115, 156]	E	4	PP	V	Poly-SiGe	1×10^9	6.5/5.9	W/TiO_2-W/TiO_2	1.4K	1.8*	h=1μm, g_0=180nm, g_d=90nm, L=27μm , W=30μm
[60]	E	4	CB	V	Poly-SiC	NR	13.5/6.3	poly-SiC-poly-SiC	NR	NR	g_0=300nm, g_d=250nm, L=12μm , W=250nm
[32]	E	3	CB	L	Pt	NR	3.3/2	Pt-Pt	0.1G	NR	h=60nm, g_0=100nm, g_d=90nm, L=3.54μm , W=70nm
[72][73]	E	3	CC	V	CNT film	NR	4.5/2.5	CNT film-Au	NR	NR	h=100nm, g_0=50nm, g_d=50nm, L=3.45μm , W=1μm
[101]	E	3	CB	L	PolySiC	2.1×10^9	NR/NR	polySiC-polySiC	NR	NR	h=400nm, g_0=200nm, g_d=175nm, L=8μm , W=200nm
[95]	E	3	SW	L	TiN	1000	34/29	TiN-TiN	NR	NR	h=1.1μm, g_0=350nm, g_d=200nm, L=7μm , W=200nm
[97]	E	3	CB	V	TiN	10	16/14	TiN-TiN	NR	NR	h=40nm, g_0=40nm, g_d=20nm, L=700nm , W=200nm
[39]	E	3	CC	V	W	NR	2.5/NR	W-Au	NR	NR	h=32nm, g_0=55nm, g_d=18nm, L=1.83μm , W=500nm

... to be continued

Table 2.2 (continued): Survey of Electro-Mechanical Relays Used for Digital Logic Applications.

Ref.	AM	No.	Geom.	AD	Structural Material	Longevity	V_{pi}/V_{po}	Contact Material	CR (Ω)	E_s (pJ)	Dimension
[81]	E	3	PP	V	Poly-SiGe	1×10^9	5.3/4	W/TiO_2-W/TiO_2	8.1K	1.1*	h=1μm, g_0=200nm, g_d=100nm, L=30μm , W=15μm
[139]	E	3	CB	L	W	1500	40/35	NR	500M	NR	h=1μm, g_{ds}=0.45μm, g_{gs}=0.55μm, L=(20-50)μm,W=0.5μm
[54, 137]	E	3	CB	L	SiO_2	NR	8/NR	PtSi	5K	0.082	g_0=75nm, L=15.5μm
[163]	E	3	CB	V	Ni	1×10^9	27.5/25	Al/Al_2O_3-Au	NR	NR	h=210nm g_0=430nm, g_d=210nm L=10μm, W=4μm
[75]	E	3	CB	V	Ni	1×10^5	22/18	Au-Au	20	NR	h=1.7μm g_0=500nm, g_d=150nm L=40μm, W=8μm
[100]	E	3	CB	V	CNT	NR	3.5/2	CNT-Au	12.5K	NR	h=60nm g_0=135nm, g_d=135nm L=1.5μm, W=60nm
[74]	E	3	CB	V	Poly-Si	NR	24.6/NR	Poly-Si-Poly-Si	NR	NR	h=1μm g_0=500nm, g_d=200nm L=30μm, W=8μm
[71]	E	3	CB	L	SiO2, Si, W	NR	0.8/0.7	W-NR	312	NR	h=1.5μ g_0=95nm, L=9.5μm, W=130nm
[71]	E	3	CB	L	SiO2, Si, W,Pt	NR	11.3/3.5	Pt-NR	NR	NR	h=1μm g_0=320nm, L=13μm, W=300nm
[98]	E	2	CC	V	TiW	20	0.4/NR	TiW-W	NR	NR	h=40nm g_0=4nm, L=1.4μm, W=300nm
[76]	E	2	CB	V	TiN	500	13/8	TiN-TiN	NR	NR	h=35nm g_0=15nm, L=1μm, W=200nm
[77]	E	2	CC	V	TiN	NR	11/6	TiN-W	NR	NR	h=30nm g_0=20nm, L=1μm, W=200nm
[77]	E	2	CC	V	Poly-SiC	NR	1.2/0.7	SiC-SiC	>5M	NR	h=25nm g_0=27nm, L=8μm, W=25nm
[88]	E	2	HS	V	NR	1×10^6	1.7/1	Au-Au	12	NR	L=100μm, W=20μm

Table 2.3: Abbreviations and definitions for Survey of Electro-Mechanical Relays contained in Table 2.2.

Parameters	Definitions
Number of Electrodes	Number of independent terminals for a relay design
Geometry	SW = sidewall perimeter beam, SS = seesaw beam with central anchor, CB = cantilever beam, DB = dual bridge, CC = clamped-clamped beam, PP = parallel plate, HS = hinge structure
Contact Material	Contact material on actuator dimples - contact material on fixed dimples
CR	Contact Resistance (Ω)
Longevity	Most cycles reported or cycles till failure
AM	Actuation Method: P = piezoelectric, E = electrostatic
Dimensions	g_0 = gate-to-bode gap, g_d = gate-to-drain, h = height, W = width, L = gate or anchor length
V_{pi}/V_{po}	Pull-in voltage/pull-out voltage
E_s	Switching energy, (*)= calculated value
AD	Actuation Direction: L = lateral (i.e., in-plane actuation), V = vertical (i.e., out-of-plane actuation)
Others	NR = not reported, K=kilo-($\times 10^3$), M=mega-($\times 10^6$), G=giga-($\times 10^9$)

As a result, many researchers are investigating electronics for harsh environment by employing MEM/NEM relays.

For example, logic gates such as XOR, AND, NAND, and NOT have been designed by only adopting a single MEM device [33]. This can reduce the device count as each CMOS-based logic gate requires 6-16 transistors. Consequently, this can help to compensate for the area penalty which arises from the fact that each individual MEM relay is significantly larger than the minimum size of CMOS transistor. Reducing the number of MEMS can improve speed, reliability, yield, and simplify the implementation of the circuits. As a result, a 1-bit multiplexer by adopting 4 AND gates, a 1-bit adder by adopting 3 XOR and 2 AND gates, and 2-bit full adder circuit by adopting 6 XOR and 4 AND gates have all been successfully validated in [33].

Other studies have presented experimental results from test chips containing circuits implemented with MEMS [156]. The results from a test chip, which contains 8 bit adder, 2-bit accumulator, SRAM, flip-flops, DRAM, 4-bit DAC, and power gates, have verified the correct functionality of MEM relay-based circuits. Fig. 2.9 shows the design style of an MEM relay-based full carry adder. It is predicted in this research that at the 90 nm technology node, MEM relay can achieve 10 times energy savings over the minimum-energy point of CMOS. Furthermore, the authors analytically demonstrated that applications which operate at maximum throughput (20 MOPS) would significantly benefit from utilizing NEMS. Noticeably, it was theoretically discussed that NEM relay can be adopted for high throughput applications by utilizing parallelism methodology. For example, a 32 parallel adder-based NEMS can operate at 0.5 GOPS with 10× savings in energy

compared to CMOS. However, due to the area-speed trade-off constraint of NEMS, this paradigm requires about a 100× increases in area overhead.

In [48], the micro-architecture of building a (7:3) compressor has also been experimentally demonstrated by adopting only 98 MEM relays. The measured results of energy and area overheads are consistent with those of the previous study [156]. Accordingly, the performance of a MEMS-based multiplier is improved by a factor of 8× over the static CMOS style and by 4× over the pass-transistor logic style. This necessitates more investigations to explore the best MEMS-based logic style in terms of performance and switching energy that can be employed in large scale digital circuits.

Other researchers have recently focused on using MEM relays as power gating devices in energy-constrained applications, due to their zero-leakage characteristics. An analytical demonstration showed that MEMS can achieve more energy saving than CMOS power switch counterparts in such applications where the T_{off}>1ms and T_{on} >100 ns [47]. However, this analysis did not explain how to provide the power gating circuitry with high pull-in voltage (V_{pi}). For accurate analysis the energy overhead of implementing a DC-DC charge pump should be added to the overall energy consumption. Furthermore, the timing analysis of the proposed power gating controller was not explained in detail. It is essential to ensure that the hold and setup time conditions are achieved during circuit switching on/off. This work also demonstrated that MEMS-based power switch at 90nm technology node can offer an order of magnitude more energy savings than CMOS counterparts, if the stand-by time T_{off}>10µs.

Other works have aimed to demonstrate the potential of adopting MEM relays as power gating devices in periodic and event-driven applications [47, 62, 63, 138, 151]. These analytical studies have demon-

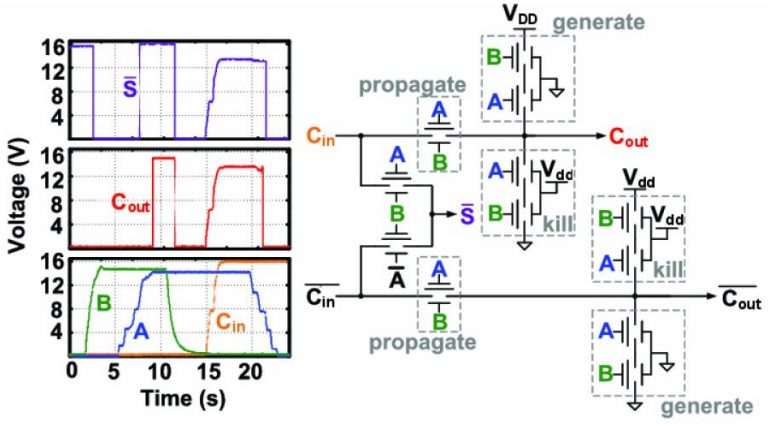


Figure 2.9: MEMS based carry adder circuit and measured waveform. This circuit is implemented on a 1µm test chip [156].

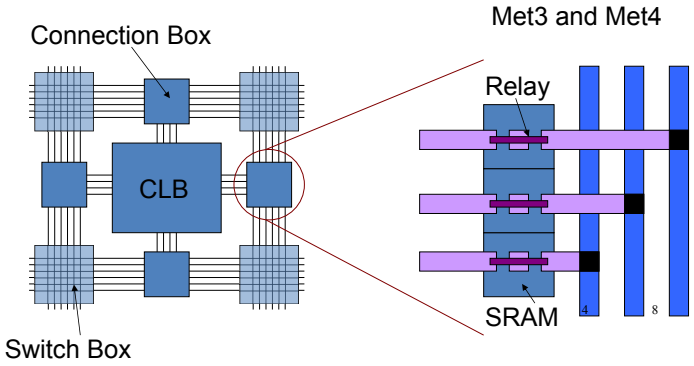


Figure 2.10: NEM relays as FPGA routing switches based on [27, 28].

strated that the MEM relay is a perfect candidate for this kind of application if the stand-by time T_{off}>100 ms.

To address limitations and drawbacks of the previous studies, a novel work based on finite element analysis (FEA) has aimed to implement MEMS in applications exhibiting bursty workloads [8, 9]. The results showed that up to 50% energy savings can be achieved when MEM relays are employed. The asynchronous power gating of an FIR filter micro-pipeline using MEMS has also been presented [10]. The results showed that the MEM relay based power gating switch favours complex architecture coupled with bursty workload behaviour.

Work in [27, 28, 154] presented an analytical study of the employment of MEMS/NEMS in programmable routing switches located in FPGAs, as shown in Fig. 2.10. This study proposed that programmable routing switches and their routing SRAM can be replaced by a single MEMS/NEMS. This is attributed to the fact that NEMS exhibits hysteresis behaviour which can be utilised as the memory element as reported in [74, 85, 114, 147, 170]. Analytical results showed that NEMS at the 22nm technology node can achieve a reduction of 28% in critical path delay, 37% leakage power reduction, and 43.6% footprint area reduction compared to CMOS counterparts. However, some points remain to be investigated in future research, including incorporating NEMS into the look-up tables of logic blocks (LB), a detailed analysis of the voltage requirements for the hybrid NEM/CMOS design to address the noise issues, and an architectural exploration of FPGAs to take advantage of the integration of NEMS.

Previous work has proposed MEMS-based flash ADC, as shown in Fig.2.11. This results showed that 350 fJ of energy is consumed in a single cycle of the flash converter [29, 104]. This means that a 6-bit 10MS/s converter requires 5.5 fJ per conversion which is an order of magnitude better than CMOS-based converters. This converter, however, suffers a from high footprint area and a low operating speed of about 10 MHz (maximum throughputs = $2\times(t_{D,on} + t_{D,off})$. It should be noted that theses findings are obtained from simulation tools, as described in Table 2.4. Others have proposed the first ring

oscillator built by adopting digital logic elements based on a curved NEMS with a size of 5×3 μm^2 and an air gap of 60 nm. Empirical results demonstrated that this ring oscillator can operate at a frequency of 6.7 MHz [13].

To simulate mixed CMOS/MEMS electronics, many studies are being conducted to build a scalable and accurate physical model simulator of the NEMS that can be plugged into the standard EDA software. Previous work focused on building these simulators either by adopting fabricated parameters or by using multiphysics tools [7, 16, 17, 24, 58, 83, 96, 131]. The present work [7] is then extended to incorporate the use of a COMOSL multiphysics tool to verify the electrical and mechanical parameters which are written in Verilog-AMS file and co-simulated in the Cadence tool, as described further in Chapter 3.

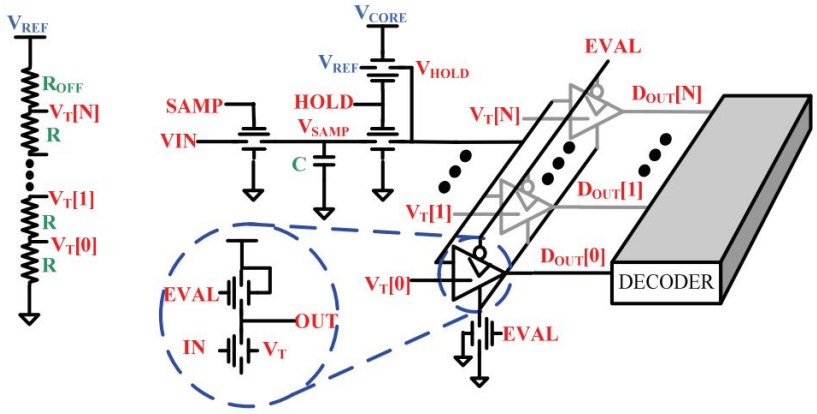


Figure 2.11: MEMS based ADC circuit diagram proposed in [29].

2.6 CONCLUSION

This chapter provides an extensive background and literature review of MEMS switches used for digital logic applications. This survey summarizes the progress and effort made in developing these relays in terms of switching energy, actuation method, material used, geometrical shape, and contact resistance. Furthermore, this work presents a novel survey based on applications which benefit using of these emerging devices. This survey briefly describes these applications with the up to date achieved results as well as the foundations/universities that published these works.

Table 2.4: Survey of Emerging Applications EmployingMEM/NEM Relays.

Approach	Application	Description	Results(simulated/fabricated)	Foundation/University
[33, 80, 116]	Harsh environment: high temperature and radiation,	Withstand temperatures greater than 200 C and absorbed radiation doses greater than 20 Mrad, 2 orders of magnitude better than radiation-hardened CMOS	Better reliability than CMOS counterparts (Fabricated)	(Bristol, California) university, NEMIAC, MIT
[8, 9]	Bursty and low throughput applications, non-invasive, including biomedical devices, sensors, RFIDs	Power delivery control for bursty applications, runtime idle energy minimization for bursty workloads in heterogeneous many-core systems	Achieved up to 1000× energy saving than CMOS (Simulated)	Newcastle university
[10, 161]	NEMS-based asynchronous circuits design/power gating of asynchronous micropipeline	Investigates ways to combine both NEMS and asynchronous design/ explore power gating in asynchronous micropipeline (FIR filter). Results illustrate that MEMS favours complex architecture coupled with fast computation implemented in non-invasive applications	Results show that 8-bit PCHB adder, 32-bit PCHB AND, and 64-bit C-element can achieve up to 1.7×, 25×, and 16× better energy-efficiency than CMOS counterparts (Simulated)	(Newcastle, Cornell) university
[57, 94, 107, 132]	DC/DC converter for stand-alone sensing systems and aerospace applications	Benefits of implementing MEMS switches in DC/DC converters are as follows: no voltage loss due to p-n junction(i.e. zero leakage current), radiation resistance, low ohm losses, capability to convert voltage up/down, capability to simultaneously charge all capacitors or based on charge transfer, prolonged operation by adopting Ir contacts	Better energy-efficiency (Simulated)	(South florida, Southampton, Newcastle) university, CQCSTC
[156, 48, 165, 34, 136, 78, 6, 5, 115, 79, 128, 49, 96, 43, 50, 126]	Integrated circuits for VLSI applications such as: 32- bit adder, multiplier, (7:3) compressor, sequential circuit, XOR, AND, buffer	Digital circuits have been implemented by using MEM/NEM relays. Theses circuits designed in such way that all relays switches at the same time to mitigate T_{mech} impact	8× and 10× lower energy consumption for multiplier and adder respectively (Simulated/Fabricated)	(California, MIT, Bristol, Texas, Stanford) university, On-Chip Power Corp., SEMATECH
[14, 65, 66]	Adiabatic logic application	NEMS-based adiabatic logic style is investigated for ultra low energy application. Three NEMS-based adiabatic architectures are proposed including: TRAG, BISAG, and SNAG	NEM-adiabatic logic can achieve at least an order of magnitude better energy saving than CMOS technology (Simulated)	IBM (NEMIAC), Delft university of technology

... to be continued

Table 2.4 (continued): Survey of Emerged applications employingMEMS/NEMS relays..

Approach	Application	Description	Results(simulated/fabricated)	Foundation/University
[47, 62, 63, 138, 151]	Highly periodic and event-driven processing, baseband processor, and battery operated systems including wirless border cameras, biomedical implants, and structural health monitoring systems	Investigates shut-off completely power supply of combinational integrated circuits when it is in stand by mode, idle, by employing MEMS/NEMS power switches	Results illustrate a potential upto $100\times$ reduction in energy consumption at off-periods $>$ 1 second (Simulated)	(California, MIT, Purdue, Virginia Tech, King Abdulla), university
[74, 85, 114, 147, 170]	NEMS based memory cell such as: mutinary RAM, NVSRAM, content addressable memory (CAM),DRAM-like mechanical non-volatile memory	Investigates the potential of implementing memory device based on MEMS/NEMS	Results show better energy-efficiency, area, and stand-by power consumption than CMOS counterparts (Simulated)	IMEC, ICREA, IIIA-CSIC (Polytechnic, California, Viswajyothi) university
[7, 17, 96, 58, 131, 83, 16, 24]	Simulation and synthesis tools for MEMS/NEMS relays	Two particular issues that these papers are addressed, firstly, the need for a scalable and accurate physical model of the MEM/NEM switches that can be plugged into the standard EDA software, Secondly, synthesis tools of MEM/NEM switches such that incorporate less number of T_{mech}	Converges, fast simulation tools	(Bristol, California, Newcastle, MIT, Cambridge) university, IBM, Coventor Inc.
[68, 69]	Thermal management of 3D many-core systems	Studies hybrid CMOS-NEMS designs of thermal buffer and power gating to reduce leakage power and thermal-runaway in 3D many-core systems	Experimental results show that MEMS/NEMS based thermal management can effectively prevent the thermal-runaway in 3D multi cache core processor (Simulated)	Nanyang university
Others [13, 24, 29, 104]	Ring oscillator, parallel A/D conversion and image processing	Design ring oscillator operates at 6.7MHz based on MEMS/NEMS, Investigate the design space of implementing image processing algorithm by employing a hybrid multiple-input NEMS-CMOS architecture	Better leakage reduction/power efficiency (Simulated/Fabricated)	(Bristol, Pennsylvania, California) university

3

A SCALABLE PHYSICAL MODEL FOR NANO-ELECTRO-MECHANICAL RELAYS

3.1 INTRODUCTION

This chapter presents the methodology, background, and results for the present research related to the MEMS-based switch mode simulator. Circuit simulation based on MEM relays enables designers to successfully optimize and validate their designs from device to system levels of abstraction. Many exploratory research projects are currently under way investigating the mechanical, electrical and logical characteristics of MEM/NEM relays. One particular issue that this thesis addresses is the need for a scalable and accurate physical model of the MEM/NEM switches that can be plugged into the standard electronic design automation (EDA) software [7].

To implement a usable behavioural model of MEM/NEM relays that can be utilized for simulating large scale MEMS-based circuits, an accurate conservative Verilog-AMS language was chosen. This choice is due to its flexible description of the device behaviour in both the mechanical and electrical domains [93, 166]. The principle of building such a simulator is basically inspired by electrostatically actuated RF MEM switches, with taking into account some differences in parameters including R_{on} and the quality factor (Q)[24]. Furthermore, efforts by [16] have led to building analytical compact model based on Verilog-AMS for in-plane electrostatic actuated cantilever ohmic MEMS.

Having a scalable model for MEM/NEM relays is crucial for designing systems which consist of mixed MEMS/CMOS electronics. To evaluate the logical characteristics of MEM/NEM relays for a range of circuits, research by Matt Spencer has led to the building of a Verilog-AMS switch model based on experimental results for MEMS and predicted parameters for NEMS [30, 156, 157]. This model is accurate and detailed. However, the simulation takes a long time and there is a convergence problem. To overcome the limitations of the model in [156], this thesis proposes an approximate model which is more scalable, fast, simple, and more stable (especially near contact discontinuities)[7]. The major contributions of this chapter are:

1. Implementing and evaluating the standard MEMS/NEMS model. The model is analysed and its pros and cons are evaluated and discussed. The model is found to be accurate. However, it is shown to suffer from a convergence problem for large designs due to the model's complexity and non-linearity [30].

2. Building a more scalable MEMS/NEMS model which can be applicable for very large scale integration (VLSI) circuits without any divergence issue.

3. Verifying the proposed model against a range of benchmarks with an average error rate of 6%. Moreover, for the same benchmarks, the proposed model does not suffer from divergence problems, while the standard model does diverge as it becomes larger and more complex.

4. Proposing a novel systematic optimization of MEMS parameters using finite element analysis (FEA) executing in the multiphysics COMSOL tool. These parameters can then be used in in our simulator to accurately simulate MEMS/NEMS in various geometric shapes, materials, dimensional sizes, and actuation directions.

3.2 ANALYTICAL MODELLING OF ONE-DIMENSIONAL MEM/NEM RELAY

3.2.1 *Terminology and definition of MEM/NEM relay parameters*

Operating in a pull-in mode (i.e, $\frac{g_d}{g_0} \geqslant 0.3$) results in the most energy-efficient relay which is preferable for digital logic applications [82]. However, the pull-in mode of operation causes the electrical characteristics of the MEMS/NEMS to exhibit some hysteresis behaviour [1]. Pull-in voltage (V_{pi}) refers to the voltage which is able to overcome the resistance of the spring-mass-damper system and to cause the relay to turn on. On the other hands, the release voltage (V_{rl}) refers to the voltage required to pull-out the device. The mechanical turn-off delay time is lower than the mechanical turn-on time ($T_{mon} > T_{moff}$). This is attributed to the fact that the electrical contact is broken rapidly as the gate electrode moves 1 nm away from the body electrode [156], while, the gate electrode needs to travel across the entire gap between gate and body to turn the device on. Noticeably, the mechanical delay time is an order of magnitude higher than the electrical delay time, as can be seen in Table 3.2.

3.2.2 *Mechanical Modelling*

The movement of gate electrode under the applied voltage is governed by a nonlinear time-variant second order differential equation [56]. The spring-mass-damper system as shown in Fig. 2.2 has been used to model the mass of the gate and flexures:

$$m_{eff}\ddot{Z} + \frac{\sqrt[2]{km_{eff}}}{Q}\dot{Z} + kZ = F_{ele}(Z) + F_{vdw}(Z) \quad , \tag{3.1}$$

where Z is the displacement of the gate, m_{eff} is the effective mass, Q is the quality factor between [0-1] for digital logic applications [82], k is the spring constant (N/m), and F_{vdw} is the Van der Waals force.

The electrostatic force F_{ele} (which is always attractive, i.e., ambipolar) is equal to the derivative of electrostatic energy stored in the C_{gb} capacitor with respect to the gap thickness [56]:

$$F_{ele} = \frac{\varepsilon_0 A_{ov} V_{gb}^2}{2(g_0 - Z)^2} \quad , \tag{3.2}$$

where ε_0 is the permittivity of free space, V_{gb} is the voltage between the gate-body electrodes, g_0 is the area gap thickness when Z=0, and A_{ov} is the overlap area between gate and body electrodes.

In a nano-scale relay, undesirable attraction forces such as Van der Waals and Casimir forces can significantly affect the pull-in stability of NEMS. It has been noted [150] that Van der Waals force is dominant over the dispersion force when the air gap thickness is several tens of nanometres, whereas, Casimir attraction force is more effective as the air gap thickness increases below several nano-metres. Therefore, in this work, only Van der Waals will be considered in the simulation since the gap thickness of the adopted NEMS is 10nm. The Van der Waals force (F_{vdw}) of the suspended gate MEMS/NEMS can be expressed in a more intuitive formula based on privious research [150] as:

$$F_{vdw} = \frac{AwL}{6\pi(g_0 - Z)^3} \quad , \tag{3.3}$$

where A is the Hamaker constant, L and w is the length and width of the suspended gate respectively.

The voltage required to switch-on the device "V_{pi}" can be derived as follows [1]:

$$V_{pi} = \sqrt{\frac{8kg_0^3}{27\varepsilon_0 A_{ov}}} \quad . \tag{3.4}$$

The mechanical delay time "$T_{mech.}$" of the MEM/NEM relays is inversely proportional to the gate over drive ($\frac{V_{gb}}{V_{pi}}$), resonant frequency $w_0=\sqrt{\frac{k}{m}}$, and actuation-gap to contact-gap thickness ratio. This is presented as [156]

$$T_{mech.} \propto \left(\sqrt{\frac{m}{k}}\right) \cdot \left(\frac{V_{pi}}{|V_{gb}|}\right) \cdot \left(\frac{g_d}{g_0}\right) \quad . \tag{3.5}$$

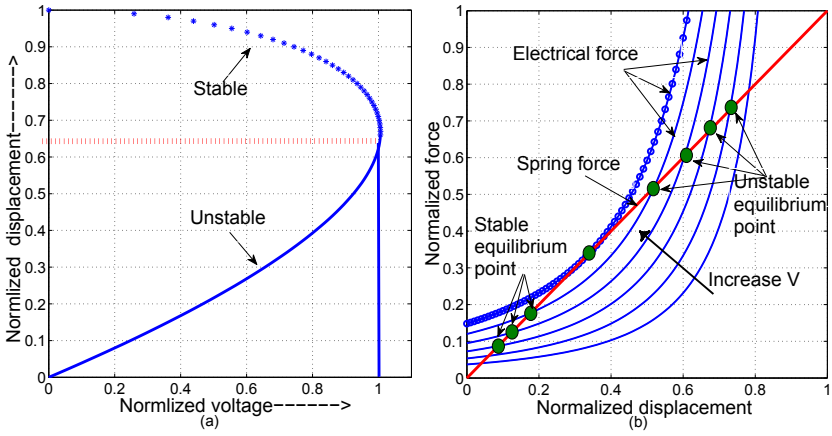


Figure 3.1: Illustrates that: (a) the solid blue line denotes the unstable actuation region, while the dotted one represents the stable region. The saddle-node bifurcation happens at 0.3 of the gap distance; (b) stability analysis, normalized displacement vs. net forces.

3.2.2.1 *Static analysis*

In this section, let us first postulate that Eq. 3.1 is reduced to mechanical and electrical forces. The first part is linear, while the second one increases as the inverse of the squared (g_0-Z). The system is thus non-linear as shown below:

$$F_{net} = kZ - F_{ele}(Z, V_{gb}) = 0 \quad . \tag{3.6}$$

By normalising gap distance such as $\xi = \frac{Z}{g_0} = \frac{g_0 - g}{g_0}$, and operating voltage by $v = \frac{V}{V_{pi}}$, then substituting it in Eq. 3.6 yields:

$$\xi - \frac{4 \times v^2}{27(1-\xi^2)} = 0 \quad . \tag{3.7}$$

To visualize the interaction between the electrostatic field with the mechanical response, and actuated voltage with corresponding displacement, normalized parameters are plotted in Fig. 3.1. The stable equilibrium point for a given voltage arises when F_{net}=0, and thereby any perturbations of the gate terminal cause it to return back to its original position, as can be seen in Fig. 3.1(b). Even though the unstable equilibrium point also occurs when F_{net}=0, however, any perturbations of the gate terminal cause it to close the gap. Fig. 3.1(a) illustrates that the pull-in phenomenon occurs when the gate terminal moves approximately $\times$0.3 of the actuating gap (g_0).

3.2.2.2 *Dynamic analysis*

Due to the electromechanical nature of a MEM relay's operation, the electrical energy applied gets converted into mechanical energy. To analyse the switching energy and how it is converted between

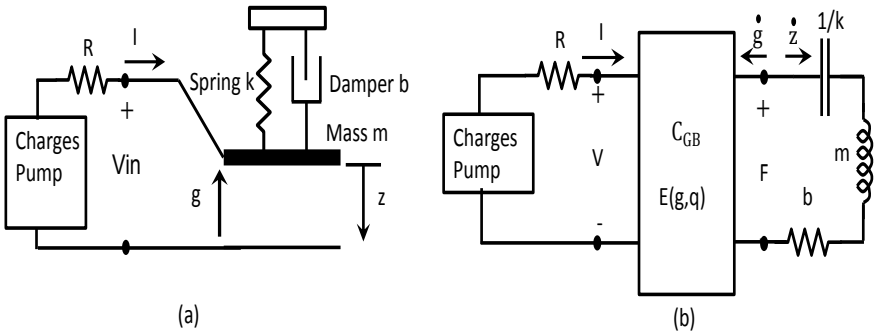


Figure 3.2: Shows: (a) the spring-mass-damper model; (b) electrical representation of the mechanical domain [9].

the two domains, a mass-damper-spring model is used to represent the MEM/NEM switch, as shown in Fig. 3.2(a). In order to achieve a powerful single representational model of the MEM/NEM relay, the mechanical parameters have been converted into electrical equivalent parameters, as shown in Fig. 3.2 (b). Applying Kirchhoff's Voltage Law (KVL) on the left and right parts of Fig. 3.2(b) yields:

$$\begin{bmatrix} \dot{q} \\ \dot{Z} \\ \ddot{Z} \end{bmatrix} = \begin{bmatrix} \frac{1}{R}\left(V_{in} - \frac{qZ}{\varepsilon_0 A}\right) \\ \dot{Z} \\ \frac{-1}{m}\left(\frac{q^2}{2\varepsilon_0 A} - k(g_0 - Z) + b\dot{Z}\right) \end{bmatrix} . \tag{3.8}$$

These non-linear states can be used to model the dynamic behaviour of the MEM/NEM relays, and this is referred to as a charge control position method [19]. In this thesis, however, non-linear states based on the voltage control position have been utilized to model the dynamics of MEMS as well as to build our simulator, as illustrated in Eq. 3.19.

The behaviour of non-linear systems can be very complex, thereby leading to so called jump-phenomena between different states [158]. Methods of linearisation about operating points are quite common in the literature and can be used to avoid the divergence caused by the non-linear behaviour of MEMS [1, 158].

Fig. 3.3 shows the operating dynamics of the fabricated MEMS found in [156]. The clear pull-in event is noted in Fig. 3.3(a), and then, once the voltage is reduced below V_{rl}, release occurs. Furthermore, the crash resets the velocity to zero at the moment where the pull-in is evident. Fig. 3.3(a-b) show a small oscillatory motion as the gate terminal pops up (is released) and oscillates about its rest position. The switching energy that the MEMS consumes during the pull-in is shown in Fig. 3.3(c) which is about 1.6pJ. Fig. 3.3(d) shows the phase plane plot of velocity vs. displacement. The cycle of pull-in, crash, and release is clearly indicated. The oscillation on release is indicated as cyclic motion about the final rest position.

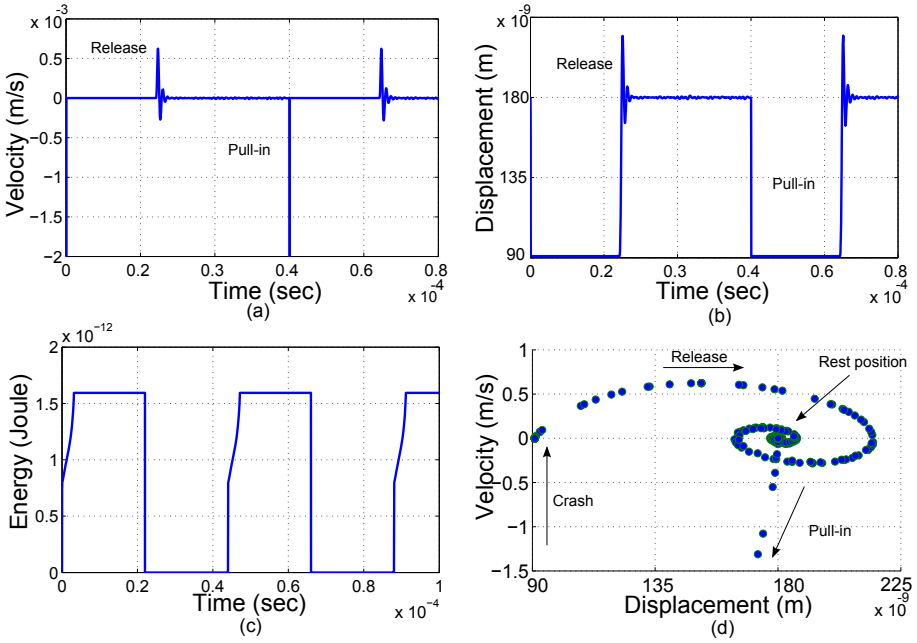


Figure 3.3: Dynamic analysis of the fabricated MEMS [156]: (a) velocity during the transient; (b) displacement at release and pull-in; (c) switching energy; (d) phase-plane of velocity vs. displacement.

3.2.3 *Distribution of energy dissipation in the MEMS/NEMS*

To analysis the distribution of switching energy of the electo-mechanical system, the principle of Lagrange's equations is used in this work. This is attributed to the fact that Lagrangian mechanics uses the energies in the system instead of forces to describe the equations of motion. It states that the dynamic system in which the work of all forces is accounted for in the Lagrangian, an allowable motion between certain setups of the system at time t_1 and t_2 is a natural motion if, and only if, the energy of the system stays constant. It should be noted that the distribution of energy dissipation in the MEMS/NEMS is based on the guideline found in [44]

The Lagrangian is a quantity that describes the equilibrium between potential and kinetic energies with excluding dissipative energies, and can be written as follows:

$$L = K_e - V_e \quad , \tag{3.9}$$

where L represents the Lagrangian quantity, V_e is the potential energy, and K_e is the kinetic energy. The Lagrange equation is introduced as:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} + \frac{\partial P}{\partial \dot{q}} = F \quad . \tag{3.10}$$

Where p represents the dissipation of the energy in the system, F denotes the generalized external forces acting on the system (i.e. for electrostatic actuated MEMS, the external force is the voltage source (V_{in})), and q denotes the coordinates of the system (i.e. for electrostatic actuated MEMS, they will be the displacement (z) and charge (q)). The kinetic energy of the MEMS can be driven as:

$$K_e = \frac{1}{2} m\dot{z}^2 \quad , \tag{3.11}$$

the potential energy (V_e) has mechanical and electrical components due to the restoring spring and the capacitance respectively.

$$V_e = \frac{1}{2\epsilon_0 A}(g-z)q^2 + \frac{1}{2}kz^2 \quad . \tag{3.12}$$

The power consumption P comprises both mechanical and electrical parts due to the squeeze film damping and source resistance respectively, and is given by:

$$P = \frac{1}{2} R\dot{q}^2 + \frac{1}{2} b\dot{z}^2 \quad . \tag{3.13}$$

To evaluate the Lagrangian, the partial derivatives of Eq. 3.10 with respect to q, z, and their first derivatives has been calculated as follows:

$$\frac{\partial L}{\partial z} = \frac{q^2}{2\epsilon_0 A} - kz, \quad \frac{\partial P}{\partial \dot{z}} = b\dot{z}, \quad \frac{\partial L}{\partial \dot{z}} = m\dot{z} \quad , \tag{3.14}$$

$$\frac{\partial L}{\partial q} = \frac{(g-z)q}{\epsilon_0 A}, \quad \frac{\partial L}{\partial \dot{q}} = 0, \quad \frac{\partial P}{\partial \dot{q}} = R\dot{q} \quad . \tag{3.15}$$

The time derivatives of the Lagrangian with respect to the generalized coordinates are given as follows:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{z}}\right) = m\ddot{z}, \quad \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0 \quad . \tag{3.16}$$

Substituting Eqs. 3.14, 3.15, and 3.16 into Eq. 3.10 yields the following equations:

$$m\ddot{z} + b\dot{z} + kz = \frac{q^2}{2\epsilon_0 A}, \quad R\dot{q} + \frac{(g-z)}{\epsilon_0 A} q - V_{in} = 0 \quad . \tag{3.17}$$

Fig. 3.4 shows the distribution of energy dissipation in the electromechanical system. It is shown that about 68% of the total energy is dissipated as potential energy, while only around 1% is consumed as kinetic energy. The energy consumption due to squeeze film damping

and source resistance is estimated to be around 31%, as reported in previous work [137].

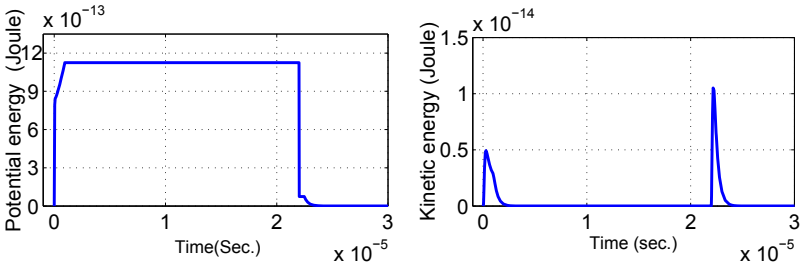


Figure 3.4: Distribution of energy consumption in the MEM relay.

3.2.4 *Electrical Modeling*

Unlike CMOS, a MEM/NEM relay based digital circuit should be designed in a large complex logic gate such that only one mechanical delay incurs at each stage. However, this significantly increases the total on-state resistance, and hence in turn leads to an increase in electrical delay time. Calculating the amount of time required to charge or discharge the load capacitance demands precise modelling of both the on-state resistance and the device capacitances.

The on-state resistance of MEMS/NEMS consists of R_{trace} (the resistance of wire leading to/and from the tungsten electrode), R_{ch} (the resistance of the channel), R_{con} (the resistance of the channel-drain/source contact), and R_{pox}(the resistance of the passive oxide which is used to improve endurance) [156].

The load capacitance of the adopted MEMS/NEMS comes from several parasitic capacitances, such as gate-body (C_{gb}), gate-source (C_{gs}), gate-drain (C_{gd}), gate-channel (C_{gc}), and channel-body (C_{cb}) capacitances. In the off-state, the gate to channel C_{gc} capacitance will not contribute to the overall capacitance as the channel terminal is floating. The electrical delay required to charge and discharge the MEMS/NEMS parasitic capacitance can be written in an approximate formula as:

$$T_{ele.} = R_{on}(C_{gb} + C_{gd} + C_{gs}) \quad . \tag{3.18}$$

3.2.5 *Simulation of Suspended Gate MEM/NEM relay*

A digital MEMS/NEMS relay can be modelled by incorporating all the mechanical and electrical effects shown in Fig. 2.2. An accurate conservative Verilog-AMS model has been adopted in this work to simulate the behaviour of MEMS based on the fabricated/predicted parameters published in [31, 156].

The Verilog-AMS has been utilized in this work as it has a multiphysics framework, which deals with different disciplines such as kinematic, electrical, thermal, and fluidic. This Verilog-AMS model is then co-simulated with the Cadence solver to simulate the mass-spring-damper system defined in Eq. 3.1 and the electrical model shown in Fig. 3.5. This model can handle the self-actuation effects of the MEMS/NEMS but does not cover the thermal impact on the electrical parameters. In order to avoid hidden states in Verilog-AMS, the state-space form has been used to rewrite the inhomogeneous non-linear differential equation Eq. 3.1. This equation describes the motion of the gate electrode according to Newton's second law. According to the state-space definition [119], first order differential equations can be used to describe the input, state, and output variables. This can be presented as follows:

$$\begin{bmatrix} \dot{Z} \\ \ddot{Z} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{-k}{m} & \frac{-\sqrt[2]{km_{eff}}}{Qm} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} (F_{ele.} + F_{vdw}) \quad . \tag{3.19}$$

The procedure of simulating the MEMS/NEMS in this work can be summarized according to Algorithm 3.1:

3.2.6 *Proposed Model*

The equivalent electrical circuit of the MEMS/NEMS can be simplified to a lower complexity model. This is advantageous to mitigate the long simulation time in the standard model while sacrificing some degree of accuracy. This assumption is inspired by the transmission line model approximation in previous study [135]. The proposed simplified paradigm of the MEM/NEM relay can be explained in the following stages:
(1) The trace resistance is very small in comparison with R_{con}, R_{pox}, and $R_{ch/2}$ therefore it can be ignored, as can be seen in Fig. 3.6(a). Where $R_{trace} \ll R_{con}$, $R_{trace} \ll R_{pox}$, $R_{trace} \ll R_{ch/2}$. Furthermore,

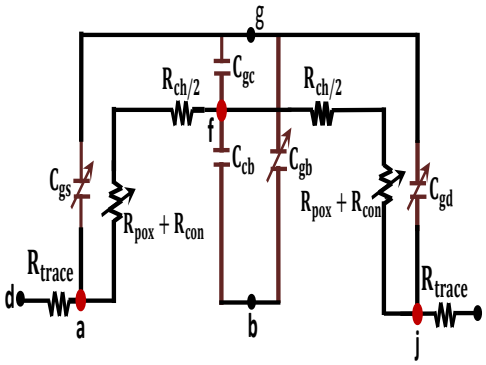


Figure 3.5: On-state electrical characterictics of MEMS/NEMS [156].

Algorithm 3.1 MEMS/NEMS Verilog-AMS model.

Define: Source:=s, Drain:=d, Gate:=g, Base:=b, Displacement:=Z.
Define: Discipline: $Electrical \leftarrow (s, d, g, b)$.
Discipline: $Kinematic \leftarrow (Z, velocity)$.
Input: (V_g, V_s, V_b).
Output: (V_d).
Define: Constant $(C_{gc}, C_{cb}, R_{trace}, R_{ch/2}, R_{pox})$.
Define: Dimple gap (g_d), Spring constant (k), mass, Damping ratio.
1: Initially $(V_g, Z) \leftarrow 0$.
2: Calculate F_e, F_s, F_{vdw} at $(Z = 0)$.
3: Calculate Z_1 by solving:
$\dot{Z} = velocity = Z_2$
$\ddot{Z} = \frac{1}{m}(-\frac{\sqrt[2]{km_{eff}}}{Qm} Z_2 - kZ_1 + F_{ele.} + F_{vdw})$.
4: If $Z_1 < g_d$ Then: (MEMS/NEMS is on)
5: Calculate $C_{gb} = \frac{\varepsilon_0 A_{ov}}{(g_0 - Z_1)}$, $C_{gd} = C_{gs} = \frac{\varepsilon_0 A_{(d)}}{(g_0 - Z_1)}$, $R_{con} = \frac{4\rho\lambda\xi H}{3F_{ele}(gd)}$, F_{ele}, F_{vdw}.
6: Find: $V(d,a) = V(j,s) = R_{trace} * I(d,a)$, $I(g,a) = I(g,j) = C_{gs} * \frac{dV_{(g,j)}}{dt}$, $V(a,f) = V(f,j) = R_{(ch/2+pox+con)} * I(f,a)$, $I(f,g) = C_{gc} * \frac{dV_{(g,f)}}{dt}$.
7: Find $I(d,s)$, T_{ele}, T_{mech}, V_{pi}.
8: Calculate Switching Energy: $E_s = (C_{gb} + C_{gd} + C_{gs})V_{dd}^2$.
9: else: $I(d,s) = 0$. (MEMS/NEMS is off)

due to the small area of overlap between the body and channel electrodes, C_{cb} is insignificant and can also be neglected. In the same way, C_{cg} can be ignored, where $(C_{cg}+C_{cb}) \ll C_{gb}$, $(C_{cg}+C_{cb}) \ll C_{gd}$, $(C_{cg}+C_{cb}) \ll C_{gs}$.
(2) The non-linear parasitic capacitances C_{gd}, and C_{gs} can be linearised by taking the maximum value ($C_{gd} = C_{gs} = 6.6$fF at $Z = 90$ nm). Then, they have been added together as shown in Fig. 3.6(b).
(3) In the same way, R_{con} can be linearised by taking the average value (200 $[\frac{\Omega}{contact}]$). Then, it has been added to the R_{pox} and $R_{ch/2}$ resistance, as shown in Fig. 3.6(c).

In terms of computational complexity, the original model needs to solve 8 non-linear equations to approach the solution. Meanwhile the proposed model can approach the solution by only solving 4 non-linear equations as shown in Algorithm 3.2.

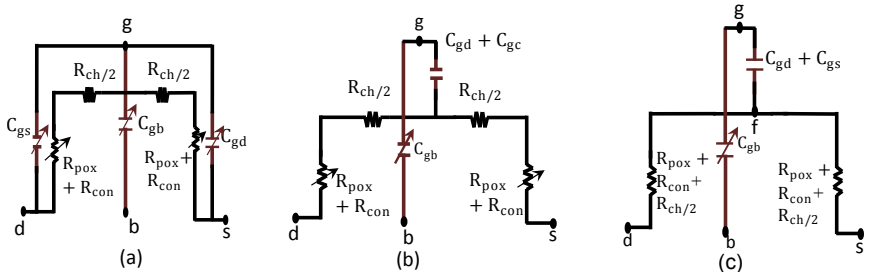


Figure 3.6: Proposed simplifications of MEMS electrical circuit.

Algorithm 3.2 Proposed MEMS/NEMS Verilog-AMS model.

Define: Source:=s, Drain:=d, Gate:=g, Base:=b, Displacement:=Z.
Define: Discipline: $Electrical \leftarrow (s, d, g, b)$
Discipline: $Kinematic \leftarrow (Z, velocity)$.
Input: (V_g, V_s, V_b).
Output: (V_d).
Define: Constant $(C_{gc}, C_{cb}, R_{trace}, R_{ch/2}, C_{gd}, C_{gs}, R_{pox}, R_{con})$.
Define: Dimple gap (g_d), Spring constant (k), mass, Damping ratio.
1: Initially $(V_g, Z) \leftarrow 0$.
2: Calculate F_e, F_s, F_{vdw} at $(Z = 0)$.
3: Calculate Z_1 by solving Eq. 3.19.
4: If $Z_1 < g_d$ Then: (MEMS/NEMS is on)
5: Calculate C_{gb}, F_{ele}, F_{vdw}.
6: Find: $V(d, f) = V(f, s) = R_{(pox+ch/2+con)} * I(d, f)$, $I(f, g) = 2C_{gd} * \frac{dV_{(g,f)}}{dt}$, $I(g, b) = C_{gb} * \frac{dV_{(g,b)}}{dt}$.
7: Find $I(d, s)$, T_{ele}, T_{mech}, V_{pi}, and E_s.
8: else: $I(d, s) = 0$ (MEMS/NEMS is off)

3.3 MODEL EVALUATION

3.3.1 *Simulation of the Proposed MEM/NEM Relay Model*

To verify the validity of this model, the error rate (latency) between the output signal of the standard and proposed models of cascade AND gates, as shown in Figs. 3.7 and 3.8, has been measured. The results indicate that a 4.6 % error rate can be noted between the output signals of the two models in the case of adopting one AND gate, while it increases to less than 7% after the cascading of 40 AND gates in series as shown in Fig. 3.8. The error rate is linearly proportional to the number of stages, which is advantageous for adopting this model with this acceptable error rate in the VLSI circuits.

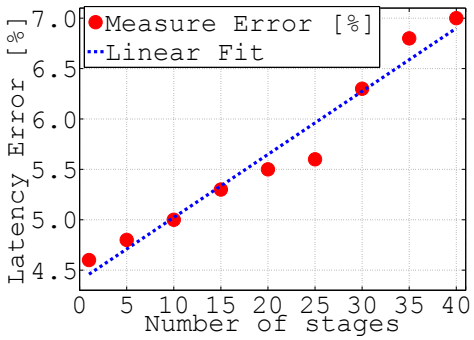


Figure 3.7: Latency error rate.

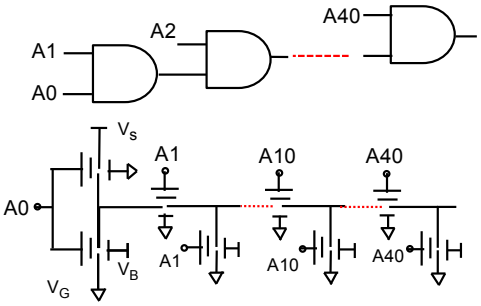


Figure 3.8: CMOS-MEM relay mapping of cascade AND gates.

3.3.2 *Evaluation with Benchmark Circuits*

To evaluate the proposed model in terms of latency, scalability, simulation time and stability, it has been checked against a range of benchmark circuits, including combinational (AND, OR, XOR), sequential (D-latch, C-Element), and arithmetic (carry save adder, carry ripple adder) as shown in Table 3.1. The Results clearly show that the proposed model is 31% faster than the standard model on average. However, as the size of the circuits increases, the simulation time of the standard model is expected to increase drastically, and hence this percentage can be considered only for the benchmark circuits used.

In terms of latency, the results indicate a small difference between the output signal of the two models (less than 6% in average). This is attributed to the presence of approximation in the electrical circuit of the proposed model. It has been noted from the results in Table 3.1 that the standard model usually diverges when the design becomes more complex and larger, such as in: the 3-inputs C-Element and 5-bit carry ripple adder. This is attributed to the fact that the accumulating error due to contact discontinuities will rise significantly as the number of stages increases. Furthermore, numerous non-linear parasitic capacitances/resistances in a big design coupled with the non-linear nature of MEM relay (as shown in Fig. 3.1) is the second reason causing the model to diverge. Moreover, the difference in scale between various state variables in the mechanical and electrical domains are prone to cause the development of numerical errors.

The proposed model has been checked against the standard one for different clock speeds and stage levels. The results shown in Fig. 3.9 indicate that a very slight impact on the simulation time of the proposed model can occur as the clock period changes from 100ns to 500ns. For example, at stage four the simulation time increases by only 10% as the clock period changes to 500ns. In contrast, the standard model shows a significant increase (which is about 26% at stage four) in the simulation time as the clock period shifts from 100ns to 500ns. This is attributed to the fact that the period of discontinuity of the switch will be shorter as the switching frequency increases, and this allows the simulator to approach the solution very rapidly. In

Table 3.1: MEMS Model Benchmark circuits

Circuit		No. of relays	Latency			Simulation Time		
			Standard	Proposed (3)	Error (%)	Standard	Proposed (3)	Improvement (%)
Logic	2-inp. AND	2	15ns	15.7ns	4.6	3m 4s	2m 23s	23
	2-inp. OR	2	15ns	15.6ns	4.0	4m 49s	3m 26s	29
	2-inp. XOR	2	15ns	15.5ns	3.3	848 ms	783 ms	7
	3-inp. Majority	12	15ns	16.1ns	7	6m 15s	3m 5s	51
Sequential	D-latch	4	15ns	15.7ns	4.6	1m 47s	1m 2s	41
	2-inp. C-Element	10	15ns	16ns	6.0	6m 28s	3m 55s	39
	3-inp. C-Element	14	Div.	17ns	—-	Div.	9m 19s	—-
Arithmetic	1-bit 3-inp. CS adder	22	15ns	17ns	13	13m 23s	9m 37s	28
	1-bit CR adder	12	15ns	16ns	6.0	6m 52s	4m 37s	33
	2-bit CR adder	24	Div.	18ns	—-	Div.	14m 34s	—-
	5-bit CR adder	60	Div	22ns	—-	Div.	31m 35s	—-
	16-bit CR adder	192	Div	24ns	—-	Div.	1h 26m	—-
Average					6			31

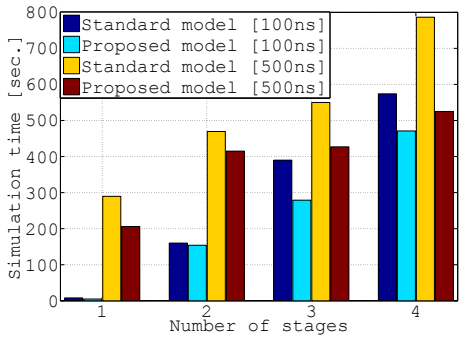


Figure 3.9: Model execution time versus clock speed and number of stages.

contrast, the simulator needs a long time to approach the solution at low switching frequency (with a long discontinuity period).

Fig. 3.10 shows the proposed schematic circuit of a 2 and 3 input C-element using NEM relays (i.e, O/P=AB+O/P(A+B)). The input/output waveforms which are obtained from our Verliog-AMS simulator of 3-input C-element are shown in Fig. 3.11. The results illustrate the correct logic functionality of the circuit with a reasonable delay time. It should be noted that these results are based on the predicted parameters published in [156].

3.4 FINITE ELEMENT ANALYSIS (FEA)

Finite element analysis (FEA) is a numerical analysis method used to solve large numbers of partial differential equation (PDE) for any design. This method is capable of handling multiphysics phenomena and accurately simulating static and dynamic behaviour. To model and capture the physical behaviour of MEMS/NEMS accurately, the COMSOL multiphysics tool has been used in this work. Fig. 3.12(a)

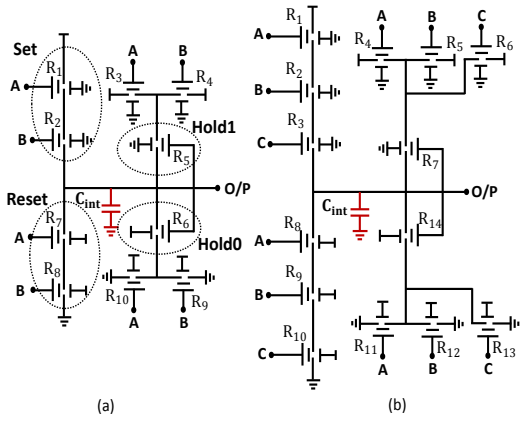


Figure 3.10: Proposed schematic circuit of C-element based NEM relay for (a) 2-input; and (b) 3-input.

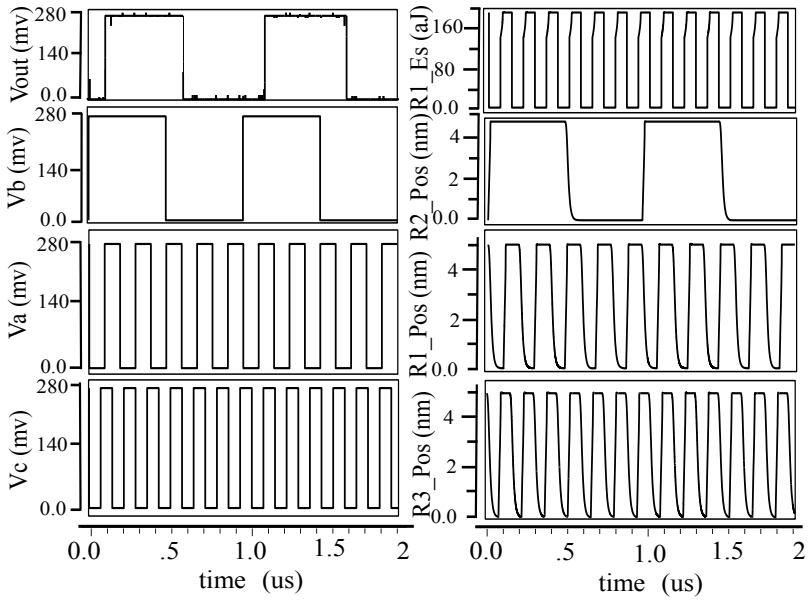


Figure 3.11: Input/output waveforms of 3-input C-element using NEM relays at V_{dd}=0.28 V.

shows the simulated pull-in voltage (11.1v) using 3D FEA, while Fig. 3.12(b) depicts the simplified sketch of the adopted MEMS in our analysis.

In this work, a comparative study of various geometric shapes of MEM relays in terms of pull-in voltage, bending out of plane, and residual stress has been investigated. These experiments were carried out using 3D FEA in COMSOL multiphysics tool. For fair analysis, the simulated relays are designed with similar area size, and material type (poly-SiGe). Fig. 3.13 illustrates that 4-terminal relays with double fold spring, as expected, are pulled-in with lower voltage and switching energy compared with that of 4-terminal relays with quadruple fold spring. In this experiment, two geometric shapes of 4-terminal relays with double fold spring are simulated as shown

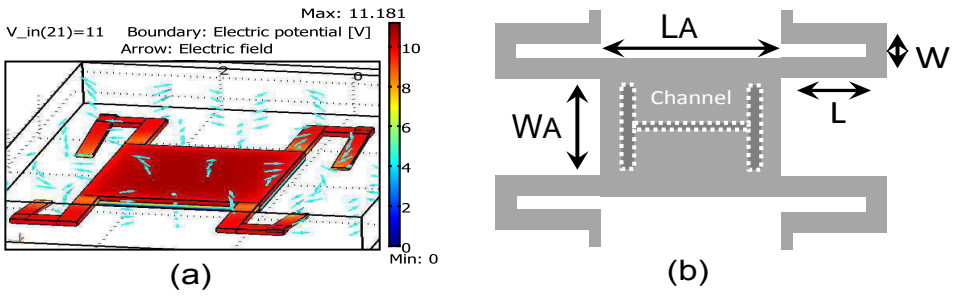


Figure 3.12: Demonstrates the: (a) FEA-simulated pull-in voltage and displacement; (b) simplified sketch, symbols L, W, L_A, W_A and h denote, respectively, spring length/width, actuation area length-/width, and thickness of the suspended gate [156].

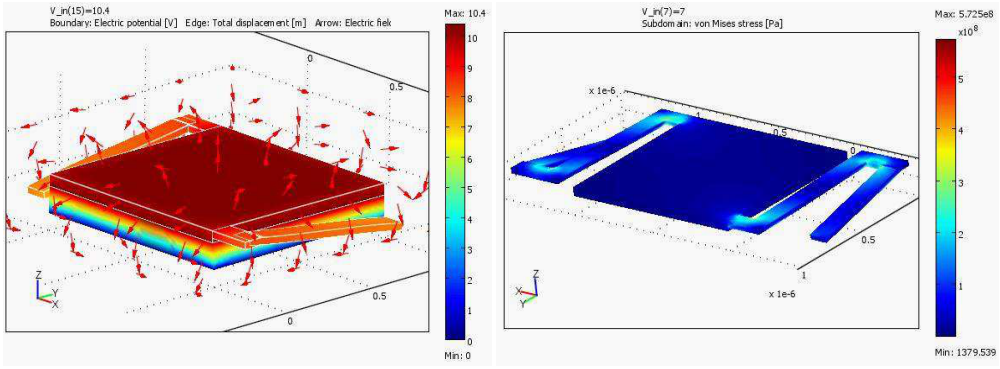


Figure 3.13: Demonstrates the 3D FEA of: (a) 4-terminals MEMS with single fold spring; (b) 4-terminals with double fold spring MEMS.

in Fig. 3.13 (a-b). A lower pull-in voltage of (7v) can be obtained with the double spring than that of single spring, as indicated in Fig. 3.13 (a-b). However, the indicated results demonstrate that 4-terminal relays with double spring suffer from high residual stress which may affect their functionality of operation. In general, it is observed that 4-terminal relays with double spring tend to bend out of plane as the gate terminal approaches the body electrode.

Consequently, 4-terminal relays are more preferable for digital logic implementations even with the expense of larger pull-in voltage, and hence higher switching energy. Alternatively, Fig. 3.14 (a-b) shows our findings of simulating MEMS including an anchor relay and 4-terminal relay with opposite spring. This result emphasises that 4-terminal relay necessitates higher actuation voltage. On the other hand, anchor relays can be pulled-in with lower actuation voltage of (9v) and residual stress. In this thesis, the MEMS in Fig. 3.12 is adopted through our analysis in the next chapters. This is due to its higher recorded on/off

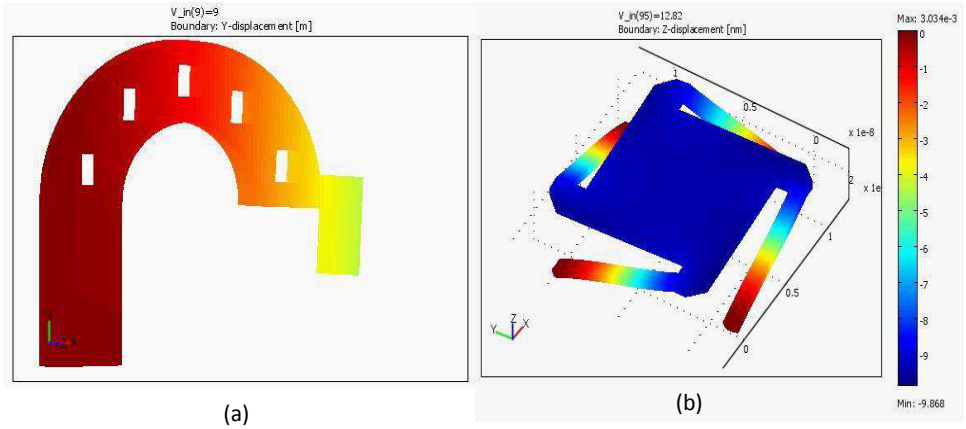


Figure 3.14: Demonstrates the 3D FEA of: (a) anchor MEMS [17, 137]; (b) 4-terminals with opposite-spring MEMS.

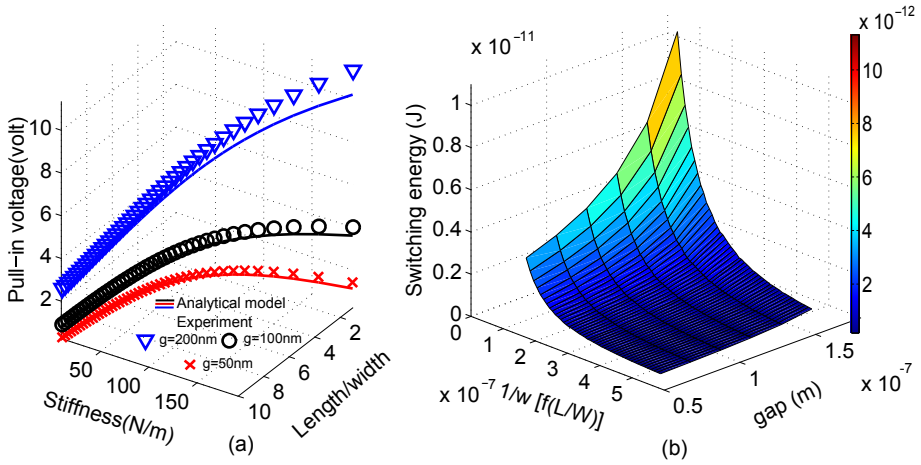


Figure 3.15: Shows the: (a) comparison of the pull-in voltage for three different gap distances obtained form 3D FEA and the analytical model; (b) switching energy at Q=1 based FEA of g_0=200nm and A=450um^2 as a function of (g_d) and resonant frequency (w).

switching cycles without exhibiting any failure. Furthermore, lower switching resistance of (R_{on} ≈1000Ω) are reported [156].

An extensive parametric sweep simulation is performed, in this work, to estimate the range of electo-mechanical parameters for both fabricated 450um^2 [156] and scaled 45um^2, and 4.5um^2 relays respectively, and thereby the energy-latency of MEMS can be optimised. These parameters can be seen in Table 3.2.

In order to obtain a precise analytical formula for the pull-in voltage, which is used later in Section 6.3, a sensitivity analysis coupled with parametric sweep has been performed, as shown in Algorithm 3.3. As a result, our analytical model for evaluating the pull-in voltage at various gap sizes demonstrates a close fit to the one obtained from FEA, as shown in Fig. 3.15. The following section describes how to evaluate energy-latency trade-offs.

3.4.1 *Structural stiffness*

The structural stiffness of MEM/NEM relays subjected to an electrostatic force is modelled using FEA. In this study, it is assumed that the MEMS

Algorithm 3.3 Pull-in analysis based FEA Parametric sweep.

Define: Spring width:=W, Spring length:=L, Actuation gap:=g.
Define: Constant: Actuation area(A), Gate thickness(h), Dimple gap:=g_d.
Output: (V_{pi})

1: Parametric sweep L = 5×10^{-6}:10^{-6}:5×10^{-5}.
2: Set W=5×10^{-6}.
3: Calculate $\frac{\partial V_{pi}}{\partial g}$, $\frac{\partial V_{pi}}{\partial W}$, $\frac{\partial V_{pi}}{\partial L}$, $\frac{\partial V_{pi}}{\partial g}$ (Sensitivity analysis)
4: $V_{pi} \simeq \sqrt[2]{\frac{\beta \times L g^3}{\varepsilon_0 W A}}$; $\beta = 3.87 \times 10^{-4}$.

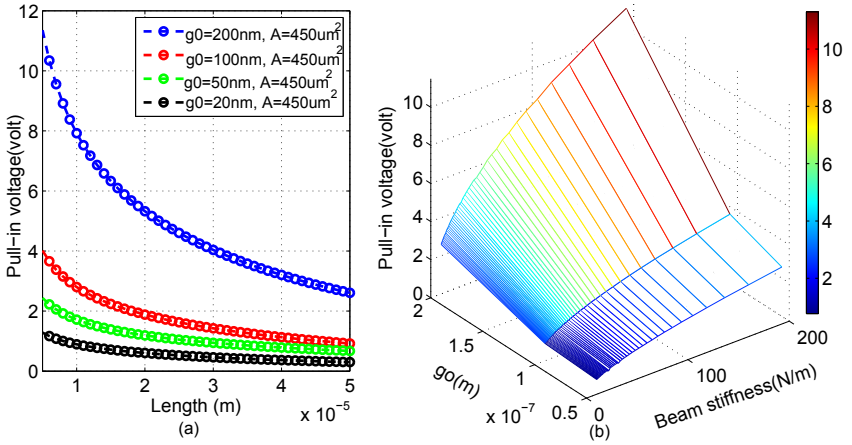


Figure 3.16: Shows the: (a) pull-in voltage for four different gap distance obtained form full finite element model; (b) pull-in voltage as a function of gap distance and beam stiffness.

exhibits a linear elastic deformation. To solve coupled problems with complex geometry, the Arbitrary Lagrangian-Eulerian (ALE) method was used by the COMSOL tool to obtain the equilibrium point between electrostatic force and mechanical structure. This method diverges as the MEMS/NEMS displacement approaches the pull-in point. This is attributed to the fact that this is the last point where, behind it, the MEMS collapses non-linearly. At this point, the electrostatic force is equal to the spring restoring force. Having calculated the pull-in voltage and corresponding displacement using the COMSOL tool, the structural spring constant can be evaluated as:

$$F_{ele.}\,|_{pullin} = F_{spring} \Longrightarrow k\,|_{structure} = \frac{V_{pi}^2\,\partial C(Z)}{2Z\,\partial Z} \quad . \tag{3.20}$$

By using the above equation, Fig. 3.16(a-b) shows the evaluated beam stiffness with the corresponding pull-in voltage for different gap distances. It is clearly shown that the pull-in voltage is highly dependent on the beam stiffness (k) and gap distance (g_d). For a simple analysis in this experiment, beam stiffness is measured by sweeping the spring length and width, while (h) is kept as a constant parameter.

3.4.2 *Damping Analysis*

Squeeze film damping is the most important damping component on the dynamic behaviour of MEMS/NEMS, especially at low ambient pressure [99]. Estimating the damping components of the MEMS/NEMS is a necessity for accurate analysis, especially at nanoscale sizes. This is attributed to the fact that the rarefied air in the gap dampen the movements of the mechanical parts. Consequently, it significantly in-

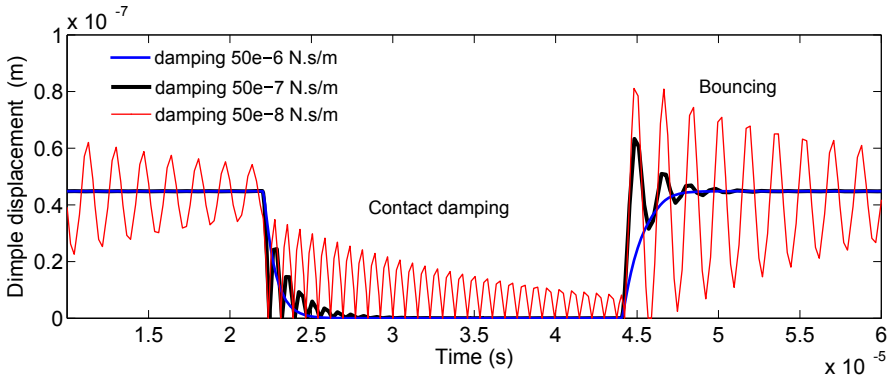


Figure 3.17: Impact of increasing structural damping coefficient on bouncing and contact damping based on a 4-terminal MEMS, A=450um^2, g_0=200nm, g_d=40nm, stiffness=150N/m, mass=0.29$\times 10^{-10}$kg.

fluences the switching time, mechanical quality factor, and impact bounce of the contact. Generally, squeeze film damping consists of viscous and electrical damping. Electrical damping due to air compression is often regarded as unimportant, especially at nano-scale geometry, and therefore it has been neglected in our analysis. Viscous damping is modelled using Rayleigh damping with the COMOSL tool as:

$$2\zeta_n\omega_n = \alpha_{dM} + \beta_{dK}\omega_n^2 \quad , \tag{3.21}$$

where α_{dM} dampens low frequency responses and β_{dK} dampens high frequency responses, while ω_n represents the natural frequency, and ζ is the damping ratio ($\zeta = \frac{1}{2Q}$).

For digital logic applications it is preferable to set $Q = \frac{\sqrt{mk}}{b}$ factor $\leqslant 1$, to avoid non-ideal switching effects such as a long settling time and contact bouncing. The effect of changing the contact damping coefficient (b) on MEMS contact damping is shown in Fig. 3.17. Introducing contact damping causes the bounce to fade away. Furthermore, increasing the damping coefficient removes increasing amounts of energy from the system, resulting in a corresponding reduction in MEMS bouncing.

3.4.3 *Natural and resonant frequency*

The natural and resonant frequencies of NEMS/MEMS are modelled by solving the 3D FEA models in the COMSOL tool with the frequency response solver. The frequency when the system vibrates naturally once it has been set into motion is called the natural frequency. The calculated natural frequencies using COMSOL tool of the relays in [10], [156], and [137] are equal to 1.2$\times 10^6$Hz, 146$\times 10^6$Hz, and 179$\times 10^6$Hz,

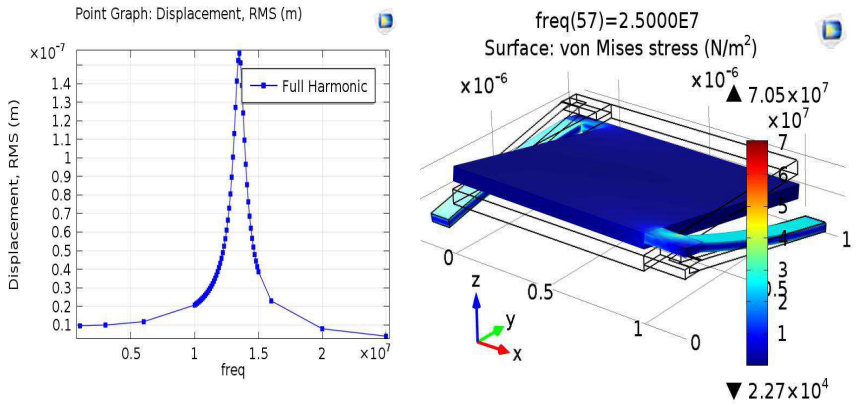


Figure 3.18: Resonant frequency of 2-terminal MEMS.

respectively. These values of frequencies represent the maximum allowable power gate switching as the MEMS will oscillate beyond these frequencies, thereby causing an operation failure. For demonstration purposes, Fig. 3.18 shows the displacement versus resonant frequencies for the 2-spring MEMS. It is clearly shown that as the natural frequency is approached, which is equal to 13×10^6Hz, the MEMS will vibrate and bend out of plane. It should be noted that for coherent analysis, these relays are simulated with comparable foot print size of $450\mu m^2$.

3.5 ENERGY-LATENCY ANALYSIS OF MEMS

The results in Fig. 3.15(b) show the switching energy consumption of MEMS by using 3D FEA as a function of the dimple gap (g_d), and resonant frequency (w). As can be seen, increasing (g_d) causes an almost linear increase in switching energy at low (w). Alternatively, switching energy increases exponentially with increasing resonant frequency (w)

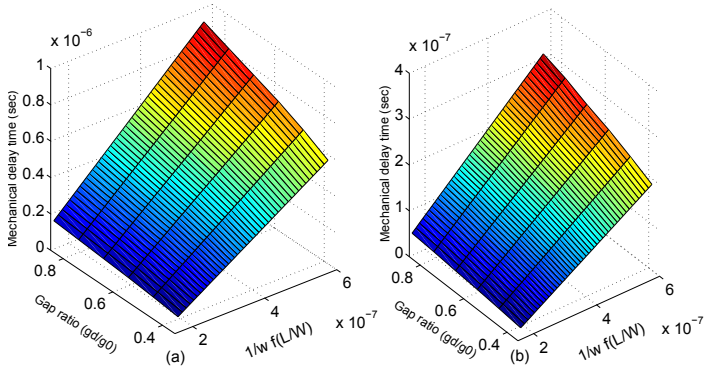


Figure 3.19: Illustrates that:(a) T_{mech} as a function of gap ratio and resonant frequency obtained from 3D FEA at Q=1 and V_{dd}=V_{pi}; (b) T_{mech} as a function of gap ratio and resonant frequency obtained from 3D FEA at V_{dd}=2V_{pi} .

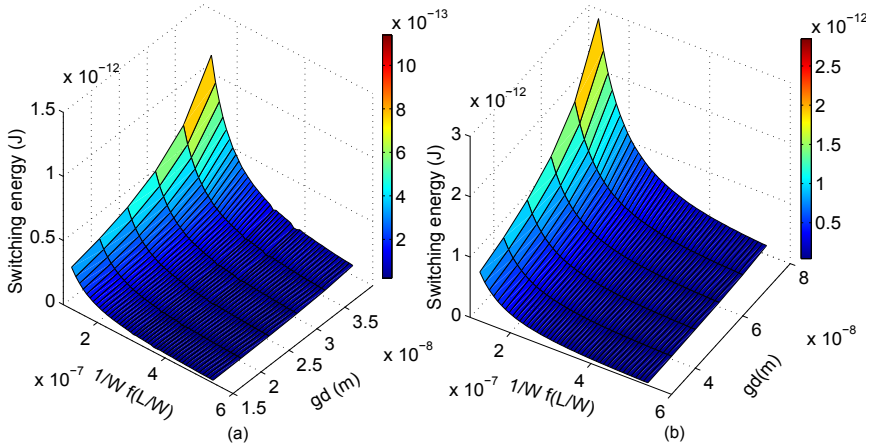


Figure 3.20: Shows the: (a) switching energy based 3D FEA at Q=1, g_0=50nm, and A=450um^2 as a function of g_d and resonant frequency; (b) switching energy based 3D FEA at Q=1, g_0=100nm, and A=450um^2 as a function of g_d and resonant frequency.

by sweeping the ratio of $(\frac{L}{W})$, at high (g_d). Fig. 3.19(a-b) shows the simulation results of mechanical delay time as a function of gap ratio $(\frac{g_d}{g_0})$, resonant frequency (w), and quality factor (Q). One observation which can be made is that T_{mech} is inversely proportional to (w), and it is linearly proportional with the increase in $(\frac{g_d}{g_0})$, which is consistent with the theoretical predictive equation in the previous study [156]. These results clearly indicate the trade-off between switching energy and mechanical delay time of MEM relays. As an example, it is found that, at $(\frac{g_d}{g_0})$=0.5, every ~3.5× increase in switching energy can be traded-off for a ~6× reduction in MEMS delay.

Fig. 3.19(b) shows the impact of doubling the value of actuating voltage on the mechanical delay of the MEM relay. It is clearly shown that increasing actuation voltage can significantly reduce the mechanical delay time by a factor of ×20 compared to the normal-actuated MEMS (i.e. V_{dd}=V_{pi}) when $(\frac{g_d}{g_0})$=0.4. However, 3D FEA simulations show that surface stress occurs, which might impact on the operation of the MEMS after a certain number of on/off cycles. Furthermore, this would exacerbate the stiction problem when the switch is required to be pulled-out due to the softening of spring stiffness.

Scaling down of the actuation gap (g_0) by factors of ×0.5, and ×0.25 can drastically improve the switching energy consumption of the MEMS, as can be seen in Fig. 3.20 (a-b). These results are obtained by performing an extensive 3D FEA coupled with parametric sweep analysis using the COMSOL tool. As can be seen, the MEMS with g_0=50nm can achieve ~8× and ~3× reductions respectively in switching energy consumption compared to the MEMS with g_0=200nm, and g_0=100nm. This comparison is obtained when $\frac{g_d}{g_0}$=0.5. With higher ratios greater energy savings can be achieved as indicated in Figs.

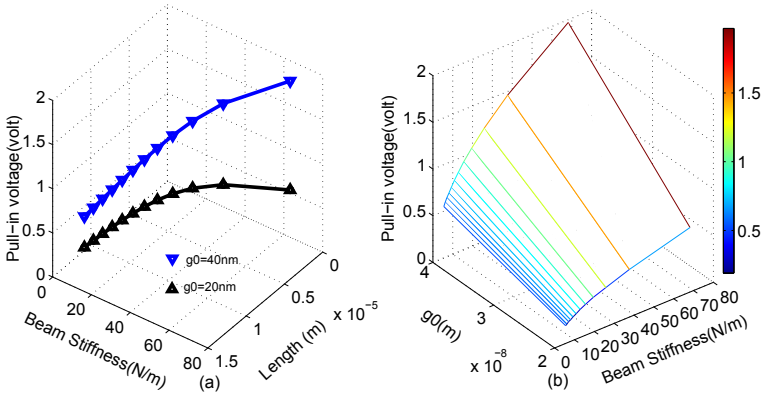


Figure 3.21: Shows the: (a) pull-in voltage for two different gap distance obtained form full finite element model at A=45um^2; (b) pull-in voltage as a function of gap distance and beam stiffness at A=45um^2.

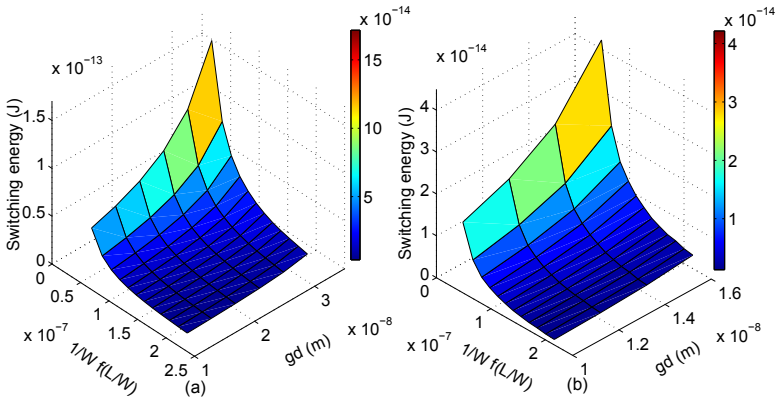


Figure 3.22: Demonstrates the: (a) switching energy based 3D FEA at Q=1, g_0=40nm, and A=45um^2 as a function of g_d and resonant frequency; (b) switching energy based 3D FEA at Q=1, g_0=20nm, and A=45um^2 as a function of g_d and resonant frequency.

3.20 and 3.15(b). However, this comes at the expense of increasing the mechanical delay.

3.6 IMPLICATIONS OF SCALING ON THE ENERGY-DELAY TRADE-OFF

Like CMOS, scaling down of MEMS parameters will lead to achieving greater energy savings. A constant field scaling methodology, analogous to the classic CMOS scaling, has recently been reported for the MEM relay [37]. Although this methodology provides useful insights into the benefits of MEMS scaling, however, it may not yield an optimal MEM relay design. Therefore, a 1-DOF optimization based on a variable scaling factors methodology has been proposed in previous work [82]. However, single DOF analysis can lead to an inaccurate outcome. As a result, 3D FEA coupled with the sweeping of scaled pa-

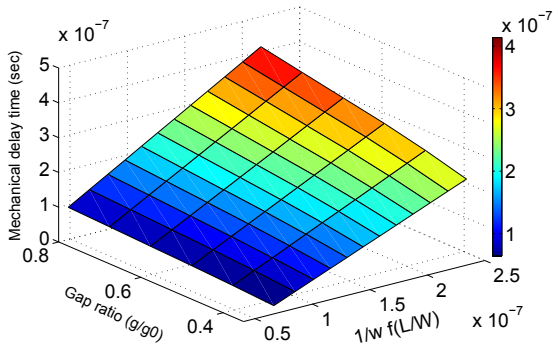


Figure 3.23: T_{mech} as a function of gap ratio and resonant frequency obtained from 3D FEA at Q=1.

rameters has been performed in the COMSOL tool. In our experiment, it is postulated that area (A), gate thickness (h), actuation gap (g_0) can be scaled by factors of $\times$0.1, $\times$0.5, and $\times$0.2 respectively. Other parameters such as g_d, spring width (W), and spring length (L) are kept variable by sweeping ($\frac{L}{W}$). Ultimately, some parameters will approach a lower limit and it may not be be possible to scale them down as readily as other parameters. For instance, g_0 is limited by nano-gap formation technology, while gate thickness (h) will be set by process technology constraints.

Fig. 3.21(a-b) shows the pull-in voltage of the scaled MEM relay as a function of spring stiffness and length. These results demonstrate a significant reduction in pull-in voltage which in turn leads to drastic reductions in the switching energy of the MEMS as can be seen in Fig. 3.22. Furthermore, these results clearly indicate a better trade-off between switching energy and mechanical delay time compared to that of the fabricated MEMS (A=450um^2). As an example, it is found that, at ($\frac{g_d}{g_0}$)=0.5, every ~5$\times$ increase in switching energy can be traded-off for a ~2$\times$ reduction in the MEMS delay, as can be seen in Figs. 3.22(a) and 3.23.

Aggressive scaling in MEMS parameters will result in drastic improvements in terms of pull-in voltage, switching energy, and mechanical delay time. Consequently, in this experiment the actuation gap (g_0), gate area (A), and gate thickness (h) are scaled further by factors of $\times$0.5, $\times$0.1, and $\times$0.5 respectively. Fig. 3.24 illustrates that, in principle, NEMS can operate at a pull-in voltage approaching 0.1V. As a result, significant improvements in switching energy can be achieved as indicated in Fig.3.25, and hence the minimum switching energy consumption at this technology node can be approached (0.1aJ).

3.7 SWITCH MODEL SIMULATOR BASED ON (FEA)

Fig. 3.26 shows the hierarchical structure of our proposed switch model simulator which is based on the lumped MEMS/NEMS param-

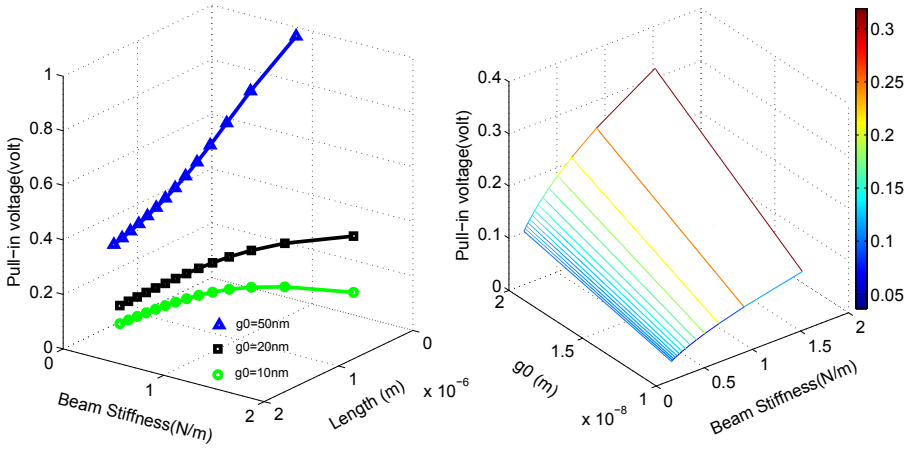


Figure 3.24: (a) Pull-in voltage for three different gap distance obtained form full finite element model at A=4.5um^2 (b) Pull-in voltage as a function of gap distance and beam stiffness at A=4.5um^2.

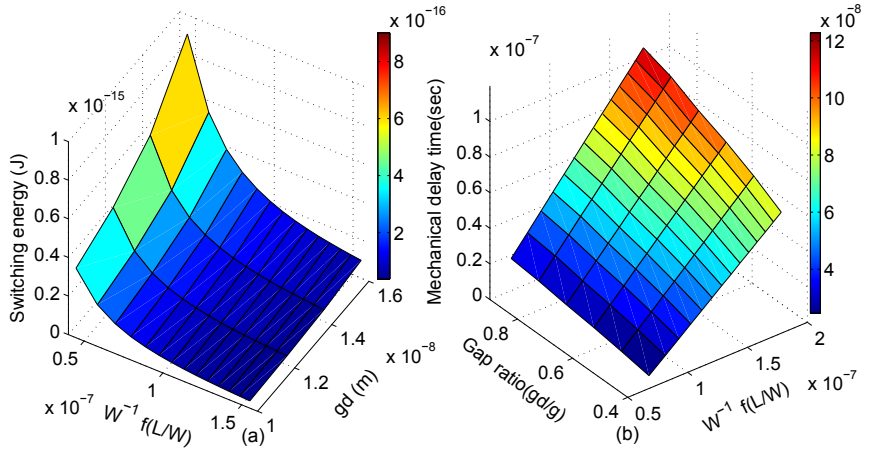


Figure 3.25: (a) Switching energy based FEA at Q=1, g_0=20nm, and A=4.5um^2 as a function of g_d and resonant frequency (b)T_{mech} as a function of gap ratio and resonant frequency obtained from FEA at Q=1 and A=4.5um^2.

eters. Lumped electrical parameters are either based on the fabricated MEMS (A=450 um^2) or predicted NEMS [156]. Alternatively, lumped mechanical parameters are evaluated by using 3D FEA, which is performed in the COMSOL multiphysics tool. As can be seen, at each gate voltage (V_g) step, the F_{ele} and F_{vdw} are evaluated by the simulator. As a result, the corresponding displacement (Z) of the gate terminal is generated and used as feedback of a new input to the design. This process is iterated until the dimple touches the drain-source terminal.

This simulator is characterized by the ease of simulation of various technology sizes by performing sensitivity analysis coupled with the parametric sweep. This can be achieved by using the equation in Algorithm 3.3 to evaluate the accurate spring constant at any given ratio of ($\frac{L}{W}$). Other mechanical parameters such as effective mass and

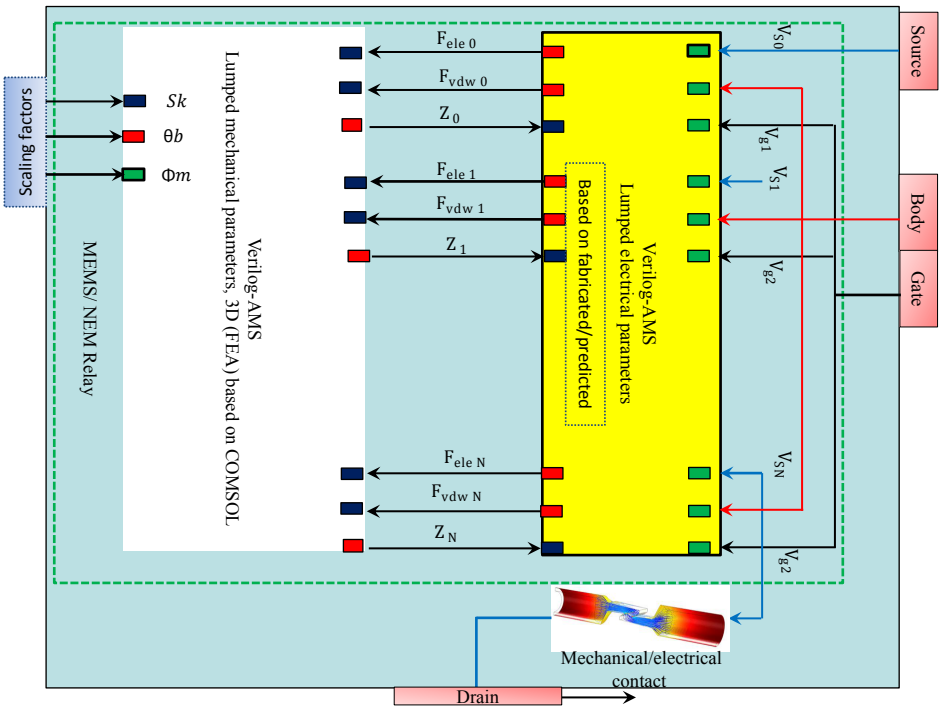


Figure 3.26: Graphic illustration of the hierarchical model of the switch simulator. The highlighted regions represent the electrical and mechanical lumped parameters, which is written in Verilog-AMS and co-simulated in The Cadance spice tool.

damping coefficient can be evaluated by using the same methodology within the COMSOL tool[55].

3.8 CONCLUSION

Scaling of CMOS has arrived at a point at which any further reduction in the threshold voltage of integrated circuits comes at the expense of higher power consumption. Furthermore, the supply voltage of the smaller node technology is not expected to be scaled less than the limit set by $\frac{kT}{q}$. Thus, an alternative device with a less energy-hungry and steeper sub-threshold regime like a MEMS/NEMS is necessary to be implemented in integrated circuits. MEMS/NEMS models are crucial to EDA design. However, the standard MEMS/NEMS model suffers from convergence problem and long execution time. This paper proposes a new simplified MEMS/NEMS model that can be used for large scale circuit simulation with insignificant error. The proposed model is evaluated with a variety of benchmark circuits and results show it can be adopted for simulating VLSI circuits with less than 6 % error rate. The proposed model can be integrated in the existing EDAs and can be used for simulating and designing of MEMS/NEMS-based VLSI circuits with better scalability and less execution time than the standard model. As we have demonstrated

in the experiments, the proposed approach could handle circuits comprised of 192 MEMS relays, while the standard model could not converge on circuits containing more than 22 MEMS.

This chapter also explores the modelling, designing, and optimizing of various MEMS/NEMS relays using 3D FEA exercised by the COMSOL multiphysics simulation tool. Furthermore, a 3D FEA-based MEMS simulator has been proposed to verify the accurate evaluation of mechanical parameters. A parametric sweep analysis is used to derive an analytical formula for beam stiffness which can be utilized to simulate various sizes of MEMS by only multiplying these parameters by a certain scaling factor.

Table 3.2: Current and scaled MEM/NEM relay physical parameters based on COMSOL multiphysics tool.

MEMS Area (um^2)	Pull-in voltage (volt)	Switching energy (pJ)	Mechanical delay (us)	Stiffness (N/m)	Mass (pg)	Viscous damping (uN.s/m)	Actuation gap (nm)	Actuation Capacitance (fF)
450	11.3-2.6	0.1-3.2	0.15-0.69	10.14-192.6	1.1-2.9	50	200	40
45 (NEMS)	0.19-1.97	0.049-0.003	0.06-0.28	5.51-68.2	0.15-0.25	0.07	40	17
4.5 (NEMS)	0.1-0.31	(0.037-0.36)$\times 10^{-3}$	(24-85)$\times 10^{-3}$	0.15-1.51	(3-3.77)$\times 10^{-3}$	0.007	20	3.54

4

MEMS-BASED POWER GATING OF ASYNCHRONOUS MICROPIPELINES FOR ULTRA LOW ENERGY DRIVEN COMPUTING

Chapter 3 presented the design, optimization, and scaling of MEM relay using 3D FEA performed by COMSOL multiphysics simulation tool. It was shown that scaling of MEMS parameters can significantly reduce actuation voltage, and hence switching energy will be minimised. Power reduction is a key design objective for embedded devices to extend battery life and is one of the major aims of this chapter. The impact of data rate and design complexity on MEMS-based power gating of asynchronous micropipeline was examined. Furthermore, the effectiveness of the proposed design is demonstrated and compared with that of previous approaches.

4.1 INTRODUCTION AND MOTIVATION

Ultra-low-power digital circuit design has become a key challenge given the growth in demand for devices that consume the minimum energy possible, energy-constrained, with less emphasise on circuit performance. These devices including active radio-frequency identifications (RFIDs), implantable medical devices, and autonomous sensors for the Internet of Things (IOTs) [117]. Traditionally, the scaling down of supply voltage has been used to reduce the dynamic energy quadratically [59]. However, operating at low voltage introduces a significant increase in propagation delay, which in turn leads to longer task completion time resulting in significant increases in leakage energy. This is even exacerbated as the technology node is scaled down further past the 90nm transistor size. Furthermore, the eventuating delays increase a circuit's sensitivity to process-voltage-temperature (PVT) variations, causing difficulty in balancing clock trees and performing timing closures in VLSI synchronous circuits. Since the target applications in this chapter generally demand non-abrupt on/off activities of the exercised workloads, the idle time between operations is non-trivial which in turn leads to more energy consumption.

To alleviate idle energy consumption, power gating techniques [140] as well as body biasing approaches [35] were used in both asynchronous and synchronous circuit designs. Of the two methods, power gating has been demonstrated to be an effective technique to mitigate leakage power consumption during the idle period. Typically, sleep transistors are inserted between the main power supply rail and combinational circuit (CC). However, these transistors themselves

contribute high leakage current making the reduction in idle energy of energy-constrained applications more challenging. In most synchronous designs, switching circuits off during idle periods is usually performed by adopting shut-off instructions within the program code. Therefore, it is difficult to employ this approach with circuit systems which operate without using instructions.

Although literature reviews have indicated that a recent proposed approach that performs power gating during active computation time within the sub-clock cycle [108], the synchronous circuits will still require proper timing analysis. It is essential to ensure that the hold and set-up time conditions are satisfied during circuit switching on/off. In asynchronous circuit designs, however, local handshake protocols may be used as control signals to power gate during the idle period, as these control signals specify when the circuit starts and ends computation. A plethora of research has been conducted to investigate power gating in asynchronous designs such as: in [102] the request signals of a four-phase bundle data are utilized as a control signal to power on/off the idle circuits in each stage of a micropipeline. A further improvement has been proposed to power gate both the combinational circuit CC block and latches within each stage of the pipeline [86]. The latter approach proposes monitoring the states of the adjacent blocks, and thereby shutting them off when required. Traditionally, depending on how the delay line (DL) block is implemented in asynchronous designs, this can be a major source of a significant amount of leakage current. Therefore in the present work, zero-leakage MEMS-based devices are employed to power gate each combinational circuit (CC) block as well as the delay line (DL) block (i.e. MEMS provides free mechanical delay time) in micropipeline stages. Furthermore, the conditions where MEMS-based power gating circuitry can achieve greater energy savings than sleep transistor counterparts are investigated, including questions of design architecture, and behaviour of operation.

Most of the work in the literature of the asynchronous power gating concentrates only on savings in leakage energy without counting the amount of energy overhead caused by the power switch circuitry on the overall energy consumption [86, 102]. In this chapter, therefore, the total dynamic and leakage energy consumed by the power gated circuit (i.e. CC and DL blocks) as well as the energy dissipated by the power gating circuitry have been evaluated and analysed. Furthermore, this work addresses the limitations and drawbacks of the previous work utilizing MEMS relays for energy-constrained implementations. Since these studies are either based on theoretical demonstrations [47, 64] or showing a lack of the model developed and simulation environment utilized [62]. Therefore, a novel work based on FEA is presented to target applications which exhibit low duty cycles as well as bursty computation workloads. It should be noted that, depending on the data rate (target throughput) and design architecture, there

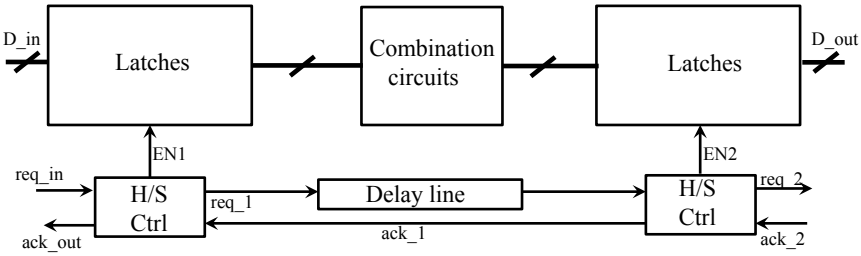


Figure 4.1: A conventional asynchronous-logic 4 phase bundle data micropipeline [102].

can be decreasing or increasing in the total energy consumed by the power gated circuits compared to cases with no power gating approach. This work investigates the extent at which power gating in the asynchronous micropipeline based MEM/NEM switches can be beneficial at different data rates and level of design architecture complexity.

4.2 ASYNCHRONOUS PIPELINE STAGE

The asynchronous micropipeline, as shown in Fig. 4.1, provides an event-driven circuit style utilizing localized handshaking signals to transfer data through the pipeline sequentially from one stage to another [159]. When the input data is ready to be sent, a request signal req_in will be generated resulting in enabling latches to capture the input data through EN1 and to raise the output request signal req_1. Consequently, the previous stage waits for an acknowledgement signal from the next stage, which is asserted once the data is computed and ready to be stored into the latches. While the latched data in the previous stage is being processed by the combinational circuits, the request signal req_1 is also passing through a matched delay element. The delay of the DL block is estimated to be not less than the worse case delay of the combinational circuits. This is to guarantee that correct data values are captured by the latches.

4.3 POWER GATING IN ASYNCHRONOUS MICROPIPELINE

4.3.1 *Conventional power gating in asynchronous micropipeline*

The req_1 signal, as shown in Fig. 4.2 (a), is used as a sleep control signal to turn-off the CC block during its idle state, or to turn it on during its active period. This will lead to the use of so called "just-in-time" computation. Consequently, leakage current can be reduced during the inactive period, as illustrated by T_{off} in the timing diagram shown in Fig. 4.2 (b). When the data is asserted, the request signal req_1 is raised to high and passed simultaneously to both sleep transistors and the DL block as a turn-on signal. This leads to powering the CC block,

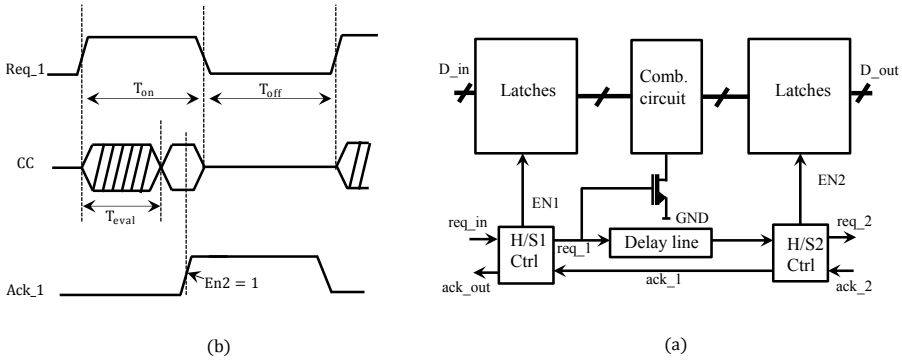


Figure 4.2: Shows the: (a) power-gated asynchronous micropipeline [102]; (b) timing diagram of the conventional power-gated micropipeline stage.

and thereby the incoming data is accepted from the data bus during T_{on}. The time required for the CC block to compute the incoming data and give stable output values is denoted as T_{eval}. As req_1 is passed through the DL block, the handshaking controller 2 (H/S2 Ctr2) generates the enable signal EN2 for the latches and Ack_1, which is transmitted back to the handshaking controller 1 (H/S1 Ctrl). Consequently, handshaking controller 1 (H/S1 Ctrl) de-asserts the req_1 signal once the Ack_1 signal is received, and thereby shuts-off the CC block. It is reported that a reduction of about 70% in the leakage energy dissipation can be achieved by utilizing this approach compared to one without a power gating technique [102]. It should be noted that this study only evaluated leakage energy reduction without taking into account the expense in energy overhead caused by the effect of the power gating circuitry on the total energy dissipation. As a result, this approach can only be energy-efficient when the energy saved is much greater than the energy consumed in switching the power switch network (PSN) on and off.

4.3.2 *Power gating in the asynchronous micropipeline based on adjacent-state monitoring*

In this approach, further improvements can be achieved by power gating both the CC blocks and the latches. Based on monitoring the status of adjacent stages, the middle block can be accordingly either switched on or off, as can be seen in Fig. 4.3 (b). For example, the power transistor of the CC1 block is controlled by monitoring the status of latches 1 and latches 2 of the adjacent stages of the micropipeline, and thereby the CC1 block can only be turned-off when both latches 1 and latches 2 are off and closed. To that end, the power transistor of latches 2 is controlled by employing logic OR on the on/off status of the CC1 block and the CC2 block in the next stage of the micropipeline. As a result, the idle time of the CC block is slightly reduced compared

with that in the previous approach [102], as can be seen in Fig. 4.3 (a). However, increasing the on-state time even when there is no computation results in high leakage energy consumption. Therefore, this approach can only be more energy-efficient than previous work [102] when the energy savings by power gating the latches are greater than the summation of increased leakage energy in the CC block and the energy overhead of complex power gating circuitry.

4.3.3 *Power gating in the asynchronous micropipeline by controlling the status of delay line blocks*

Typically, power gating can only be an efficient approach when the duration of the idle state is long enough to outweigh the increased energy overhead due to recharging the CC block as well as the power gating circuitry. In previous work, the delay line (DL) blocks (which are typically large) still consume a significant amount of energy even when the CC block is in the idle state. This approach, therefore, proposes to power gate both of the CC and DL blocks [121]. Consequently, the DL block of each micropipeline stage is added to the corresponding power domain, as shown in Fig. 4.4 (a). Simulation results showed that the DL block consumes a significant amount of energy ranging from 12% to 5% of the total energy at 1 to 200 KHz respectively [121].

4.3.4 *Proposed MEMS-based power gating in asynchronous micropipeline*

Further improvements to the previous paradigm can be achieved by employing MEMS/NEMS to power gate both the DL as well as CC blocks, as can be seen in Fig. 4.4 (b). The following two benefits can be observed by adopting MEMS/NEMS. Firstly, unlike sleep transistors, MEMS/NEMS switches exhibit zero leakage current. To avoid any performance degradation, especially in ultra low power applications, the

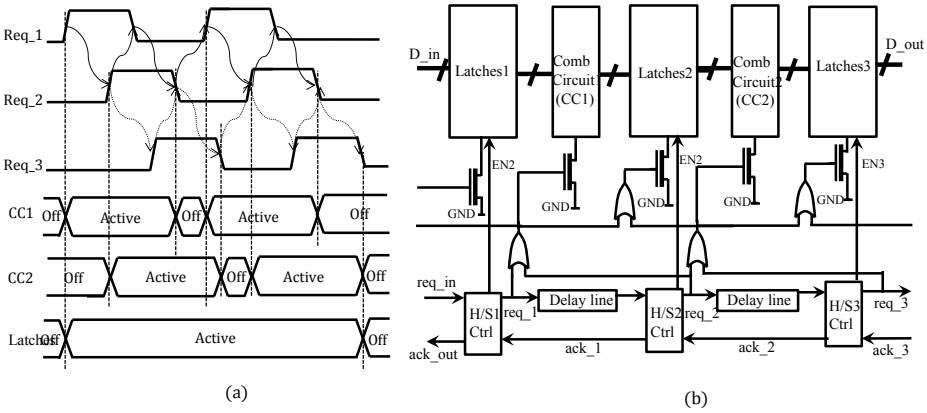


Figure 4.3: illustrates the: (a) timing diagram; (b) power-gated asynchronous micropipeline based on state monitoring approach [86].

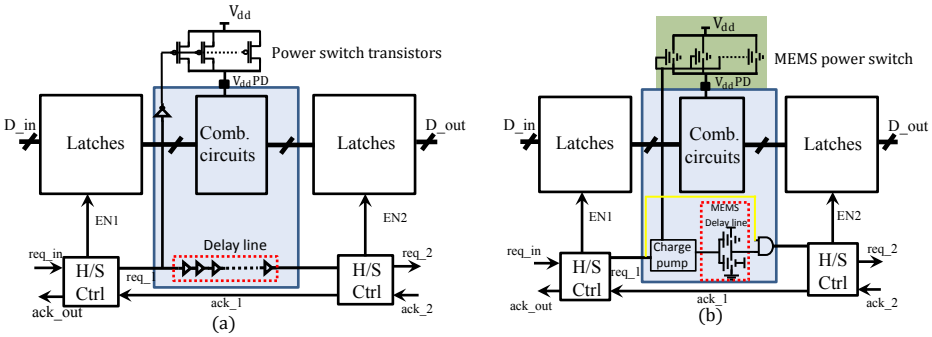


Figure 4.4: Shows the: (a) asynchronous power gating in[121]; (b) proposed MEMS-based synchronous power gating.

width of sleep transistors has to be made larger, which in turn leads to significantly greater leakage current. Finally, fewer MEMS/NEMS than CMOS counterparts are required in both the DL block as well as the PSN, due to its lower R_{on} than the sleep transistor counterpart. When req_1 is de-asserted, the DL block will be floating due to the disconnectivity period caused by the long T_{mech} delay to turn-on the other MEMS/NEMS in the buffer circuit. Therefore, employing the AND gate, as can be seen in Fig. 4.4 (b), is essential to ensure that the output of the DL block is always driven by either logic low or logic high and that it is never floating. Using FEA, T_{mech} of the MEMS/NEMS in the DL block can be carefully tuned to be higher than the summation of the T_{mech} of the MEMS/NEMS in the PSN and the worst case computation delay time of the CC block. The next section describes in detail how to hide the overhead of mechanical delay time in the asynchronous micropipeline.

4.4 ON-CHIP CHARGE PUMP FOR MEMS IMPLEMENTATIONS

An on-chip charge pump (CP) is required to switch on/off the current existing technology of the MEMS (A=450 μm^2). This is attributed to its high switching voltage, as shown in Table 3.2. Typically, charge pumps are used in integrated circuits (ICs) devoted to special kinds of applications, including non-volatile memories, RF antenna switch controllers, operational amplifiers, and piezoelectric actuators. Adding a charge pump, however, incurs the design with an extra energy, latency, and area overhead. In this work, the traditional dc-dc converter has been optimised to meet the requirement of the pull-in voltage for the existing technology of MEMS. The design parameters involved in the simulation of CP are as follow: V_{dd}=1.2V, V_{out} = 6V, f=40 MHz, α =0.1, I_L=10^{-15}A, and hence current consumption of the MEMS

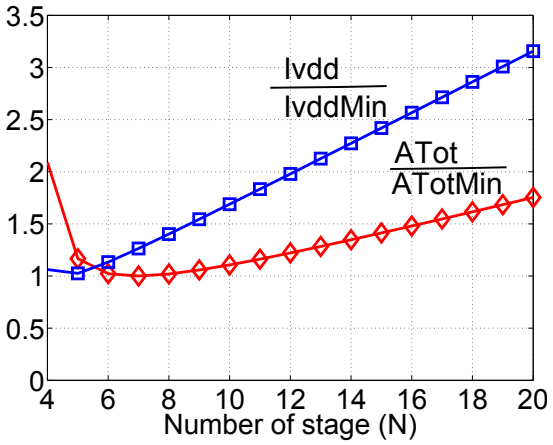


Figure 4.5: Current consumption and area normalization to their minimum values versus N.

has been proven to be in the range of 10^{-15}A [38]. Consequently, the charge pump capacitance can be calculated as follows [123]:

$$C = \frac{N_{opt}\ T\ I_L}{(N+1)\ V_{dd} - V_{out}} \quad . \tag{4.1}$$

Fig. 4.5 illustrates the optimized number of stages, for the given design parameters, versus area and current consumption. The number of stages which are required to deliver the pull-in voltage, by adopting the methodology presented in [123], has been calculated to be approximately of 6 as can be seen in Fig. 4.6.

Due to the hysteresis behaviour characteristic of pull in phenomena [156], once the MEM relay is switched on a lower voltage level is required to keep it in the on state. This hysteresis window depends on the device parameters. Consequently, an adaptive charge pump that changes its number of stages when the output voltage approaches a

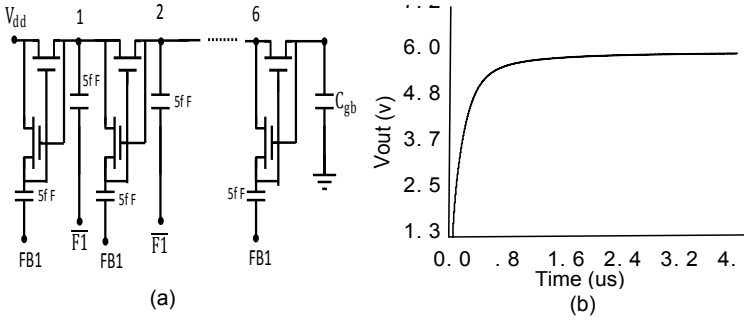


Figure 4.6: Illustrates the: (a) schematic circuit of the simulated charge pump; (b) simulated output voltage with respect to time for N=6 stages and φ = 40 MHz.

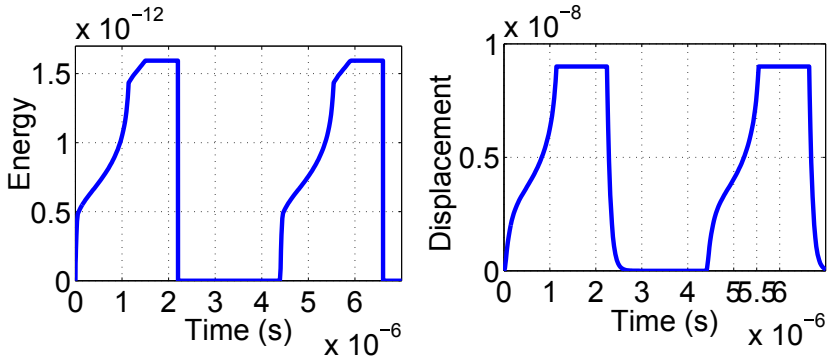


Figure 4.7: Impact of charge pump on the MEMS relay energy-latency.

certain level could be advantageous for a more energy-efficient design [124].

Adaptive charge pump has recently been proposed for MEMS-based implementation to reduce the power consumption By self-adjusting of the pulse amplitude voltage [23]. However, this method can lead to incur the design with complex hardware, and thereby increasing circuit area and energy overhead.

Previous work has investigated the energy-latency overhead caused by adopting the MEMS without taking into account the on-chip charge pump impacts on the model [137]. Consequently, for a more accurate analysis, this study addresses this issue as can be seen in Fig. 4.7. Adopting a charge pump increases the delay overhead by (0.5 µs), while it has a slight impact on the energy overhead (0.5 pJ).

4.5 EVALUATION SETUP

To validate the proposed approach as well as to find the conditions at which power gating in asynchronous micropipelines using MEM/NEM power switches becomes beneficial, 8- and 32-tap FIR

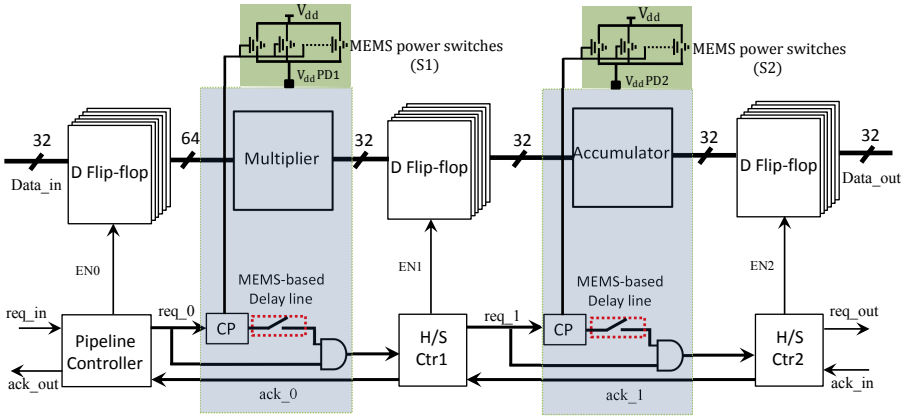


Figure 4.8: Asynchronous 32-tap FIR filter implemented based on the proposed approach.

filters were designed. These FIR filters are implemented using 90nm CMOS technology node, as shown in Fig. 4.8. They consist of two combinational blocks including an accumulator and multiplier. Each combinational block is powered by its own power domain, so that the multiplier is powered by power domain_1 (V_{dd}PD1) while the accumulator is powered by power domain_2 (V_{dd}PD2). These two power domains are powered by the main power supply voltage (V_{dd}) through an array of power switch networks (PSN) including S_1 and S_2, as indicated in Fig. 4.8. The delay line (DL) blocks in each stage of the micropipeline are connected now with the power domain of the corresponding combinational circuit (CC) block. The acknowledgement signal `ack_out` is only generated when the controller receives a request signal `req_in` and performs N handshake cycles with the micropipeline. Consequently, when the `req_0` signal is placed high the switch array S_1 will turn on, thereby connecting the supply voltage V_{dd} to power domain_1 (V_{dd}PD1). This leads to powering up the multiplier and its corresponding delay line (DL) block. One branch of the `req_0` signal bypasses the delay line (DL) block and will be ANDed with its delayed signal, resulting in generating the signal to be passed into handshaking controller_1 (H/S Ctrl). The power domain_1 can be powered down when the `req_0` signal is de-asserted, and this leads to placing one input of the AND gate in the floating state. However, the output logic of the AND gate will be set at logic low by `req_0`.

The flowchart of our evaluation process is presented in Fig. 4.9. Firstly, MEMS switch has been modelled and designed using 3D FEA performed by COMSOL multiphysics tool. Pre-defined specifications including dimension, geometric shape, material, and actuation type are used as an input to the COMSOL. As a result, the evaluation of the mechanical and electrical lumped parameters can be then obtained by performing frequency, transient, and parametric sweep analysis. It is worth mentioning that, in our work, modelling contact resistance of MEMS is based on the previous fabricated results [156], since it is only possible to model the contact resistance of the MEMS/NEMS under applying mechanical force, as described in [55]. From the other side, VHDL code is written for 32-tap FIR filter which consists of two combinational blocks 32-bit adder and 32-bit multiplier, as can be seen in Fig. 4.8. Using Synopsys simulator the area, dynamic and static energy has been optimized. Finally, the combined MEMS/CMOS can be simulated using the Cadence simulation tool.

4.6 RESULTS

Our approach was evaluated and compared with various set-ups in previous work [102] [121]. All these set-ups are powered by supply voltage with V_{dd}=0.6V. The number of PMOS sleep transistors used in the PSN are S_1=20 with width=4um, and S_2=15 with width=4um.

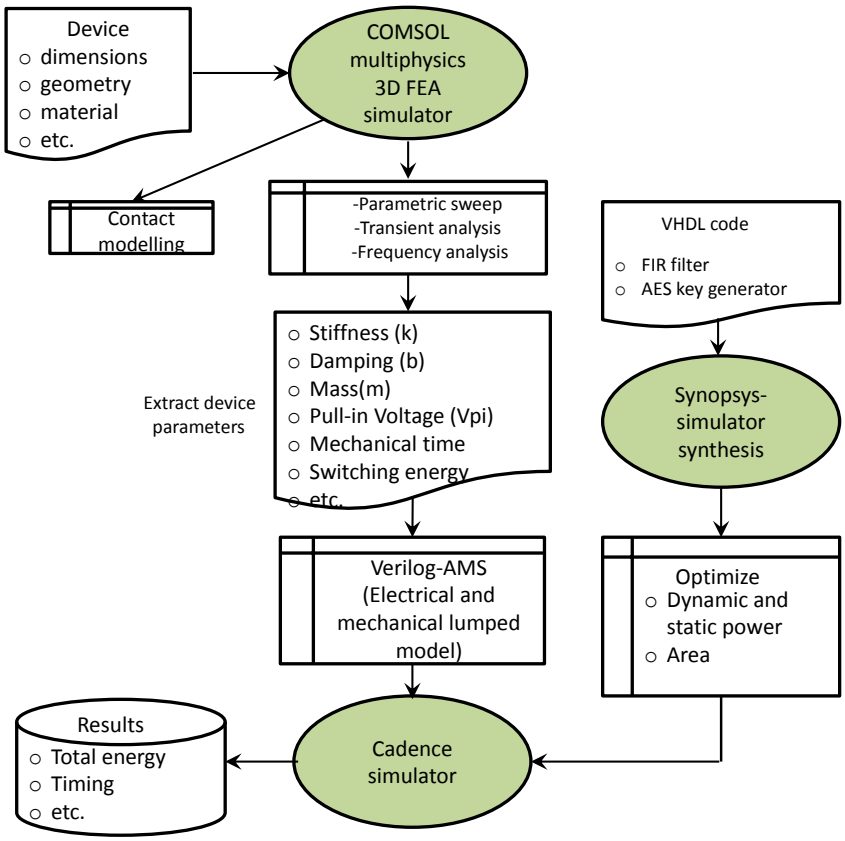


Figure 4.9: Flowchart of our evaluation process of the mixed MEMS/CMOS design.

Table 4.1: Total energy per computation for various asynchronous power gating configurations

	No PG		With PG in [102]	
Data rate (KHz)	Total Leakage (pJ)	Total energy (pJ)	Total energy (pJ)	Saving (%)
1	1351	2812	1940	31.0
10	134.3	283	217	24.9
100	12.90	41.0	33.0	18.14
400	3.277	19.8	20.6	-2.95
800	1.678	17.0	17.7	-7.56
1000	1.301	16.0	17.35	-8.40

Although MEMS relays have low R_{on} and their V_{gs} is independent of I_{ds}, compared with sleep transistors, it is assumed in our analysis to double the required number of MEMS in the PSN. As a result, the number of adopted MEMS in the PSN are S_1=10, and S_2=8. This conservative assumption is made due to the fact that R_{on} of MEMS will become higher after a number of on/off cycles and the unrevealed maximum current that MEMS can handle. The energy characteristics of the 8-tap FIR filer were evaluated by applying 6 sample points at the input. Consequently, the total energy dissipation was evaluated and energy per computation was recorded. Furthermore, the total leakage energy (i.e. caused by DL, and CC) of each set-up as well as the overall energy overhead of the proposed approach caused by adding MEMS relays and the charge pumps were evaluated. It should be noted that this filter only works for a range of data rates (Hz) which

Table 4.1 (continued): Total energy per computation for various asynchronous power gating configurations.

	With PG in [121]		Proposed PG	
Data rate (KHz)	Total energy (pJ)	Saving (%)	Energy (overhead) (pJ)	Total energy (pJ)
1	1601.9	43.0	3171	4640
10	172.6	39.0	336	464.9
100	30.9	24.0	52.5	81.70
400	18.6	7.0	28.8	45.90
800	16.8	-2.43	24.9	40.60
1000	16.5	-3.12	24.1	39.60

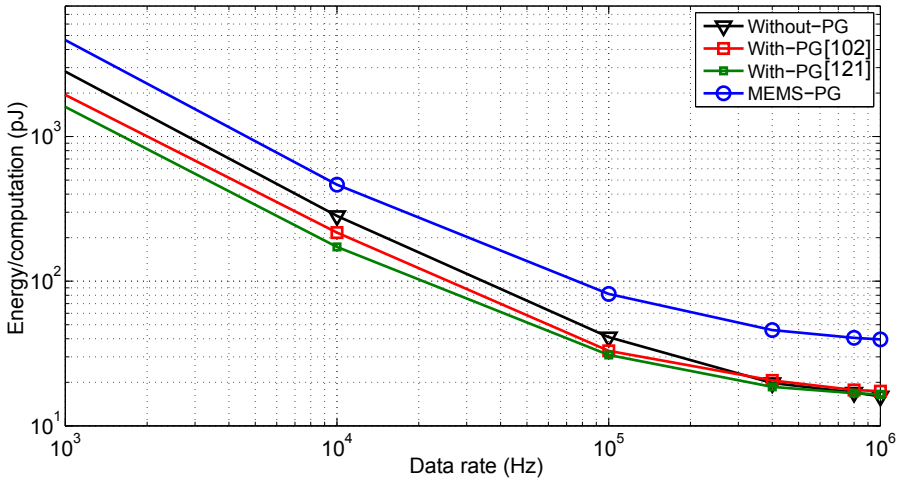


Figure 4.10: Shows the total energy consumption per computation for a various date rate with different power gating configurations

must not exceed the natural throughput of the filter at V_{dd}=0.6V. Table 4.1 shows the simulation results of the filter exercising at data rate with pulse width 50% (duty ratio (D)=0.5).

It can be deduced from these results that decreasing the data rate from 1MHz to 1KHz will result in an increase in the dissipated energy per computation of the four presented set-ups, as can be seen in Fig. 4.10. This is attributed to the fact that leakage energy in the circuit increases as the time required to complete a single computation increases, leading to a longer circuit idle time. Although MEMS relays exhibit zero leakage current, decreasing the data rate (i.e, with longer computation time) will lead to a significant increase in the energy consumption, as indicated in Fig. 4.10. This is attributed to the fact that increasing the switching energy of the MEMS-based power gating circuitry due to the long computation time will outweigh its leakage energy reduction benefits. It can be seen from Table 4.1 that large leakage energy can be significantly mitigated by power gating using sleep transistors. Conversely, our findings show that MEMS power switches may not be a perfect candidate to power gate in a similar fashion to CMOS sleep transistors. Therefore, three observations can be made from the results obtained in Table 4.1. Firstly, MEMS prefers a longer idle time to outweigh the high energy consumption of its power gating circuitry. Secondly, MEMS-based power gating circuitry prefers a bursty nature of operation (i.e, with short T_{on}). Lastly, in order to achieve greater energy savings, unlike sleep transistors, MEMS-based power gating circuitry favours the power gating of a complex CC block. This is due to the fact that driving a big capacitance load necessitates sleep transistors with a larger width, and hence leakage current significantly increases. Therefore, MEMS can achieve greater energy improvements compared with sleep transistors when the design architecture becomes complex and big.

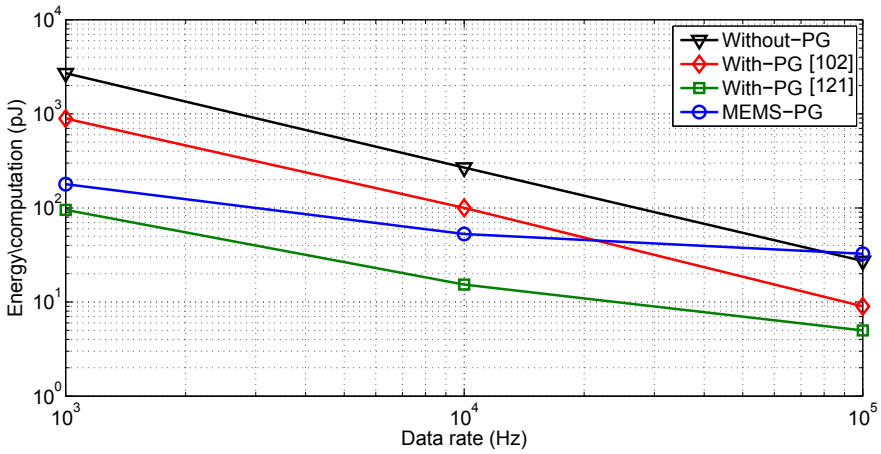


Figure 4.11: Shows the total energy consumption per computation for a various date rate with different power gating configurations at D=0.01.

Fig. 4.11 shows the simulation results of an 8-tap FIR filter performing at various data rates with pulse width of 1% (duty ratio (D)=0.01). These results indicate that significant improvements in the MEMS-based power gating paradigm can be achieved by reducing the duty ratio. Furthermore, the graph indicates that the MEMS-based power gating design intersects with the design Without-PG and With-PG [102] at about 92KHz and 40KHz, respectively. This means that beyond this points, further increases in the data rate (Hz) will result in negative savings since the energy consumed by the MEMS-based power gating circuitry will be greater than the savings gained by eliminating the FIR filter's leakage current. It is clearly shown that, even with the low duty cycle greater energy savings are still achieved by [121] compared to our proposed design.

Table 4.2 and Fig.4.12 show the total energy consumption of the 32-tap FIR filter implemented at various power gating set-ups and data rates. All these set-ups are powered by supply voltages with V_{dd}=1V. The number of PMOS sleep transistors used in the PSN are S_1=30 with width=8um, and S_2=20 with width=8um. The total energy consumption caused by the DL and CC of each set-up as well as the overall energy overhead of the proposed approach caused by adding MEMS relays and the charge pumps were evaluated. These results indicate that at low data rate our approach can achieve greater energy savings about 69% compared with the one without-PG and 29.5% with-PG [121]. This can be attributed to the significant increase of the leakage energy in the DL, and CC blocks. It is evaluated that the leakage energy dissipation of DL blocks equal to 0.68nJ at data rate about 1KHz. However, increasing data rate will lead to increase switching energy of the MEMS-based power gating which outweighs its leakage power reduction benefit. Therefore, our proposed approach achieve

Table 4.2: Total energy consumption for 32-tap FIR filter at various asynchronous power gating configurations

	No PG	With PG in [121]		Proposed PG	
Data rate (KHz)	Total energy (nJ)	Total energy (nJ)	Saving (%)	Total energy (nJ)	Saving (%)
1	6.778	4.10	39.5	2.10	69.0
10	0.80	0.49	38.0	0.30	62.5
100	0.60	0.41	31.0	0.31	48.3
400	0.45	0.38	15.0	0.315	30.0
800	0.389	0.35	10.2	0.32	17.7
1000	0.34	0.345	-1.4	0.32	5.80
10000	0.3	0.38	-26.0	0.45	-50.0

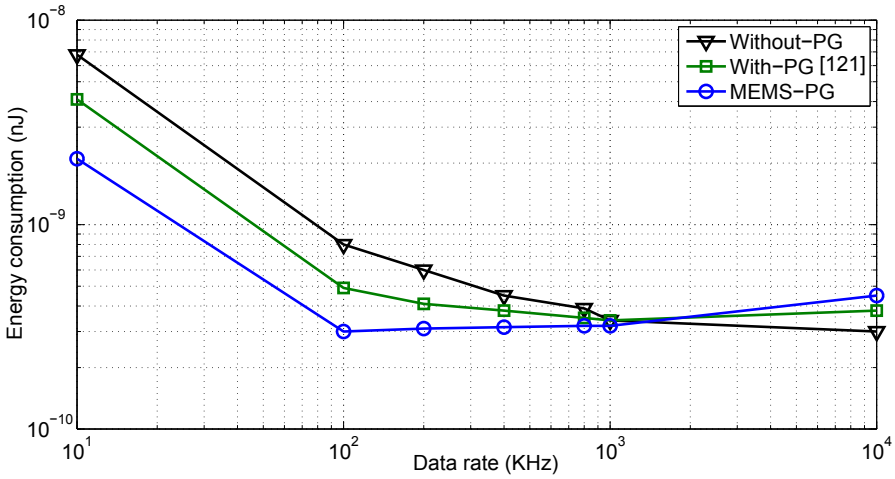


Figure 4.12: Shows the total energy consumption for 32-tap FIR filter at various date rate and different power gating configurations.

a negative energy savings when the data rate approaches 10MHz, as can be seen in Fig.4.12.

4.7 ZERO DELAY RIPPLE TURN ON (ZDRTO)

This section proposes a zero delay ripple turn on (ZDRTO) power gating control technique to hide T_{mech} latency of the hybrid MEMS-based and CMOS-based power gating in asynchronous micropipeline. This approach postulates that the T_{mech} latency of the downstream micropipeline stages, which is power-gated by MEMS, can be concealed by the computation delay time of the upstream stages, which is power-gated by sleep transistors. In synchronous designs, upon system wake-up, computation can only start after stabilisation of the

input voltage so that timing requirements are met. In the asynchronous micropipeline, on the other hand, ZDRTO kick-starts processing just as the system is waking up [122]. This enables the micropipeline to be shut-off when idle and to wake-up efficiently to process data without incorporating the T_{mech} overhead. As a result, employing MEMS-based power switch with ZDRTO in the asynchronous micropipeline can be an attractive approach for power-cycling the baseband processor which can be idle for milliseconds when waiting for correlated data from modules near the fronted [160].

Fig. 4.13 shows the power gating scheme of a 128-bit Advanced Encryption Standard (AES) asynchronous micropipeline based on mixed NEMS/PMOS power switches. In the present work, AES is chosen because of its wide data path, complexity, and low duty cycle. Since encryption engine are typically inactive for long periods of time. Typically, the AES round operation consists of four operations including add round key (AK), shift rows (SR), byte substitute (BS), and mix column (MC), as can be seen in Fig. 4.13. It is postulated that the small combinational blocks, add round key and shift rows, can be executing at low frequency to hide the mechanical delay time of the complex blocks, byte substitute and mix column, which preferably execute at high frequency. This necessitates the development of a tool to automatically cluster the micropipeline stages as well as to choose the best power gating techniques for each cluster by determining the static power requirements and maximum allowable wake-up time. This postulation is left for future investigation.

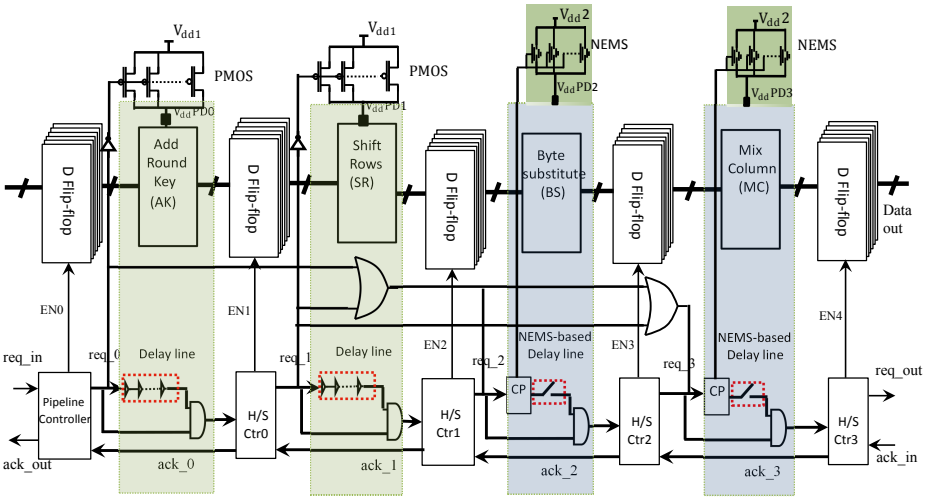


Figure 4.13: Shows the power gating scheme of AES asynchronous micropipeline based on mixed NEMS/PMOS power switch.

4.8 CONCLUSION

This chapter has presented an investigation into power gating techniques implemented on asynchronous micropipelines. This study demonstrated the threshold at which these techniques can achieve greater energy savings in relation to the design architecture and data rate of the input. Our proposed paradigm offers 69% energy improvements at a data rate about 1KHz compared to 39% in the previous paradigm.

5

MEMS-BASED POWER DELIVERY CONTROL FOR BURSTY APPLICATIONS

5.1 INTRODUCTION

The proliferation of digital VLSI devices has generally been associated with the development of high throughput microprocessors, owning mainly to progress in technology scaling over the last four decades. In recent years, however, the growth of portable battery-operated computing such as with cellphones, PDAs, wireless sensor networks, baseband processors and biomedical implants has become overwhelmingly popular in the VLSI market. These devices either operate at regular medium frequency (several hundreds of MHz) or in bursty usage patterns. Therefore, increasing leakage power in these gadgets due to a low duty cycle is a major challenge. This is attributed to the constraints of small batteries and energy harvesting systems as well as difficulties in replacing batteries.

Let us postulate a computing system like a biological system as an example of an energy-efficient portable gadget architecture, as illustrate in Fig. 5.1. Such systems typically have two types of operation: regular and bursty. Regular activities happen most of the time, and are intended to serve the needs of the entire system, and are determined by the dynamics of the system and the overall structure. Bursty activities are usually not those that are regularly initiated by normal periodic cycles, but rather are triggered by or in response with the demand to interact with external changes or conditions [9].

A battery-operated system, like biological systems, spends a non-trivial proportion of time in standby mode and then shifts into a regular or bursty active mode. As a result, leakage current can drastically lower battery lifetime. To that end, this work proposes a new paradigm to drastically reduce leakage current in idle periods while utilizing the dynamic voltage scaling (DVS) in the bursty/regular active mode. This paradigm has been adopted due to the fact that, in highly energy-constrained applications, voltage scaling and power gating are not two independent knobs [8].

Designing an efficient power delivery network (PDN) in a system-on-chip (SoC) that supports DVS is a challenging task. This is because the PDN must deliver power at appropriate voltage levels to different function blocks while incurring the minimum power loss when the voltage level of an function block (FB) is changed. Furthermore, this PDN must switch off the power completely when an FB is in the idle state. To that end, a PDN which incorporates an on-chip power grid,

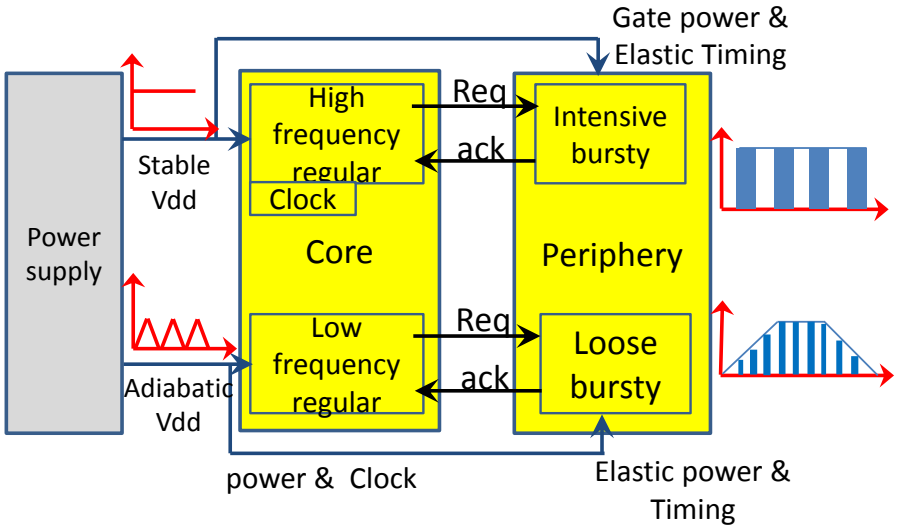


Figure 5.1: Proposed portable system architecture [9].

power conversion network, and MEMS power switch layer is proposed in this work. The major contributions of this chapter are to:

1. Propose an efficient power delivery network architecture to support applications with low-throughput, extremely energy-constrained, and bursty/regular operational patterns.

2. Develop a mathematical analysis that describes the impact of MEMS/NEMS and CMOS on the power consumption and supply voltage for given target throughputs.

5.2 BACKGROUND

5.2.0.1 *Relay-Based Power-Gating*

Despite its high mechanical delay time compared to other switches, the thermal actuated relay has recently been proposed for the power gating of throughput-aware applications [62]. Such applications include wireless sensor networks and biomedical implants. It has been shown that, for low throughput, the MEMS favour a complex and fast circuit architecture and leakage power is always zero. However, the results illustrated in this work are basically theoretical and no full explanation is provided of the model developed, the benchmark circuit, and the simulation environment utilized [151]. Other previous work has proposed a thermal switch with dimensions of 0.5 µm × 0.5 µm for power gating the baseband processor circuitry of a cell phone [138]. Consequently, the battery lifetime effectiveness was significantly increased by 50 % more than adopting the sleep transistor technique.

To overcome the drawbacks of thermal actuated switches, including high mechanical delay time and hot-spot production in the electronic circuit,[62] was adapted for an electrostatic MEM/NEM relay in power-

gating of highly scalable periodic and event-driven processing [63]. The results showed that a MEMS-gated processor is ideal for applications with more than 100 ms standby time. However, analytical results showed that the MEM relay can achieve greater energy reduction benefits than the CMOS for idle periods of $T_{off} > 10\mu s$ [47].

5.2.0.2 *Transistor Based Power-Gating*

Power gating is one of the most effective techniques to combat leakage current. This technique is implemented by adding sleep transistors either between the power source or ground network and the stacked logic gates. However, these sleep transistors themselves leak, which means that leakage current is mitigated but not totally eliminated. Furthermore, sleep transistors cause a performance penalty. Consequently, adopting power gating in ultra-low voltage applications may lead to inoperable circuits as the CMOS approaches its fundamental limits. Basically, a wider transistor causes high leakage (low R_{on}), while a smaller transistor degrades performance (high R_{on}). Therefore, a trade-off exists between area, leakage and performance.

5.3 PROPOSED ARCHITECTURE OF COMPUTER SYSTEMS

Biological systems usually have two kinds of operational behaviours, regular and bursty, and manage to regulate their mode of operation in robust and an energy-efficient way, which also maintains natural propensity for survival [171]. Regular activities occur most of the time, and are intended to serve the needs of the entire system. Bursty activities are not usually those constantly initiated by normal periodic cycles of the system, but rather in response to the demand of interaction with external changes or conditions. In the present study, it is postulated to build a computer system in a similar fashion of biological systems, such that all fast processing has to be built in specialised (periphery) units whose operation is bursty and slow processing has to be done in a constantly active part whose operation is regular. Thus, let's shift to the world of computer systems, and attempt to consider them as systems which we would like to lead balanced life and harmonious, and execute their operations in such a way that they can both (i) fulfil their own demands of remaining alive and functioning efficiently (regular activities), and (ii) fulfil the demands of their environment and/or user which pilots them with specific operation (bursty activities) [171].

As the first example, lets examine the functionality of a simple microprocessor. It is clear that the major instruction processing control has to be constantly active if its primary goal is to sustain life in all sections of the system at any time. But does it have to be executing the instruction control very fast? well, it should be fast if these instructions are driven by the demands of the environment and/or users. But if our presumption is that only bursty activities can be running fast, then

the environment and/or users should be bursty, and hence instruction processing control is not regular activities. In our proposed approach, where we should to be naturalistic about robustness and resource efficiency, we can assume that the entire system cannot have both regular operation and performing high rate activities [171].

A many-core multiprocessor system can be considered as our second example. In such a system, the responsibility of the regular activity organ allots presumably to the communication infrastructure and power distribution, which should always be empowered and prepared for action, while switching the cores on and off is depending on the demands generating from the application tasks. Conversely, creating the interconnect fabric slow is not permanently possible, and actually where we must to strive for throughput we have to empowered it for fast operation [171]. So, how the conflict between regularity and high speed can be resolved? well, some sacrifices should be made. To that end, a prominent amount of redundancy in the interconnect should be added, either to generate 'cold spare', which can be replaced when the hot sections fail, or have a dynamic reconfiguration system in which the artificial "quasi-burstiness" is introduced, thereby fast executing organs are given time to rest. This for instance could be achieved by certain time or space division demultiplexing [171].

A schematic depiction of the idea of an activity, regular and bursty, balanced system between regularly gating core and bursty pulsing peripheral core is shown in Fig. 5.1. It presents a general scenario in which two kinds of power supply are used, one adiabatic and the other standard with stable level of voltage (V_{dd}). This enables support of the performing of tasks that occur in different classes [171].

Regular tasks/assignments specify the "functional face" of the entire proposed system in terms of energy, reliability, and performance. They are framing the dynamics of the system corresponding to its main structural characteristics. Regular tasks can be categorized based on the level of granularity of periodicity into two types. Firstly, which is refereed to as "high-frequency regular assignments", and are postulated to be performed at the guaranteed level of timing and performance, and hence are regulated by clock while power level is constantly remained stable. Different approaches including voltage scaling can be implemented in this class to switch between various modes of the system level performance. Secondly, which is refereed to as "low-frequency regular assignments", and are performed in an adiabatic fashion, consequently their timing is controlled by the dynamics of the system's supply [171].

In comparison, bursty tasks/assignments are more sporadic and pulsed on demand as specified either by the necessity to certain kinds of computation functions pulsed by the commands from the core part (i.e performing instructions) or to interact to environmental stimuli. Bursty tasks can be categorized into two classes. Firstly, which is

refereed to as "intensive bursty", and are driven by the level of the power supply. Therefore, in this class timing elasticity is manifest due to various voltage levels and acknowledgement-request interaction with the core modules. Secondly, which is refereed to as "loose bursty", and which is enabling the maximum elasticity, because the delivered power level is not constant and is varied based on the waveform of the adiabatic power source. It should be noted that the timing of each cycle of computation functions is specified by the instantaneous level of the power supply waveform. Therefore, these assignments are somewhat elastic both in voltage supply levels and time [171].

In the present work, a novel power delivery network (PDN) using zero-leakage MEMS relays has been implemented to serve such a system, to capitalise on leakage energy reduction of burst activities, thus improving the total energy-efficiency. The next sections will explain our proposed PDN architecture that support bursty computation workloads.

5.4 VOLTAGE REGULATOR NETWORK OPTIMIZATION

In a complex system on chip (SoC) design, there are many function blocks (FBs) providing various functionalities. As an example of processing elements are CPU, DSP, and GPU. Examples of other function blocks (FBs) are RF front-end, random logic blocks, on-chip memory, custom signal (audio or video) signal processing blocks, and various controllers [11]. A voltage regulator module (VRM) design must meet the needs of all FBs which are powered-on by it. In an system on chip SoC with DVS option, the level of supply voltage of some FBs is dynamically adjusted in order to lower the total energy dissipation while meeting the performance demands [120]. An on-chip power controller decides when to switch the SoC power performance state (PPS) , where each PPS corresponds to a particular combination of voltage level (and associated clock frequency) assignments to various FBs in the SoC.

In the traditional technique to support DVS for various FBs, which is shown in Fig. 5.2, each function block has its own Voltage regulator VRM with multiple output voltage levels [22, 120]. The power switch controller selects the supply voltage level that VRM_i provides to FB_i.

This approach, despite its simplicity, has several drawbacks, including firstly, the number of voltage regulators VRM used in the PDN is equal to the number of function blocks FBs that can accept multiple voltage levels becomes large, and thereby causes to increase the chip cost and area. Secondly, designing a multiple output VRM is quite challenging and its cost is higher than that of fixed output voltage level VRM . Finally, the power conversion efficiency of variable V_{out} voltage regulator VRM varies as a function of the selected V_{out} and may degrade severely from one V_{out} level to another.

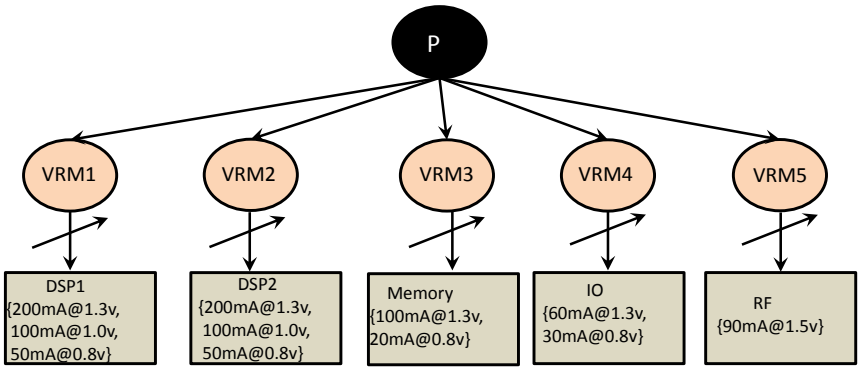


Figure 5.2: Shows the VRM tree in providing an appropriate voltage level for each function block (FB) [11].

Based on these observations, a new approach to address the problem of traditional PDN has been proposed [11]. In this approach, the PDN consists of two layers including power conversion network (PCN), where VRMs are used to generate all voltage level that may be required by various FBs in the SoC design. This is achieved by adopting fixed-V_{out} VRMs; so, if $\boldsymbol{u}$ is the set of all voltage levels needed by any FBs, then must there be at least $|\boldsymbol{u}|$ VRMs in the PCN. The second layer, a power switch network (PSN) is adopted to connect the power supply of each FB to the appropriate VRM output in the PCN [11].

5.5 PROPOSED ARCHITECTURE OF THE PDN

A new PDN that address some of the typical PDN's problems and supports DVS has recently been proposed in [11]. It was shown that the proposed architecture reduces the power losses of the PDN by 34 % while reducing its cost by an average of 8%. However, this paradigm has shortcomings in terms of supporting applications with low throughput and bursty operation. This is attributed to the leakage current of the CMOS transistor in the PSN, and especially the wider transistor used to meet the hight current requirements. Furthermore, the number of leaky paths between the power transistor and FB is increased. Therefore, MEM/NEM relays in the present study have been utilised and compared with their CMOS counterpart in terms of delay overhead and energy consumption for a different duty ratio, as shown in Fig. 5.3.

5.6 POWER SWITCH NETWORK ARCHITECTURE

The power switch network (PSN) performs the function of switching the supply voltage level of the FBs when a new PPS is commanded by the power manager. Fig. 5.3 depicts a PSN used to deliver various voltage levels to different FBs. The switches in the PSN are controlled by a power switch controller (PSC) which is zero-hot coded, i.e., at

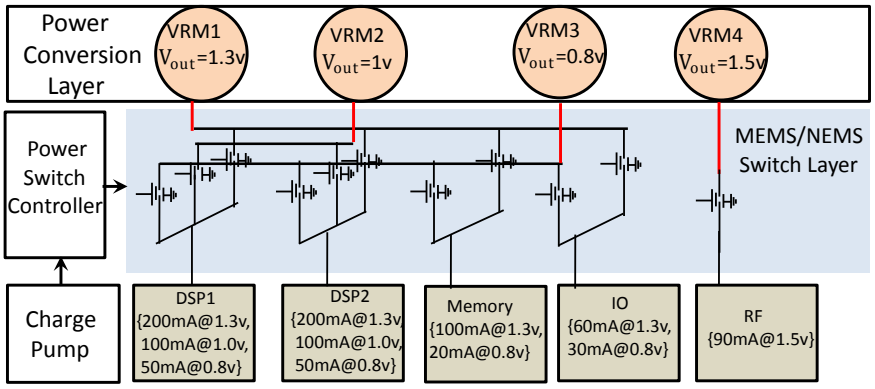


Figure 5.3: The modified architecture of the PDN to support low-throughput and Bursty operated applications.

any given time only one of its outputs is zero, and hence, only one MEMS/NEMS is on.

In this study, the Cadance simulations are used to optimise the width W(f,v) and number U(f,v) of PMOS switches which are required to deliver the voltage level v to a FB. While, a multiphysics data in Table 3.2 have been utilized to determine the number of parallel switches for the MEMS/NEMS power gate. As an example, Fig. 5.4 shows the optimized width and number of PMOS and MEMS to deliver the required voltage and current level to DSP1 at state S3.

5.6.1 *PSN energy consumption*

When the state of the system changes from PPS S_i to S_j, some energy is consumed to turn on/off some of the power gate switches. It is assumed that the power switch controller (PSC) changes the state of the system based on the state graph in Fig. 5.5. If C_M and C_R is the total capacitance which is charged or discharged during the transition

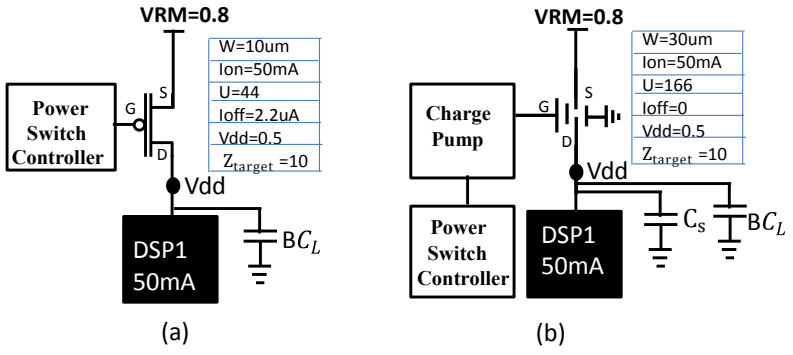


Figure 5.4: An optimised PDN that meets the target impedance (a) PMOS transistors; (b) MEMS.

for MOSFET and MEMS respectively, then the switching energy for this transition can be calculated as [156]:

$$E_S(R) = UC_R V_{gb}^2 = U\frac{\epsilon A}{g-z}V_{gb}^2 \quad , \tag{5.1}$$

$$E_S(M) = UWC_M V_g^2 \quad . \tag{5.2}$$

For a given amount of time that the CMOS logic is in active (regular or burst) or sleep modes, the energy per power gate switching cycle is:

$$E_R(V, R_t) = \frac{I_{on} V_{VRM}}{R_t} + E_S(R) + (\beta C_L + C_s)V_{dd}^2 \quad , \tag{5.3}$$

$$E_M(V, R_t) = \frac{I_{on} V_{VRM}}{R_t} + WUT_{off}P_{leak} + E_S(M) \quad + \beta C_L V_{dd}^2 \quad , \tag{5.4}$$

where the value of the external voltage regulator V_{VRM} is set by the desire on-chip supply V_{dd} and the drop voltage IR through the power gate switches:

$$V_{VRM} = V_{dd} + \frac{I_{on} R_{on}}{U} \quad . \tag{5.5}$$

R_t is the target throughput, C_s is adopted in our study to deliver current in a momentary period to an FB when a new PPS happens. This is attributed to the mechanical delay time that MEMS needs to switch to the new PPS.

5.6.2 *Power Switch controller(PSC)*

In our modelling framework, it is assumed that the transition of the system into different PPS can be described as a time-homogenous Markov chain. In each state, the supply voltage level of all FBs is specified as shown in the test bench of Fig. 5.5. In this study, a power switch controller based on Markov chain has been utilised to switch the PPSs of different FBs for 100 cycles.

5.7 RESULTS

Using Eqs. 5.3 and 5.4 with parameters from a standard 90nm complementary metal-oxide-semiconductor (CMOS) and Table 3.2, Fig. 5.6 shows the energy ratio versus T_{off} for fixed T_{on} (100 cycles). For

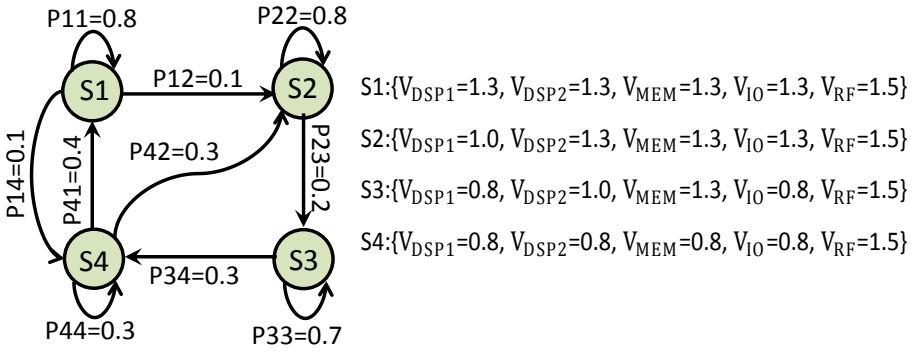


Figure 5.5: Test bench of operating states and state transition based on [11].

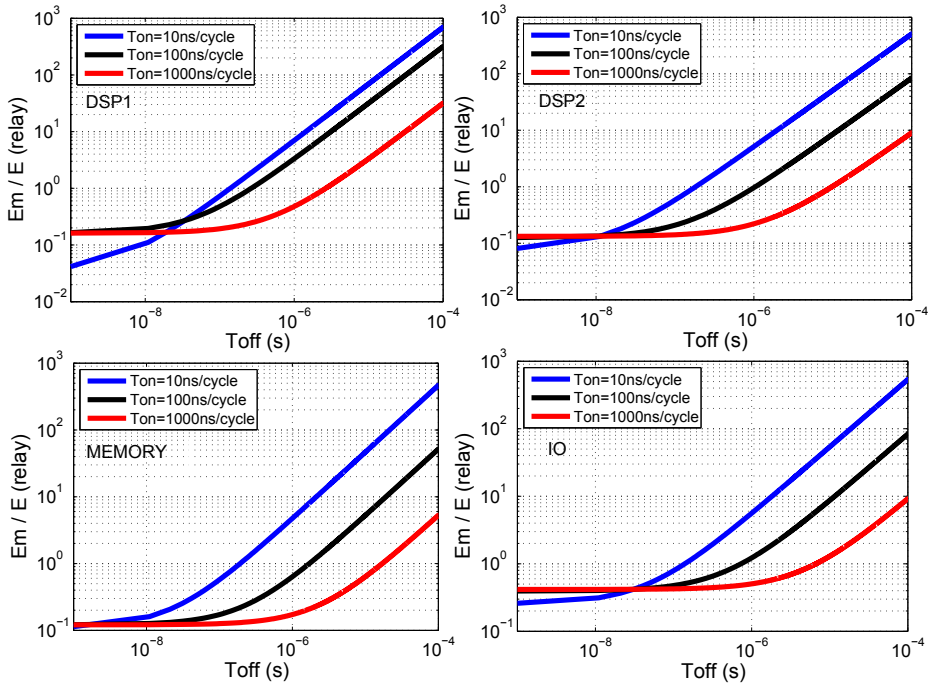


Figure 5.6: Energy ratio vs. T_{off}, for various T_{on}/cycle, for design power gated by a 90 nm PMOS transistor and current MEMS.

short T_{off}, for instant in digital signal processing (DSP)2 ($\leqslant 10\mu$), the increased switching energy of the relay based power gate and/or state changes outweigh its leakage power reduction benefit. However, even with the current existing MEMS, if T_{on}=1000 ns/cycle and $T_{off} \geqslant 10\mu$, the relay's negligible leakage constantly reduces the total energy as the T_{off} is increased.

As shown in Fig. 5.7, scaling the MEMS in order to reduce their parasitic capacitance and switching energy enables to begin accruing relatively high energy saving at a substantial lower T_{off} of 10^{-8} (about 100 $\times$ improvement). As a consequence, results in Fig. 5.7 clearly shown that power gating based NEMS is always energy-efficient regardless of target throughput.

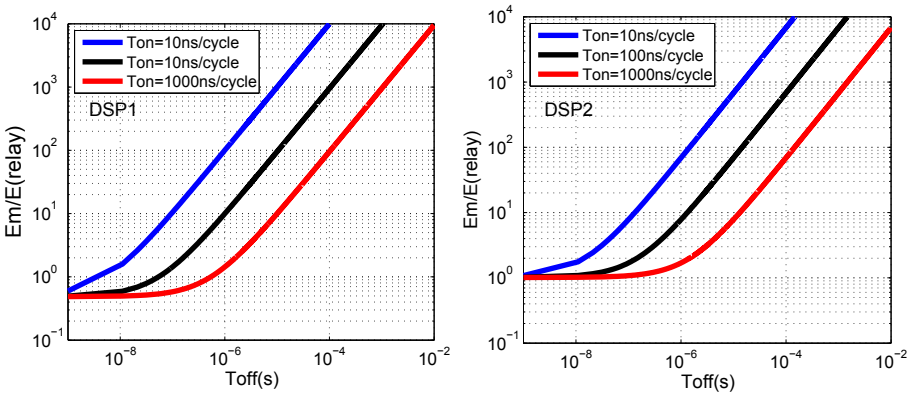


Figure 5.7: Energy ratio vs. T_{off}, for various T_{on}/cycle, for design power gated by a 90 nm PMOS transistor and scaled MEMS ($4.5\mu m^2$).

5.8 CONCLUSION

This work proposed an efficient PDN for battery operated devices which can drastically reduce leakage current in idle periods while utilizing the DVS in the bursty/regular active mode. This paradigm is essential for devices with mixed duty cycling, where some parts are required to work regularly with low-throughput while other parts are activated spontaneously, i.e. in bursts. Results have shown that the proposed PDN can achieve 1000x saving in energy compared to its CMOS counterpart for low duty cycle. However, even though adopting MEMS relay in the PSN adds only one mechanical delay overhead to the design, it produces the design no area overhead if the MEMS layer fabricated on top of CMOS layer using back-end of line process.

6

MEMS-BASED IDLE ENERGY MINIMIZATION FOR BURSTY WORKLOADS IN HETEROGENEOUS MANY-CORE SYSTEMS

Chapter 5 presented a MEMS-based power delivery control for bursty computation workloads targeting zero-leakage energy paradigm. Furthermore, a novel embedded system architecture that support bursty nature of operation was proposed. It was shown that significant energy reduction can be achieved under certain conditions, while incorporating in the design acceptable mechanical delay latency and no area overhead penalty if the MEMS layer is fabricated by using back-end on line process. This chapter examines the impact of the leakage energy savings of a heterogeneous many-core platform performing real applications, suitably chosen from a pool of available benchmarks, including memory-intensive, CPU-intensive, and other combinations. A novel non-invasive MEMS-based power gating design is presented using performance-energy states modelled through the feedback from performance counters. The aim is to eliminate the idle energy consumption of heterogeneous systems that exhibits either completely idle state scenario or executing burst computation workload, which necessitates to perform core allocation and dynamic voltage and frequency scaling (DVFS). The effectiveness of the proposed approach is compared with that of sleep transistors using different core allocations, operating frequency, and duty cycles.

It should be noted that this chapter contains collaborative work with Ali Aalsaud and others researchers in the micro system design group related to power and performance measurements from the Odroid platform. The outcome of this collaborated work is presented in the paper listed in publications on Page x. As the first author I am responsible for the majority of the research. Specifically, the topics indicating in Section 6.2.2 and 6.2.3 are collaborative efforts and the rest of the work described in this chapter pertains to my individual research.

6.1 INTRODUCTION AND MOTIVATION

The impetus of high throughput at low energy cost is at the core of design and implementation of many core embedded systems. To manage the trade-offs between throughput and energy an effective technique is to allocate heterogeneous computing resources on these systems. The Exynos 5422 big.LITTLE octa-core heterogeneous platform, which

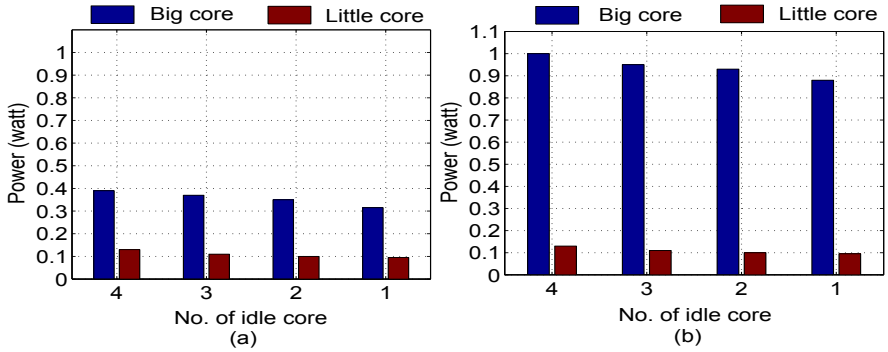


Figure 6.1: Experimental measurements of idle power consumption by adopting Odroid-XU3 big.LITTLE platform: (a) 1400MHz big.LITTLE; (b) 2000MHz big, 1400MHz LITTLE.

includes 4 big (ARM A15), and 4 LITTLE (ARM A7) cores, is a typical example[155].

Over the years significant research has been carried out to address energy minimization in heterogeneous embedded systems [172]. This research typically control the core allocation, coupled with DVFS decisions to react to workload variations [3]. When higher workload is encountered more numbers of cores are allocated with suitably determined DVFS. Conversely, when the workload is lower, fewer cores are executed with reduced voltage/frequency levels.

From a core-level viewpoint, such controls render bursty workloads under continuous runtime management. Bursty workloads are typically characterized by frequent switching between high activity followed by no activity. The period of inactivity leads to idle energy consumption as the clock and supply voltage remain operational. Fig. 6.1 depicts the idle power consumption (i.e. energy per second) measurements on the Odroid-XU3 big.LITTLE platform for different core allocations and frequencies. The following two observations can be made from the figure. Firstly, with increasing number of inactive cores (big or little) the idle power consumption increases. As an example, the idle power of 4 big inactive cores at 2000 MHz is 1 Watt, while it is 0.8 Watt when only 1 big core is inactive. Secondly, the idle power is also dependant on the operating frequency. For instance, in the case where threads are allocated to LITTLE cores only, the idle power dissipation of 4 big inactive cores rises from 0.39 Watt at 1400 MHz to 1 Watt at 2000 MHz.

Idle power contributes to unuseful energy consumption, essentially reducing the battery operational life time. To reduce the idle power, a key traditional approach is to use power gating. The basic principle is to adopt a number of sleep transistors that disconnect the supply voltage rail for shutting down the inactive cores. Table 6.1 summarizes contributions of the existing power gating approaches. A hardware-

based stateless load balancing for homogeneous multi-core scheme is evaluated in terms of power and thermal behaviour in [112]. In this approach, a power reduction is achieved by switching off the idle cores. In [108], a sub-clock power gating technique is proposed to reduce static power during the sub-clock cycle of ARM Cortex-M_0. This technique uses intrusive redesigning of the power gating paradigm.

Among others, Charles et al. [26] implemented per core power gating in mainstream homogeneous processor (Intel Core i7). They showed that extra power headroom from power gating idle cores can be diverted to the active cores to increase their voltage and frequency without violating the power and thermal envelop. Similarly, diverting the saved power of idle cores into active cores was investigated in [105] by adopting a homogeneous many core AMD Opteron 6168 processor. The experimental results of this paper are based on manually tuned dynamic voltage scaling (DVS) coupled with power gating.

Minimizing idle power using the existing CMOS-based approaches (Table 6.1) still remains largely challenging. This is because with increasing capacitive loads, consisting of many cores, the gate dimensions of sleep transistors are becoming wider. As a result, for an effective idle energy minimization MEMS-based non-invasive technique is promising due to zero-leakage current. For example, a MEMS-based approach has been demonstrated in [47], highlighting simulation results that show potential energy reduction benefits over CMOS counterpart (for off-periods > 1 ms). Further, others illustrated methods of forming electromechanical power switch on top of integrated circuits (ICs) device for controlling idle energy consumption of CPU/GPU, I/O interface, and memory controller [111].

Despite its promises, the full-scale implementation for power gating remains unresolved due to engineering challenging. In this chapter, we propose a novel MEMS-based non-invasive idle energy controller for

Table 6.1: Features and limitations of the existing approaches.

Approach	Architecture	Validation	Working level	Key method
[26]	Homogeneous	Hardware	System	Power gating, (Nehalem)
[105]	Homogeneous	Hardware	System	Power gating, AMD Opt. 6168, DVS (manually)
[108]	Homogeneous	Hardware	Micro-architecture	Power gating, ARM Cortex-Mo
[112]	Homogeneous	Hardware	System	Task mapping, power gating
Proposed	Heterogeneous	Hardware+ simulation	System	MEMS-based power gating+ DVFS

bursty workloads executed on Odroid-XU3 heterogeneous platform. A MEMS-based power gating solution is being investigated for the following two reasons: firstly, MEMS-based controller itself contributes zero-leakage current unlike CMOS based power gating. Secondly, these controllers can be integrated by using back-end metallization layers with no penalty to the overall die area.

In our proposed approach, we make the following main *contributions*:

1. Propose a MEMS-based non-invasive power gating controller to support bursty workloads in heterogeneous many-core systems.
2. Core to the controller is an integrated sleep mode management based on the performance-energy states modelled using the feedback from performance counters.
3. Validate using a number of real application benchmarks to demonstrate the comparative advantages and trade-offs of our controller under bursty workloads scenario.

To the best of our knowledge, this is the first approach that investigates: (a) power gating control using non-invasive MEMS-based solution for heterogeneous many-core systems; (b) a systematic optimization of MEMS-based relay through parametric sweep in COMSOL tool. The rest of this chapter is organized as follows: the proposed system approach is described in Section 6.2. Section 6.3 presents the experimental results, while Section 6.4 concludes the chapter.

6.2 PROPOSED APPROACH

Using the optimized relay design (Chapter 3) a MEMS-based power gating controller is proposed. Fig. 6.2 shows our proposed MEMS-based power controller coupled with Exynos 5422, used as a case-study heterogeneous system. As can be seen, our proposed approach interacts with performance-energy state management to suitably identify opportunities for switching the idle big cores off under bursty workload scenarios. This is enacted through a charge pump connected to the power switch network (PSN) based-on MEMS. In the following sections we briefly describe our approach, highlighting the platform interactions.

6.2.1 *Hardware experimental platform*

In the previous chapter, a computer architecture in a similar fashion of biological system was proposed, such that a constantly active part has to be relatively slow and all the fast processing has to be built in specialised (periphery) units, whose operation is bursty. To that end,

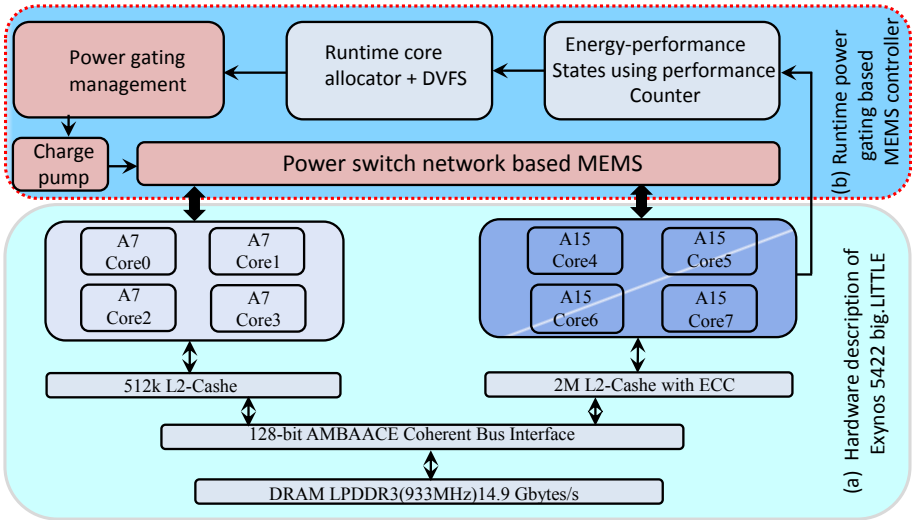


Figure 6.2: (a) Exynos 5422 block diagram; (b) proposed runtime power gating based MEMS.

the Odroid-XU3 board (Exynos 5422 big.LITTLE) supports techniques such as DVFS, core disabling, and affinity, typically used to optimize system operation in terms of energy consumption and performance has been chosen in the present study [155]. The Odroid-XU3 board is a small energy efficient octa-core computing device. The board can run Ubuntu 14.04 or Android 4.4 operating systems. The main component of Odroid-XU3 is the 28 nm Application Processor Exynos 5422. The architecture of the processor is shown in Fig. 6.2(a). This System-on-Chip is based on the ARM big.LITTLE heterogeneous architecture and consists of a high performance Cortex-A15 quad core processor block, a low power Cortex-A7 quad core block, a Mali-T628 GPU and 2GB DRAM LPDDR3. The board contains four real time current sensors that give the possibility of the measurement of power consumption on the four separate power domains: big (A15) CPUs, little (A7) CPUs, GPU and DRAM. In addition, there are also four temperature sensors for each of the A15 CPUs and one for the GPU. On the Odroid-XU3, for each power domain, the supply voltage (Vdd) and clock frequency can be tuned through a number of pre-set pairs of values. The performance-oriented Cortex-A15 block has a range of frequencies between 200 MHz and 2000 MHz with a 100 MHz step, whilst the low-power Cortex-A7 quad core block can scale its frequencies between 200 MHz and 1400 MHz with a 100 MHz step. On the other hand, The GPU can be clocked at seven various voltage-frequency set-ups ranged from 177 MHz to 600 MHz as shown in Table 6.2.

Two types of experiments were explored by using this real state-of-the-art mobile application platform, as shown in Fig.6.3, to demonstrate the functionality of dynamic frequency scaling. Fig. 6.4(a) shows the voltage-frequency characteristics by performing only the Ubuntu 14.04 OS on the Cortex-A7 and Cortex-A15. It is observed that in a

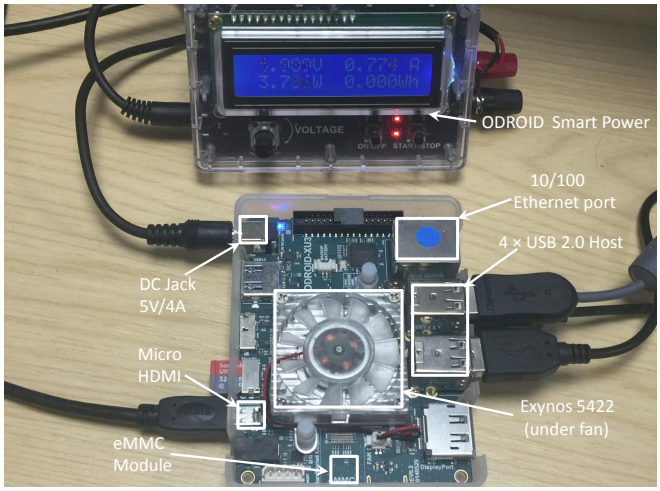


Figure 6.3: Experimental setup using Odroid-XU3 mobile platform with Samsung Exynos 5422 SoC.

certain frequency ranges the voltage remains constant, however, the voltage will increase linearly above these ranges. For an example, Cortex-A15 has a voltage of 0.912v at frequencies ranged from 200 MHz to 700 MHz, meanwhile Cortex-A7 operates at a voltage of 0.913v at frequencies ranged from 200 MHz to 500 MHz, as can be seen in Fig. 6.4(a). This experiment clearly illustrates the voltage-frequency dependencies in Odroid-XU3 platform. Linux governor renders the utility *cpufrec-set*, which is adopted to alter the frequency of all cores in the cluster of either A15 or A7. As an example, *cpufrec-set -u 1400MHz -c 6* sets maximum frequency 1400MHz for CPU core 6.

It should be noted that it is only possible to change the frequency of all cores in a cluster at the same time, all four cores of Cortex-A15 will execute the workload at frequency equal to 1400MHz. For instance, Fig. 6.4(b) depicts the voltage-frequency characteristics of the Odroid-XU3 platform by performing 50×10^7 square-root operations. Thread affinity was implemented to run the program on the pre-set CPU core. To assign the task to a particular CPU core the *taskset* Linux instruction was adopted. As an example, *taskse SqrtStress -c 2* executes the SqrtStress program on CPU core 3 (the last core of Cortex-A7 processor).

To enable the monitoring of energy-performance states, we designed a custom system software routine following ARM's technical specification manual that can report different performance counter values at pre-defined regular intervals. The routine can be used as a wrapper

Table 6.2: The available voltage/frequency scaling set-ups for GPU

Voltage (v)	0.76	0.97	0.85	0.87	0.91	0.96	0.97
Frequency (MHz)	177	266	350	420	480	543	600

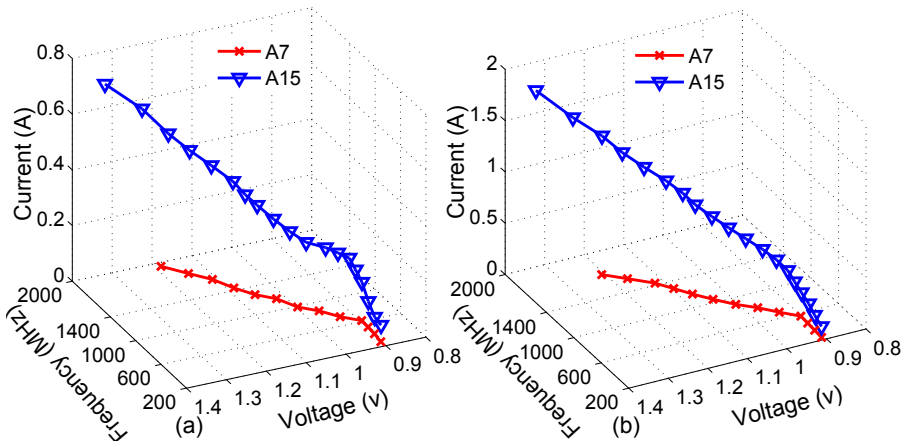


Figure 6.4: Voltage-frequency characteristics of Cortex-A7 and Cortex-A15 at: (a) without workload; (b) with exercising workload.

around the application binaries. This routine together with its libraries is currently being considered for a public release [3].

6.2.2 *Energy-performance state models*

Experiments were carried out on the Odroid-XU3 platform to investigate the energy consumption under different voltages and operation frequencies. The frequency of each cluster can be changed independently using utility programs and the system accordingly scales the operating voltage of the cluster to fit the chosen frequency. The eight cores in the Odroid-XU3 are numbered as follows: core 0, core 1, core 2 and core 3 are A7 cores, core 4, core 5, core 6 and core 7 are A15 cores.

Fig. 6.5 depicts the power consumption of the *ferret* application used in the study as an example for different thread to core allocations and operating frequencies. The power measurements were obtained through our performance counter routine (Section 6.2.1). As expected the power consumption increases as the operating frequency is increased, and as more cores are allocated for the given application. Fig. 6.6 shows the power consumption and execution time when the system is operating at the maximum frequency. The apparent power saturation is caused by the system engaging in automatic thermal throttling.

In this study, the power distribution between the cores for various PARSEC application scenarios is measured, as shown in Fig. 6.7. The key characteristics of such applications, according to [18], which are used in this work, can be seen in Table 6.3. Three applications (*fluidanimate*, *ferret*, and *bodytrack*) are chosen to represent memory-intensive, CPU-intensive, and CPU- with memory-intensive respectively. Such a classification decreases the effort of model description for combinations of concurrently exercising applications. The following two ob-

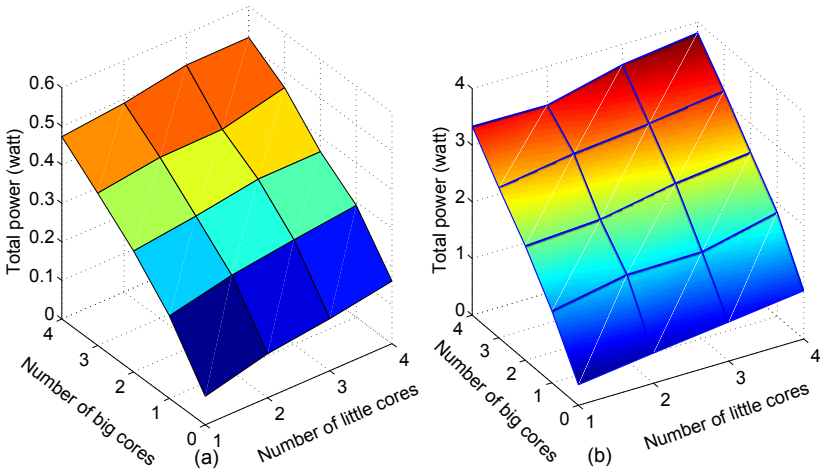


Figure 6.5: Total power for *ferret* application at: (a) 200 MHz; (b)1400 MHz.

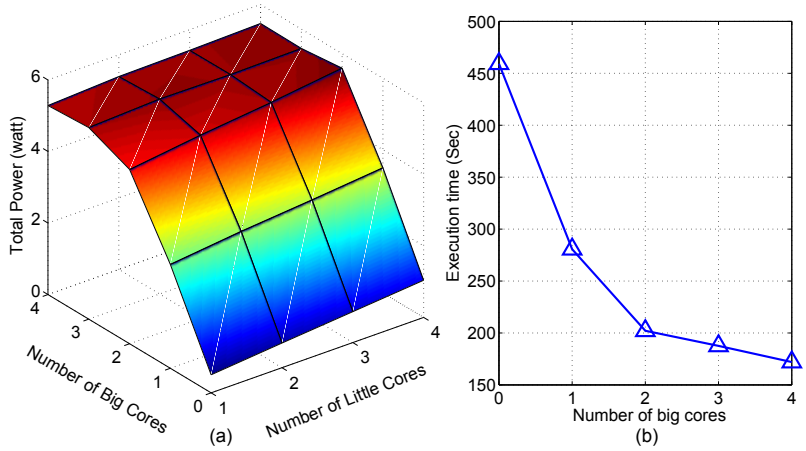


Figure 6.6: (a) Total power of *ferret* application at 2000 MHz big-cores and 1400 MHz little-cores; (b) execution time when 4 little cores fully operated with various big cores number.

servations can be made from the Fig. 6.7. Firstly, it is clearly indicated that the total power consumption for A7 and A15 for CPU-intensive application (*ferret*) is higher than memory intensive application (*fluidanimate*). Secondly, exercising concurrent applications such as *ferret* with *fluidanimate* consumes approximately similar power as running of *ferret* application alone.

The relationship between energy consumption Figs. 6.5 and 6.6, types of cores (big, little), frequency, and number of cores of the Odroid-XU3 platform can be rendered as [3]

$$E(V,f) = \frac{N_{A7} I_{A7} V_{A7}}{f_{A7}} + \frac{N_{A15} I_{A15} V_{A15}}{f_{A15}} + \varepsilon_1(x) \quad , \tag{6.1}$$

where N_{A7} and N_{A15} are the numbers of little and big cores, V_{A7} and V_{A15} are the voltages of A_7 and A_{15} cores, I_{A7} and I_{A15} are

Table 6.3: Qualitative description of the inherent key features of PARSEC benchmarks [18].

Program	Application Domain	Application Type	Parallelization Model Granularity		Work Set	Data Usage Sharing Exchange	
bodytrack	Computer Vision	CPU and memory intensive	Data-parallel	medium	medium	high	medium
ferret	Similarity Search	CPU intensive	pipeline	medium	unbounded	high	high
fluidanimate	Animation	memory intensive	Data-parallel	fine	large	low	medium

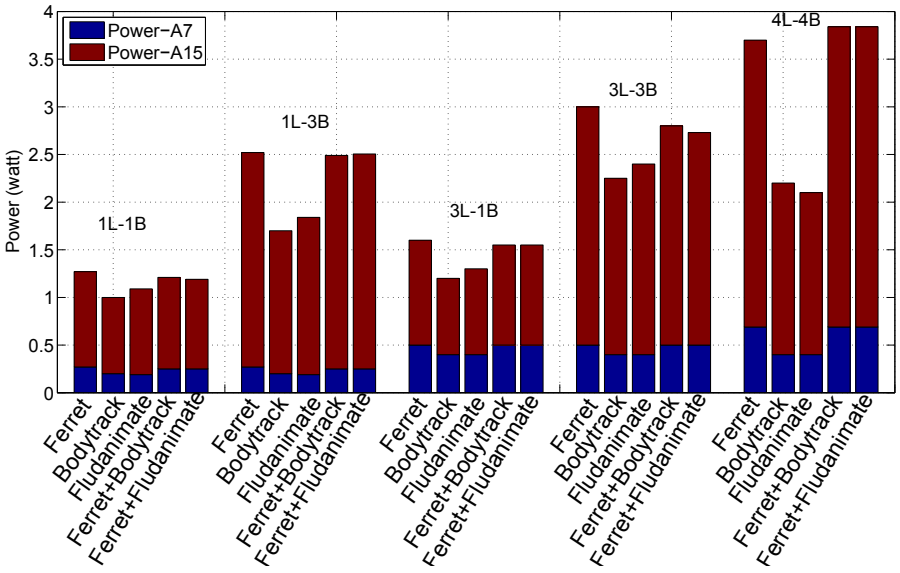


Figure 6.7: Total power for single and current applications in various set-ups exercising at 1400MHz .

the currents of A_7 and A_{15} cores, respectively, $\varepsilon_1(x)$ represents the background energy due to leakage, interconnects and memory access. Eq. 6.1 can be used to model energy consumption for all applications, with high accuracy up to 5% error rate. The detailed modelling results can be found in [3].

To enact power gating in Odroid-XU3, the PSN also adds energy consumption due to charging and discharging transition. If C_R and C_M are the total capacitances, which are charged or discharged during the transition of MEMS and CMOS respectively, then the switching energy of this transition can be evaluated as:

$$E_S(R) = U_R C_R V_{pi}^2 \simeq U_R \frac{\epsilon A}{g-z} V_{pi}^2 \simeq U_R \frac{\beta \times Lg^3}{W(g-z)} \quad , \tag{6.2}$$

$$E_S(M) = U_M W_M C_M V_g^2 \quad , \tag{6.3}$$

where U_R and U_M are the numbers of parallel MEMS and CMOS power switch, respectively. W_M represents the width of sleep transistor. For a given amount of time that cores A_{15} and or A_7 are in active or sleep mode, the energy per power gate switching cycle is:

$$E_R(V,f) = \frac{N_{A7} I_{A7} V_{A7}}{f_{A7}} + \frac{N_{A15} I_{A15} V_{A15}}{f_{A15}} + E_S(R) \quad , \tag{6.4}$$

$$E_M(V,f) = \frac{N_{A7} I_{A7} V_{A7}}{f_{A7}} + \frac{N_{A15} I_{A15} V_{A15}}{f_{A15}} + E_S(M) + \varepsilon_3 \quad , \quad (6.5)$$

where ε_3 represents energy consumption due to leakage current of sleep transistors. Eqs. (6.4) and (6.5) will be used to evaluate energy overhead caused by power gating.

6.2.3 *Measurements of Power Normalized performance (IPS/Watt)*

Using the measurements of power from the sensors and performance (IPS) from the performance counters, the normalized power-performance IPS/Watt merit was evaluated. These experiments demonstrate that the optimal system set-up is represented by the highest IPS/Watt value. The experimental data sets of the power normalized performance IPS/Watt for various applications case studies and architectural setups are shown in Fig. 6.8. It can be observed from the results that single applications including *bodytrack*, which is CPU and memory intensive, is displayed the highest power normalized performance IPS/Watt. As the number of cores allocation is increased, the power normalized performance IPS/Watt of the *bodytrack* indicates a significant increasing trends. As an example, when 4 Little cores and 4 big cores are allocated, a maximum IPS/Watt of 3.8 $\times 10^9$ is exhibited at operating frequency of 800MHz. As the operating frequency is increased the power dissipation will correspondingly increase, thereby lowering its IPS/Watt. It should be noted that similar illustrations can be made for *ferret* and *fluidanimate* single application scenarios. The concurrent application scenarios including *ferret+bodytrack* and *ferret+ fluidanimate* are left for the future work as indicated in Chapter 7.

6.2.4 *Power-gating management*

Based on the opportunities exposed by runtime control, regardless of which runtime approaches are used [3, 148, 172] our proposed techniques are able to find bursty opportunities and save leakage energy. Power gating of cores is enacted by the interface shown in Fig. 6.9. As can be seen, on every interval a number of flag registers are overwritten by the system software depending on the number of idle big cores. As an example, when two big cores (core 6 and core 7) are free the corresponding flag bits are set to 1 indicating the opportunity of power gating. These bits are then used to enable the charge pumps, which are used for shutting those cores.

In our work we mainly consider big cores as candidates for shutting down as: (1) they are typically used to offload computing units by the LITTLE cores running system software, and (2) the typical power saving achieved by shutting down the big cores is significantly higher

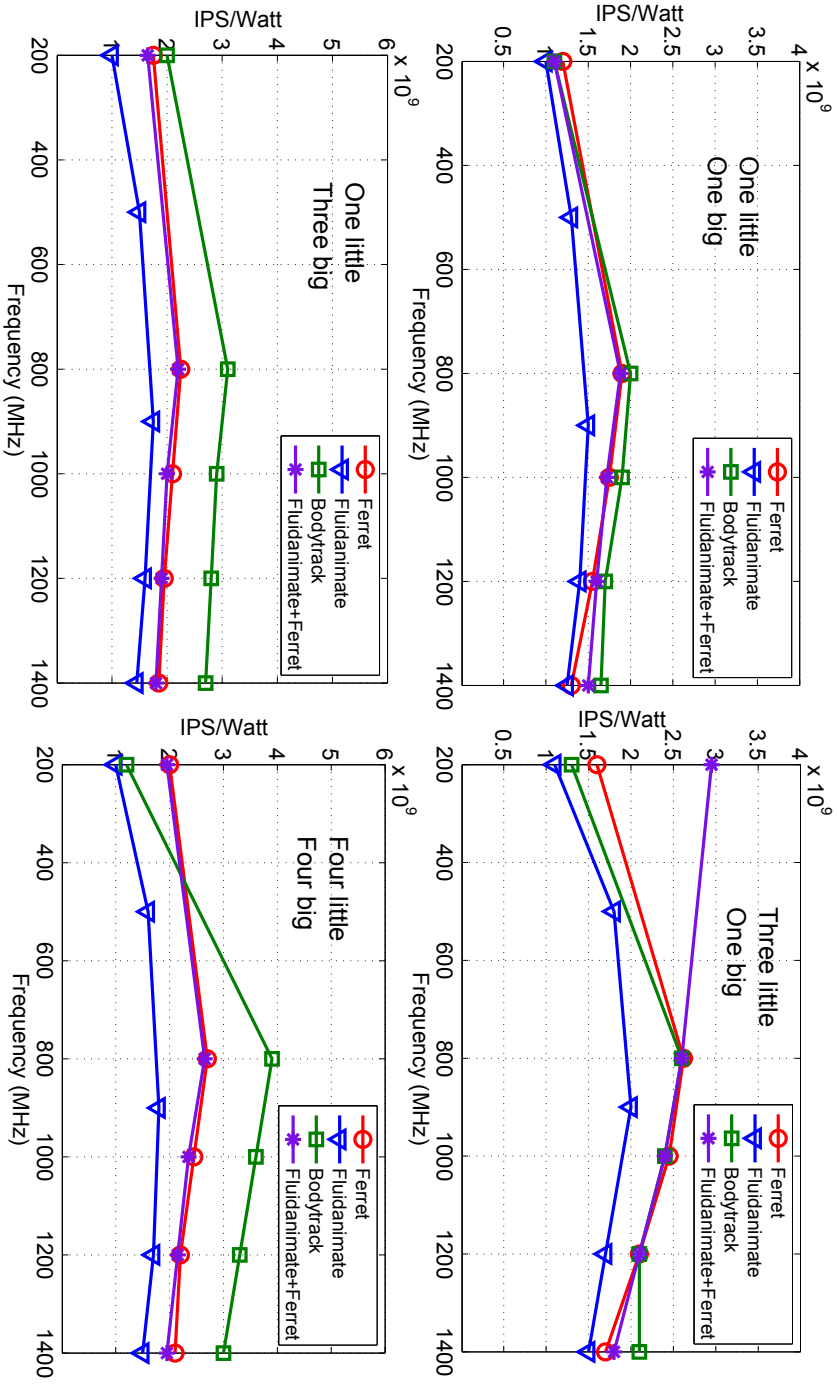


Figure 6.8: Total IPS/Watt for various applications exercised at different core allocations and frequencies.

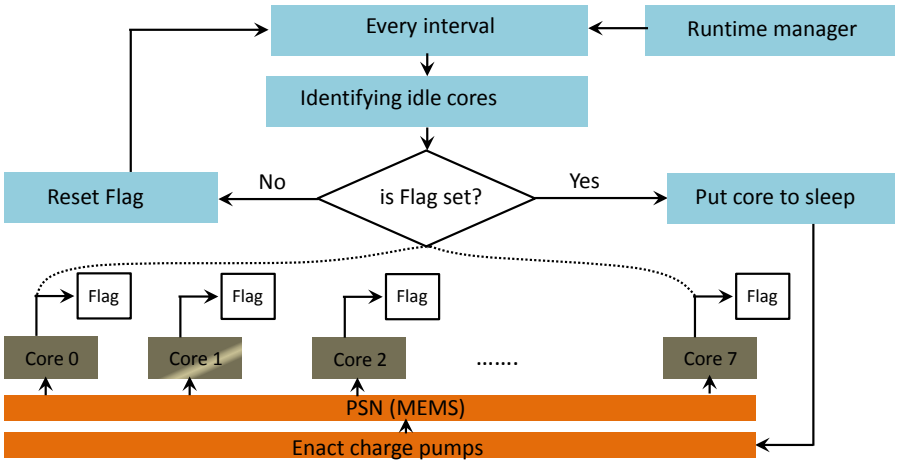


Figure 6.9: Hardware flowchart of the power gating management interface.

than that of LITTLE cores (Fig. 6.1). However, using our runtime interface circuitry the LITTLE cores (except core 0, which execute operating system routine) can also be shut-down.

6.3 RESULTS

A number of experiments are carried out in emulated environment in COMSOL tool, which are further cross-validated on the Exynos 5422 platform. The evaluation setup is first explained highlighting this environment, followed by extensive application case studies and trade-offs analysis.

6.3.1 *Evaluation setup*

6.3.1.1 *Energy measurement of power gating circuitry*

Fig. 6.10 shows the emulation environments used to evaluate both CMOS- and MEMS-based power gating circuitries. As can be seen, both setup consist of PSN coupled with the heterogeneous cores (only core A15 is shown for demonstration purpose). The CMOS-based emulation environment has been developed using Cadence Spice tool, while that of MEMS-based environment has been developed using COMSOL multiphysics tool. A key aspect for effective emulation is to determine the target impedance (Z_{target}) of the active cores. Establishing target impedance of the active core, that should be met over a broad of frequency band, can be computed by assuming a 5% allowable ripple in the core virtual voltage (VV_{dd}), and a 50% drawn current in the rise and fall time of the processor clock [11].

$$Z_{target} = \frac{0.1 \times VV_{dd}}{I_{peak}} \tag{6.6}$$

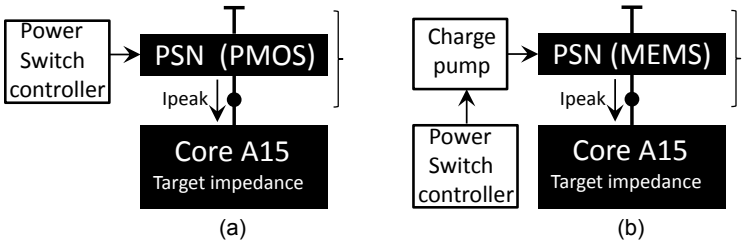


Figure 6.10: Power gating circuitry that meets target impedance: (a) PMOS transistors; (b) MEM relays.

In our experiment, the maximum current drawn by the A15 per core in the case of (for CPU-intensive application) is measured to be I_{peak}=1A at f=2GHz. For other operating frequencies and workload types the rated current ($I < I_{peak}$) can also be accurately estimated. Furthermore, for fair comparison between MEMS and CMOS based power gating circuitry, it is assumed that the allowable voltage drop (δ), as show in Fig. 6.10, is around 0.1volt. Therefore the number (U_M), width (W_M), and switching energy Es(M) of power transistors in the PSN are tuned so that it can deliver the maximum current with the allowable voltage drop (δ). Alternatively, PSN of MEM relays are evaluated as indicated in Table 3.2.

6.3.1.2 *Idle power measurement of Exynos 5422*

Using the setup (Section 6.2.1) a case study application (*ferret*, part of PARSEC benchmarks [18] as described in detail in Table 6.3) is executed in single cortex A15 (core7). The aim is to demonstrate in details the application state dependences over different times and frequencies. The application execution consists of two key states. State 1 characterizes core idle state at low frequency, while state 2 shows active state when the application is instantiated and exercised, as can be seen in Fig. 6.11.

6.3.2 *Application case studies*

Fig. 6.12 shows the state transition diagram of two different applications resulting from experimental user-space case studies of core allocation and DVFS. For demonstration purposes two different applications have been chosen: Fig. 6.12(a) for CPU-intensive application, and Fig. 6.12(b) for memory intensive application. The state transition in *ferret* starts from states 1 (idle) to s23 (active 2B-1L at 200MHz). These states are then followed by s21 (active 3B-1L at 400MHz) and s22 (active 4B-1L at 400MHz). The application returns back upon completion. The corresponding execution time of each state is normalised and annotated in percentage on the transition edge. Since the application remains in s23 most of the time, it gives the opportunity of shutting

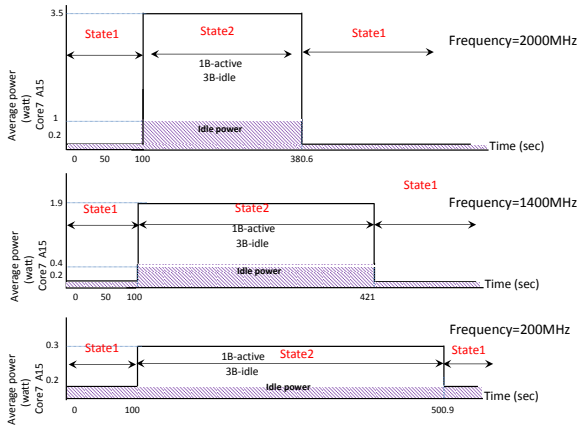


Figure 6.11: Idle power dissipation of Exynos 5422 big.LITTLE octa-core heterogeneous platform exercising *ferret* application in only one big core.

down the idle big cores, thereby saving higher energy, as can be seen in Fig. 6.12(a).

In the case of the memory-intensive application, it is expected that runtime controller will tend to allocate LITTLE cores at higher frequency. This gives opportunity for our MEMS-based to disable the big cores and achieving significant energy reduction. For example, since the application execute in s21 state most of the time it benefits from disabling three big cores and to achieve 32% energy reduction.

Fig. 6.13 shows the comparative energy consumptions of 3 different applications: *ferret*, *fluidanimate*, and *bodytrack*. These applications are executed with three different controllers. The first controller is an existing on-demand governor typically available in modern Linux operating system. The second controller is the user-space core allocator and DVFS without using any power gating approach. The third is our proposed controller featuring the MEMS-based power gating circuitry. From the Figure, two key observations can be made. Firstly, the on-demand governor, which is agnostic of core allocation management, only controls operating frequencies based on CPU usage. As such there is no power gating opportunity of the cores, resulting in high dynamic and leakage energy consumption. The user-space controller examines the IPS/Watt for various core allocation set and DVFS. However, due to no power gating, an effective energy minimization is limited. Our proposed approach integrates MEMS-based energy reduction and achieved upto 22% less energy consumption on top of 18% savings achieve by the user-space experiment when memory intensive application is exercised. The second observation is related to power gating opportunities exposed by different applications. As can be seen the best energy savings (40%) is achieved by memory-intensive applications. This is because these application favour allocation of LITTLE cores, and hence generate along idle periods for big cores.

Leakage energy reduction due to power gate idle cores throught state transition

Transition		Leakage energy reduction (%)
S21	S22	2
S22	S1	3
S23	S21	8

Leakage energy reduction due to power gate idle cores throught state transition

Transition		Leakage energy reduction (%)
S21	S22	32
S22	S1	10
S1	S21	16

Figure 6.12: Test bench of state transition based: (a) CPU-intensive *ferret* application; (b) Memory-intensive *fluidanimate* application.

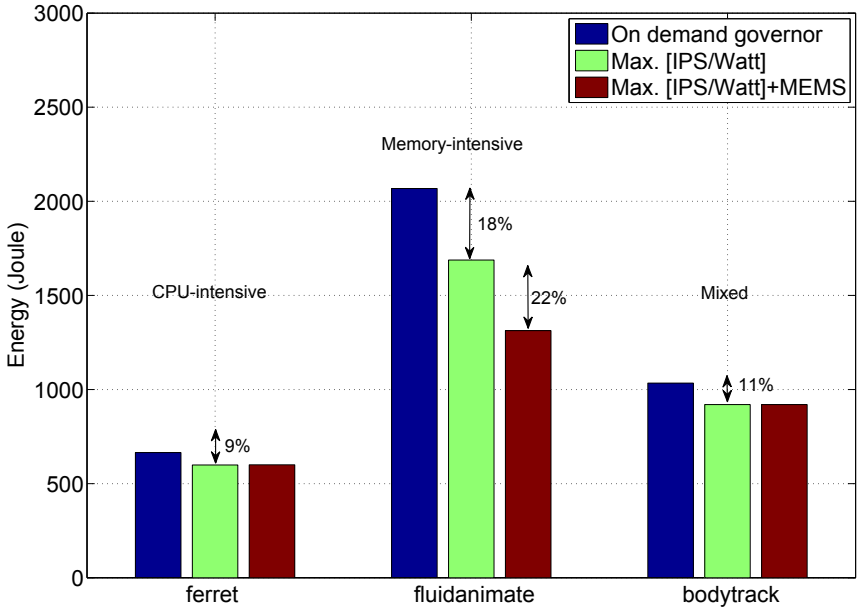


Figure 6.13: Comparative energy dissipations of *ferret*, *fluidanimate*, and *bodytrack* applications.

Fig. 6.14 shows a case study scenario when the applications (executing/observing) time is 10%. From the Figure, two points can be observed. Firstly, as expected the best energy savings (55.5%) is achieved by memory-intensive applications. Secondly, CPU- and mix-intensive applications start gaining significant energy savings compared to that of the previous scenario as the (executing/observing) time is reduced further.

Using Eqs. (6.4) and (6.5), the normalized energy overhead caused by PMOS power gating circuitry over that of MEMS has been evaluated as shown in Fig. 6.15. Two observations can be made. Firstly, increase the number of power gated cores causes a reduction in energy savings of MEMS due to the high switching energy of the adopted MEMS compared with that of CMOS. Secondly, increasing the core execution frequency leads to improve energy savings of MEMS power gating circuitry.

6.3.3 *Trade-off analysis*

ARM platforms are capable of managing power dissipation dynamically, where some cores can be deployed in various low-power modes of operation (ranging from power gating to simple wfi mode) using the OS power management (PM) policies. The CPU modes demonstrating the range of dynamic idle modes that a system can target at runtime. This can be determined via device tree bindings rendering the parameters which is demanded to enter/exit particular idle modes on a given platform [2]. In this work, the system can be placed in

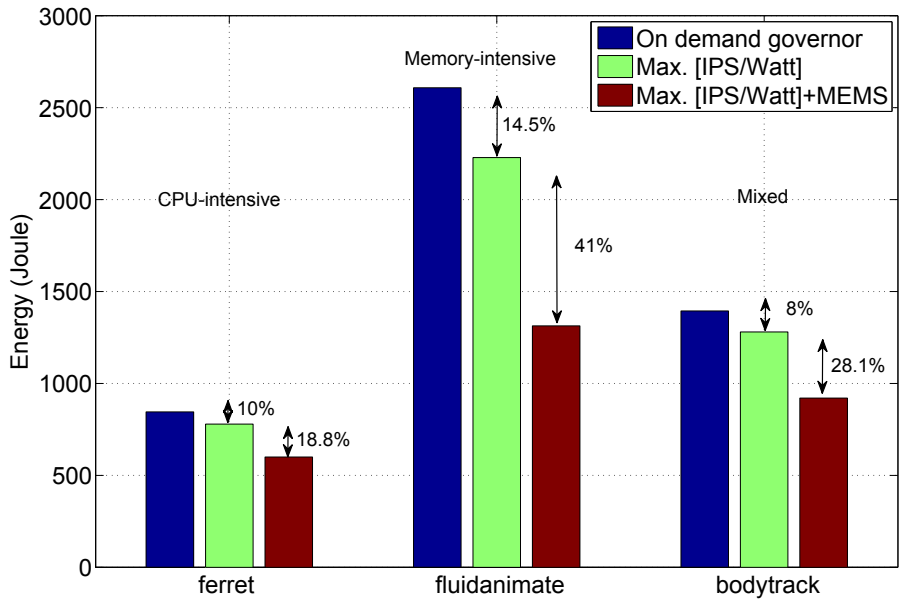


Figure 6.14: Comparative energy dissipations of *ferret*, *fluidanimate*, and *bodytrack* with applications (exercising/observing) time equal 0.1.

Table 6.4: Latency overhead

A15 (core/cluster) wake-up latency	A7 (core/cluster) wake-up latency	Charge pump latency/energy
(600/2230)us	(250/1650)us	0.5us/0.5pJ

off-state when exhibiting long idle period and/or bursty workloads. The wake up latency of sleep mode is experimentally measured to be 250µs for enabling a single Cortex-A7 core and 1659µs for waking up the entire cluster of A7. As expected, the wake up latency of the disabled core of Cortex-Aa5 is approximately 600µs, while enabling the entire cluster of A15 requires 2230µs for the cores to be stabilised, as shown in Table 6.4. Furthermore, the latency of using the charge pump (described in Section 4.4) is illustrated in Table 6.4.

The energy savings in our MEMS-based approach is achieved at the cost of latency overheads. However, the actual impact of these overheads will depend on the nature of the burst workloads. As an example, implementing our approach in a system that spend most of the time in the idle mode or a system exhibit regular burst workloads can achieve significant energy saving with a considerable latency overhead.

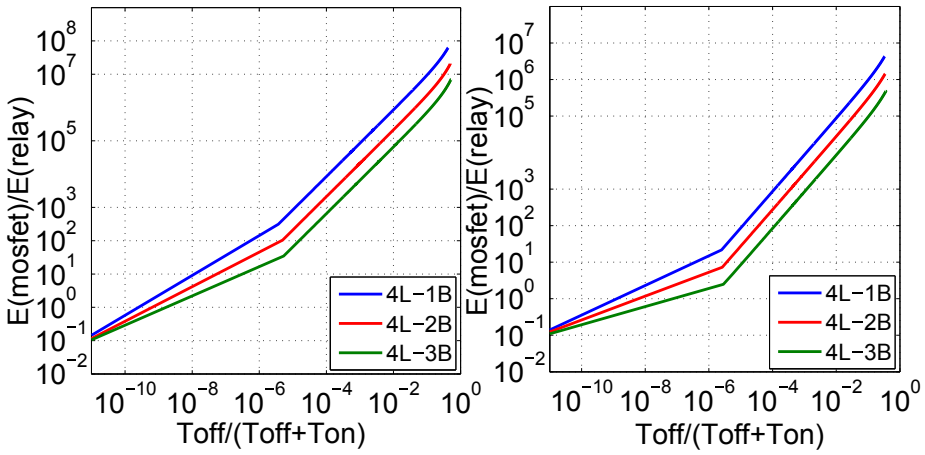


Figure 6.15: Energy gain vs (1-D) for various core allocations for a design power gated with MOSFETs and MEM relays while exercising *ferret* application at: (a) 2000MHz; (b) 200MHz.

6.4 CONCLUSIONS

A MEMS-based power gating scheme targeting heterogeneous many-core platforms running bursty applications is proposed. Theoretical analysis leads to concrete models which drove the MEMS design, with energy optimal results through the use of FEA techniques.

In our work, a novel controller based on workloads classification has been proposed. Experiments show that MEMS-based power gating approach can lead to significant improvements in the energy per computation metric, over that of no power gating using user-space controller and the Linux Ondemand governor. A number of PARSEC benchmark applications are used as case studies of bursty workloads, including CPU- and memory- intensive ones. Our findings demonstrate that MEMS-based power gating can achieve up to 55.5% greater energy savings compared to On-demand governor.

7 CONCLUSIONS AND FUTURE WORK

7.1 SUMMARY AND CONCLUSION

The shrinking of technology size is providing an opportunity for integrating higher numbers of cores on a single die, and hence silicon density is constantly increasing. This leads to ever-escalating of energy dissipation in VLSI designs, resulting in a substantial proportion of future chips is mandatory to be powered-off to comply with power budgets. In the present work, Micro/Nano-Electro-Mechanical MEM/NEM relays have been used as effective power gating switches to completely eliminate leakage energy in idle digital components. In order to adopt these emerging devices in power gating circuitry, the mechanical, electrical and logical characteristics of MEM/NEM relays have been investigated. One particular issue that this thesis addresses is the need for a scalable and accurate physical model of the MEM/NEM switches that can be plugged into the standard EDA software. As a result, an accurate Verilog-AMS model based on the published parameters of the suspended gate MEMS was developed. This model covers the self-actuation characteristics of the MEM/NEM switches but does not take into account the mechanical-thermal, contact-resistance variations, surface stress, and contact stiction impacts. The convergence problem caused by non-linear behaviour of MEMS, large disparity in variables scale of the electrical and mechanical domain, and contact discontinuity was mitigated by linearising the system behaviour.

To model and optimise different types of MEMS at various geometrical shape, dimensions, and material, 3D FEA simulation using the COMSOL multiphysics tool has been performed. This has led to building a switch model simulator based on the parameters obtained from the multiphysics analysis. Results revealed that MEMS at $45\mu m^2$ size can be actuated at drastically lower V_{pi}, T_{mech}, and switching energy compared to those of the existing technology making it of great interest for many future studies. Therefore, this section summarises the design exploration of the MEMS/NEMS parameters and presents its major conclusions.

Reducing idle energy consumption is becoming a critical concern with the rapid shrinking of technology size coupled with the increased demand for energy-constrained devices used for various implementations. In this thesis, power gating using MEM/NEM switches and/or sleep transistors for bursty and low duty-cycle implementations has been explored. The asynchronous power gating of 8- /32-tap FIR filter micropipeline has been investigated, and it is concluded that the

design architecture and data rate are the major factors in achieving greater energy savings when MEM/NEM switches are used. The proposed paradigm shows about 69% savings in energy consumption at a data rate of 1KHz compared to 39.5% in previous work. Furthermore, a zero delay ripple turn-on (ZDRTO) power gating control technique has been proposed to hide the mechanical delay time of the micropipeline stages power-gated by NEM switches. This approach postulates that the computational latency of the upstream stages, which are power gated by sleep transistors and exercised at low frequency, can be utilized to conceal the $T_{mech.}$ of the downstream stages, which are power-gated by NEMS and exercised at high frequency.

In this work, the system architecture of an energy-efficient battery operated gadget that operate in similar fashion of the biological systems has been proposed. It is postulated that such a system has two mode of operation: regular and bursty. It is revealed that the power gating of a system exhibiting bursty behaviour (in accordance with the need to respond to the demands of the environment) leads to non-trivial idle energy consumption, thereby providing an opportunity to shut it off. To that end, a MEMS-based power delivery network (PDN) that support non-invasive on/off activities coupled with bursty computation workloads has been proposed. Results have shown that the proposed PDN can achieve 1000× energy savings compared to that of power gating circuitry using sleep transistors.

This thesis also presents a novel runtime controller based on workload classification which are performed on the Exynos 5422 big.LITTLE octa-core heterogeneous platform. Experiments show that the MEMS-based runtime power gating approach leads to significant improvements in the energy per computation metric, compared to a case with no power gating using either runtime controller or the Linux Ondemand governor. Core to the proposed approach is integrated sleep mode management based on the performance-energy states and bursty workloads indicated by the performance counters. A number of real applications, PARSEC benchmark, are used as case studies of bursty workloads, including CPU-intensive, memory-intensive, and mixed applications. The Results revealed that up to 55.5% energy savings can be gained compared with the on-demand governor.

7.2 FUTURE WORK

Many research trends can be followed based on the architectures, techniques, and algorithms proposed in this work. Some recommendations for expanding the present work in this thesis are provided in this section.

The proposed switch model simulator can be expanded by taking into account the impact of mechanical contact on the on-state resistance (R_{on}) for a particular number of on/off switch cycles with/without

passing current through the drain-source terminal. A library could be built in the switch mode simulator that covers various kinds of MEM relays with their corresponding parameters variation at each scaling factor using 3D FEA. Furthermore, the creation of CMOS compatible devices could be investigated by exploring the trades-off involved in reliability (contact-stictions), on-state resistance, and current handling capability. Moreover, explore how the on-state resistance of MEM relays can be scaled with scaling the feature size.

The voltage supply design of MEMS/NEMS-gated CMOS circuits is also worth exploring, since it will potentially experience higher peak power and current spikes compared with circuits using sleep transistors. Furthermore, develop a tool that automatically cluster the asynchronous micropipeline stages so as to choose the best power gating approach for each cluster by specifying the idle power requirements and the maximum allowable wake-up time.

Idle power minimization of heterogeneous many-core platform using zero-leakage MEM relays can be further investigated by power gating of workloads which exhibit concurrent applications. Furthermore, a comprehensive analysis of idleness behaviour of modern GPU workloads of heterogeneous many-core platform, which can provide an opportunity of implementing power gating approach, is left for the future research. Moreover, worthy topics of future investigation include the prediction of the burst computation workloads occurrence as well as the duration of the idle events by using runtime machine learning algorithm including linear regression method. This necessitates the categorization of bursty computation workloads into periodic and aperiodic.

Backmatter

Part II

Thesis Appendices

A

VERILOG-AMS CODE FOR MEMS/NEMS

```
**************************************************************************
This is a scalable model of a Nano/Micro-Electro-Mechanical Relay. It
can be used for digital logic implementations.
**************************************************************************
'include "constants.vams"
'include "disciplines.vams"
module scalable_physical_model_relay(g, s, b, d, z);
inout g;
electrical g;
inout s;
electrical s;
inout b;
electrical b;
inout d;
electrical d;
inout z;
kinematic z;
**************************************************************************
Technology parameters of the program
**************************************************************************
% L_A: length of the spring
% W_A: width of the spring
% A_ov: overlap area between the gate-body electrode
% A_ch: overlap area between the gate-channel
% R_air: resistance of the air gap
% L: length of the gate
% W: width of the gate
% t: thickness of the gate oxide
parameter real L_A= 10,....,50e-6;
parameter real W_A= 5e-6 ;
parameter real A_ov=7.5e-10
parameter real A_ch=58e-12;
parameter real L= 27e-6;
parameter real W= 30e-6;
parameter real t= 1e-6;
parameter real g_0=180e-9 from (0:inf); % nominal gap height [m]
parameter real g_d=90e-9 from (0:inf); % nominal gap height minus
dimple thickness [m]
**************************************************************************
Definition of the internal program variables
**************************************************************************
```

```
// f: contact points between source/drain and channel
//c: channel voltage
// cs: channel at the source side
//cd: channel at the drain side
//mgr: mechanical ground, the position of the body (substrate)
//nv: normalized velocity of the shuttle
// nz: normalized position of the shuttle
electrical cd, c, f, cs;
kinematic mgr, nz, nv;
*****************************************************************************
Electrical parameters of MEMS
*****************************************************************************
% R_con: resistance of the channel touching contact
% R_pox: contact resistance to the drain/source electrode
% R_c: resistance through the channel
% R_air: resistance of the air chamber
% C_db: capacitance from the drain to the body
% C_sb: capacitance from the source to the body
% Q_f: quality factor
parameter real R_con= 1.0 from [0:inf);
parameter real R_pox= 500.0 from [0:inf);
parameter real R_c= 6000 from [0:inf);
parameter real R_air= 1e15 from [0:inf);
parameter real C_db= 5.0e-17 from (0:inf);
parameter real C_sb= 5.0e-17 from (0:inf);
parameter real Q_f= 1;
*****************************************************************************
Internal Variable declarations
*****************************************************************************
real P_g, P_EPS0, m_e, P_c, q, B_c;
real C_i, C_gc, k, m, Qcd, Qgf, Qcs, C_dsg0, Qgs, Qgd;
real b, Qgc, wo, Qgb, V_pi ;
analog begin
*****************************************************************************
Constants parameters
*****************************************************************************
% P_EPS0: permittivity of free space 8.85e-12 F/m
% m_e: mass of electron
% P_c: Planck's constant over 2pi
% q: charge on an electron
% B_c: Boltzmann's constant
% P_g: relative permittivity of gate oxide
% A: Hamaker constant (Joul)
P_EPS0= 8.85e-12;
m_e= 9.1e-31;
P_c= 1.055e-34;
```

q= 1.6e-19;
B_c= 1.38e-23;
P_g= 3.9;
A=1.0e-19;
B_c=3.87e-4;
****
Derived values
****
% C_i: initial capacitance of the gap
% k: structure's spring constant
% C_{gc}: gate to channel capacitance
% m: mass of the channel
% V_{pi}: pull-in voltage
% Q: overall quality of the beam
% b: viscous damping coefficient
C_i=P_EPS0×(Aov)/(gap);
k= B_c×L_A ×pow(g,3)/(2×z×pow((g-z), 2);
C_{gc}=P_EPS0×P_g×Ach/(t);
m=A_{gate} ×t×p;
b=pow(k×m/Q_f, 0.5);
wo=pow(k/m, 0.5);
V_{pi}=pow(B_c×L×pow(gap,3)/(P_EPS0×W×A), 0.5);
C_{dsc0}=P_EPS0×Acds/gap;
C_{dsg0}=P_EPS0×Adsg/gap;
F_e=P_EPS0× A_{ov} × pow(V(g,b),2)/2×1/pow(g_0- Pos(nz,mgr), 2);
Fs=k×Pos(nz,mgr);
F_{vdw}=A_{ov} × A/(6× pi×pow(g_0-Pos(nz,mgr),3))
****
Electrical circuit equations
****
% I(s,f): contact_channel_pox resistances from source side
I(s,f) <+ V(s,f)/(R_{con}+R_{pox}+R_c/2);
% I(g,b): gate-body variable capacitor
Qgb=V(g,b)×C_i×(1/(1-Pos(nz,mgr)));
I(g,b)<+ddt(Qgb);
% I(g,f): gate to drain/source capacitor
Qgf=V(g,f)×2×Cdsg0×(1/(1-Pos(nz,mgr)));
I(g,f)<+ddt(Qgf);
If (Pos(nz,mgr)<g_d)
I(d,s)<+ V(d,s)/R_{air};
else
% I(d,f): contact_channel_pox resistances from drain side
I(d,f)<+V(d,f)/(R_{con}+R_{pox}+R_c/2);
****
Mechanical state equations
****

```
Pos(mgr)<+0.0;
Pos(nv,mgr):ddt(Pos(nz,mgr))==Pos(nv,mgr);
Pos(nz,mgr):ddt(Pos(nv,mgr))==1/m*(-b*Pos(nv,mgr)-F_s+F_e+F_vdw);
end
endmodule
```

Part III

Thesis Bibliography

BIBLIOGRAPHY

[1] *Microsystem Design*. Springer (India) Pvt. Limited, 2006. ISBN 9788181285461. URL http://books.google.co.uk/books?id=aLXJQwAACAAJ.

[2] http://infocenter.arm.com/help/index.jsp.

[3] Ali Aalsaud, Rishad Shafik, Ashur Rafiev, Fie Xia, Sheng Yang, and Alex Yakovlev. Power–aware performance adaptation of concurrent applications in heterogeneous many-core systems. In *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*, pages 368–373. ACM, 2016.

[4] Amit Agarwal, Saibal Mukhopadhyay, Chris H Kim, Arijit Raychowdhury, and Kaushik Roy. Leakage power analysis and reduction for nano-scale circuits. *System-on-Chip: Next Generation Electronics*, 18:415, 2006.

[5] K. Akarvardar, D. Elata, and *et al.* Design considerations for complementary nanoelectromechanical logic gates. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 299–302, 2007. doi: 10.1109/IEDM.2007.4418930.

[6] K. Akarvardar, C. Eggimann, and *et al.* Analytical modeling of the suspended-gate fet and design insights for low-power logic. *Electron Devices, IEEE Transactions on*, 55(1):48–59, 2008. ISSN 0018-9383. doi: 10.1109/TED.2007.911070.

[7] H. Alrudainy, A. Mokhov, and A. Yakovlev. A scalable physical model for nano-electro-mechanical relays. In *Power and Timing Modeling, Optimization and Simulation (PATMOS), 2014 24th International Workshop on*, pages 1–7, Sept 2014. doi: 10.1109/PATMOS.2014.6951889.

[8] H. Alrudainy, Ali Alsaaud, Rishad Shafik, A. Mokhov, Fei Xia, and A. Yakovlev. Mems-based runtime idle energy minimization for bursty workloads in heterogeneous many-core systems. pages 93–100, 2016.

[9] H. Alrudainy, A. Mokhov, N. S. Dahir, and A. Yakovlev. Mems-based power delivery control for bursty applications. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 790–793, May 2016. doi: 10.1109/ISCAS.2016.7527359.

[10] Haider Alrudainy, Andrey Mokhov, and Alex Yakovlev. Mems-based power gating of asynchronous micropipelines for ultra

low energy driven computation. In *unpublished work, IEEE International Symposium on*, 2017.

[11] B. Amelifard and M. Pedram. Design of an efficient power delivery network in an soc to enable dynamic power management. In *Low Power Electronics and Design (ISLPED), 2007 ACM/IEEE International Symposium on*, pages 328–333, Aug 2007. doi: 10.1145/1283780.1283850.

[12] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[13] C. L. Ayala, D. Grogg, A. Bazigos, M. F. B. Badia, U. T. Duerig, M. Despont, and C. Hagleitner. A 6.7 mhz nanoelectromechanical ring oscillator using curved cantilever switches coated with amorphous carbon. In *2014 44th European Solid State Device Research Conference (ESSDERC)*, pages 66–69, Sept 2014. doi: 10.1109/ESSDERC.2014.6948759.

[14] Christopher L Ayala, Antonios Bazigos, Daniel Grogg, Yu Pu, and Christoph Hagleitner. Ultra-low-energy adiabatic dynamic logic circuits using nanoelectromechanical switches. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2596–2599. IEEE, 2015.

[15] Sung-Yong Bang, Kwanhu Bang, Sungroh Yoon, and Eui-Young Chung. Run-time adaptive workload estimation for dynamic voltage scaling. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(9):1334–1347, 2009.

[16] Antonios Bazigos, Christopher L Ayala, Montserrat Fernandez-Bolanos, Yu Pu, Daniel Grogg, Christoph Hagleitner, Sunil Rana, Tyson Tian Qin, Dinesh Pamunuwa, and Adrian M Ionescu. Analytical compact model in verilog-a for electrostatically actuated ohmic switches. *IEEE Transactions on Electron Devices*, 61(6): 2186–2194, 2014.

[17] Antonios Bazigos, Christopher L Ayala, Sunil Rana, Daniel Grogg, Montserrat Fernandez-Bolaños, Christoph Hagleitner, Tian Qin, Dinesh Pamunuwa, and Adrian M Ionescu. Electromechanical design space exploration for electrostatically actuated ohmic switches using extended parallel plate compact model. *Solid-State Electronics*, 99:93–100, 2014.

[18] Christian Bienia and Kai Li. Parsec 2.0: A new benchmark suite for chip-multiprocessors. In *Proceedings of the 5th Annual Workshop on Modeling, Benchmarking and Simulation*, volume 2011, 2009.

[19] Elena Blokhina, Sergi Gorreta, David Lopez, David Molinero, Orla Feely, Joan Pons-Nin, and Manuel Dominguez-Pumar. Dielectric charge control in electrostatic mems positioners/varactors. *Journal of microelectromechanical systems*, 21(3):559–573, 2012.

[20] Shekhar Borkar. Design challenges of technology scaling. *IEEE micro*, 19(4):23–29, 1999.

[21] Shekhar Borkar. Thousand core chips: a technology perspective. In *Proceedings of the 44th annual Design Automation Conference*, pages 746–749. ACM, 2007.

[22] Thomas D Burd and Robert W Brodersen. Design issues for dynamic voltage scaling. In *Proceedings of the 2000 international symposium on Low power electronics and design*, pages 9–14. ACM, 2000.

[23] Sangdon Byeon, Sanghak Shin, Jae-Sang Song, Son Ngoc Truong, Hyun-Sun Mo, Seongsoo Lee, and Kyeong-Sik Min. Process-variation-adaptive charge pump circuit using nem (nano-electromechanical) relays for low power consumption and high power efficiency. *JOURNAL OF SEMICONDUCTOR TECHNOLOGY AND SCIENCE*, 15(5):563–569, 2015.

[24] K. Van Caekenberghe. Modeling rf mems devices. *IEEE Microwave Magazine*, 13(1):83–110, Jan 2012. ISSN 1527-3342. doi: 10.1109/MMM.2011.2173984.

[25] Ji Cao, Ling Li, Kimihiko Kato, Tsu-Jae King Liu, and H-S Philip Wong. Sub-5 nm gap formation for low power nem switches. In *Energy Efficient Electronic Systems (E3S), 2015 Fourth Berkeley Symposium on*, pages 1–3. IEEE, 2015.

[26] James Charles et al. Evaluation of the intel core i7 turbo boost feature. In *Workload Characterization, 2009. IISWC 2009. IEEE International Symposium on*, pages 188–197. IEEE, 2009.

[27] C. Chen, R. Parsa, N. Patil, S. Chong, K. Akarvardar, J. Provine, D. Lewis, J. Watt, RT Howe, H.S.P. Wong, et al. Efficient fpgas using nanoelectromechanical relays. In *Proceedings of the 18th annual ACM/SIGDA international symposium on Field programmable gate arrays*, page 273–282. ACM, ACM, 2010.

[28] Chen Chen, W.S. Lee, R. Parsa, Soogine Chong, J. Provine, J. Watt, R.T. Howe, H.-S.P. Wong, and S. Mitra. Nano-electro-mechanical relays for fpga routing: Experimental demonstration and a design technique. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2012*, pages 1361–1366, March 2012. doi: 10.1109/DATE.2012.6176703.

[29] F. Chen, Hei Kam, and *et al.* 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.

[30] Fred Chen. *Energy-efficient wireless sensors: fewer bits, Moore MEMS*. PhD thesis, Massachusetts Institute of Technology, 2011.

[31] I-Ru Chen, Louis Hutin, Chanro Park, Rinus Lee, Rhesa Nathanael, Jack Yaung, Jaeseok Jeon, and Tsu-Jae King Liu. Scaled micro-relay structure with low strain gradient for reduced operating voltage. *ECS Transactions*, 45(6):101–106, 2012.

[32] Soogine Chong, Byoungil Lee, Kokab B Parizi, J Provine, Subhasish Mitra, Roger T Howe, and H-S Philip Wong. Integration of nanoelectromechanical (nem) relays with silicon cmos with functional cmos-nem circuit. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 30–5. IEEE, 2011.

[33] Faisal K Chowdhury, Daniel Saab, and Massood Tabib-Azar. Single-device xor and and gates for high speed, very low power lsi mechanical processors. *Sensors and Actuators A: Physical*, 188: 481–488, 2012.

[34] F.K. Chowdhury, D. Choe, T. Jevremovic, and M. Tabib-Azar. Design of mems based xor and and gates for rad-hard and very low power lsi mechanical processors. In *Sensors, 2011 IEEE*, pages 762–765, Oct 2011. doi: 10.1109/ICSENS.2011.6127242.

[35] L. T. Clark, M. Morrow, and W. Brown. Reverse-body bias and supply collapse for low effective standby power. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(9): 947–956, Sept 2004. ISSN 1063-8210. doi: 10.1109/TVLSI.2004.832930.

[36] Nicholas J Conway, Zachary J Traina, and Sang-Gook Kim. A strain amplifying piezoelectric mems actuator. *Journal of Micromechanics and Microengineering*, 17(4):781, 2007.

[37] Harold G Craighead. Nanoelectromechanical systems. *Science*, 290(5496):1532–1535, 2000.

[38] David A Czaplewski, Gary A Patrizi, Garth M Kraus, Joel R Wendt, Christopher D Nordquist, Steven L Wolfley, Michael S Baker, and Maarten P De Boer. A nanomechanical switch for integration with cmos logic. *Journal of Micromechanics and Microengineering*, 19(8):085003, 2009.

[39] BD Davidson, SM George, and VM Bright. Atomic layer deposition (ald) tungsten nano-electromechanical transistors. In *Micro*

Electro Mechanical Systems (MEMS), 2010 IEEE 23rd International Conference on, pages 424–427. IEEE, 2010.

[40] Alexis De Vos. *Reversible computing: fundamentals, quantum computing, and applications*. John Wiley & Sons, 2011.

[41] Robert H Dennard, Fritz H Gaensslen, V Leo Rideout, Ernest Bassous, and Andre R LeBlanc. Design of ion-implanted mosfet's with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9(5):256–268, 1974.

[42] David Duarte, Narayanan Vijaykrishnan, Mary Jane Irwin, H-S Kim, and G McFarland. Impact of scaling on the effectiveness of dynamic power reduction schemes. In *Computer Design: VLSI in Computers and Processors, 2002. Proceedings. 2002 IEEE International Conference on*, pages 382–387. IEEE, 2002.

[43] S. Dutta and V. Stojanovic. Floating-point unit design with nano-electro-mechanical (nem) relays. In *Nanoscale Architectures (NANOARCH), 2014 IEEE/ACM International Symposium on*, pages 145–150, July 2014. doi: 10.1109/NANOARCH.2014.6880487.

[44] Jason Edwards. Modeling and feedback control of a mems electrostatic actuator. Technical report, Cleveland State University, 2009.

[45] M. Enachescu, G. R. Voicu, and S. D. Cotofana. Is the road towards "zero-energy" paved with nemfet-based power management? In *2012 IEEE International Symposium on Circuits and Systems*, pages 2561–2564, May 2012. doi: 10.1109/ISCAS.2012.6271826.

[46] Hadi Esmaeilzadeh, Emily Blem, Renee St Amant, Karthikeyan Sankaralingam, and Doug Burger. Dark silicon and the end of multicore scaling. In *Computer Architecture (ISCA), 2011 38th Annual International Symposium on*, pages 365–376. IEEE, 2011.

[47] H. Fariborzi, M. Spencer, V. Karkare, Jaeseok Jeon, R. Nathanael, Chengcheng Wang, F. Chen, Hei Kam, V. Pott, Tsu-Jae King Liu, E. Alon, V. Stojanovic, and D. Markovic. Analysis and demonstration of mem-relay power gating. In *Custom Integrated Circuits Conference (CICC), 2010 IEEE*, pages 1–4, Sept 2010. doi: 10.1109/CICC.2010.5617380.

[48] H. Fariborzi, F. Chen, and *et al.* 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.

[49] H. Fariborzi, F. Chen, R. Nathanael, I. R. Chen, L. Hutin, R. Lee, T. J. K. Liu, and V. Stojanovic. Relays do not leak - cmos does. In *Design Automation Conference (DAC), 2013 50th ACM/EDAC/IEEE*, pages 1–4, May 2013.

[50] Jun Fujiki, Nuo Xu, Louis Hutin, I-Ru Chen, Chuang Qian, and Tsu-Jae King Liu. Microelectromechanical relay and logic circuit design for zero crowbar current. *IEEE Transactions on Electron Devices*, 61(9):3296–3302, 2014.

[51] Marius Gligor, Nicolas Fournel, and Frédéric Petrot. Adaptive dynamic voltage and frequency scaling algorithm for symmetric multiprocessor architecture. In *Digital System Design, Architectures, Methods and Tools, 2009. DSD'09. 12th Euromicro Conference on*, pages 613–616. IEEE, 2009.

[52] Ricardo Gonzalez and Mark Horowitz. Energy dissipation in general purpose microprocessors. *IEEE Journal of solid-state circuits*, 31(9):1277–1284, 1996.

[53] Michael K Gowan, Larry L Biro, and Daniel B Jackson. Power considerations in the design of the alpha 21264 microprocessor. In *Design Automation Conference, 1998. Proceedings*, pages 726–731. IEEE, 1998.

[54] D Grogg, U Drechsler, A Knoll, Y Pu, C Hagleitner, and M Despont. Curved cantilever design for a robust and scalable microelectromechanical switch. In *Proc. 56th Int. Conf. on Electron, Ion and Photon Beam Technology and Nanofabrication*, 2012.

[55] Structural Mechanics Module User Guide. Version 4.3. *COMSOL, May*, 2012.

[56] R.K. Gupta and S.D. Senturia. Pull-in time dynamics as a measure of absolute pressure. In *Micro Electro Mechanical Systems, 1997. MEMS '97, Proceedings, IEEE., Tenth Annual International Workshop on*, pages 290–294, 1997. doi: 10.1109/MEMSYS.1997.581830.

[57] Christoph H Haas and Michael Kraft. Modelling and analysis of a mems approach to dc voltage step-up conversion. *Journal of Micromechanics and Microengineering*, 14(9):S114, 2004.

[58] W. Haaswijk, L. Amaru, P. E. Gaillardon, and G. De Micheli. Nem relay design with biconditional binary decision diagrams. In *Proceedings of the 2015 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCHÂ¿Â¿15)*, pages 45–50, July 2015. doi: 10.1109/NANOARCH.2015.7180585.

[59] S. Hanson, B. Zhai, K. Bernstein, D. Blaauw, A. Bryant, L. Chang, K. K. Das, W. Haensch, E. J. Nowak, and D. M. Sylvester.

Ultralow-voltage, minimum-energy cmos. *IBM Journal of Research and Development*, 50(4.5):469–490, July 2006. ISSN 0018-8646. doi: 10.1147/rd.504.0469.

[60] Tina He, Rui Yang, Srihari Rajgopal, Swarup Bhunia, Mehran Mehregany, and Philip X-L Feng. Dual-gate silicon carbide (sic) lateral nanoelectromechanical switches. In *Nano/Micro Engineered and Molecular Systems (NEMS), 2013 8th IEEE International Conference on*, pages 554–557. IEEE, 2013.

[61] Weste Neil HE et al. *Cmos Vlsi Design: A Circuits And Systems Perspective, 3/E*. Pearson Education India, 2006.

[62] M.B. Henry and L. Nazhandali. From transistors to mems: Throughput-aware power gating in cmos circuits. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, pages 130–135, March 2010. doi: 10.1109/DATE.2010.5457224.

[63] M.B. Henry, R. Lyerly, L. Nazhandali, A. Fruehling, and D. Peroulis. Mems-based power gating for highly scalable periodic and event-driven processing. In *VLSI Design (VLSI Design), 2011 24th International Conference on*, pages 286–291, Jan 2011. doi: 10.1109/VLSID.2011.66.

[64] Michael B Henry. *Emerging power-gating techniques for low power digital circuits*. PhD thesis, Virginia Tech, 2011.

[65] S. Houri, G. Billiot, M. Belleville, A. Valentian, and H. Fanet. Limits of cmos technology and interest of nems relays for adiabatic logic applications. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 62(6):1546–1554, June 2015. ISSN 1549-8328. doi: 10.1109/TCSI.2015.2415177.

[66] Samer Houri, Christophe Poulain, Alexandre Valentian, and Hervé Fanet. Performance limits of nanoelectromechanical switches (nems)-based adiabatic logic circuits. *Journal of Low Power Electronics and Applications*, 3(4):368–384, 2013.

[67] Chenming Hu. *Modern semiconductor devices for integrated circuits*. Prentice Hall, 2010.

[68] Xiwei Huang, Hao Yu, and Wei Zhang. Nems based thermal management for 3d many-core system. In *Nanoscale Architectures (NANOARCH), 2011 IEEE/ACM International Symposium on*, pages 218–223, June 2011. doi: 10.1109/NANOARCH.2011.5941507.

[69] Xiwei Huang, Chun Zhang, Hao Yu, and Wei Zhang. A nanoelectromechanical-switch-based thermal management for 3-d integrated many-core memory-processor system. *IEEE Transactions on Nanotechnology*, 11(3):588–600, 2012.

[70] Safeen Huda, Muntasir Mallick, and Jason H Anderson. Clock gating architectures for fpga power reduction. In *2009 International Conference on Field Programmable Logic and Applications*, pages 112–118. IEEE, 2009.

[71] Hiro Iizuka and Takahito Ono. Low-voltage electrostatically driven nanoelectromechanical-switches. In *2015 Transducers-2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, pages 560–563. IEEE, 2015.

[72] Min-Woo Jang, Chia-Ling Chen, Walter E Partlo, Shruti R Patil, Dongjin Lee, Zhijiang Ye, David Lilja, T Andrew Taton, Tianhong Cui, and Stephen A Campbell. A three-terminal single-walled carbon nanotube thin film mems switch for digital logic applications. In *2011 16th International Solid-State Sensors, Actuators and Microsystems Conference*, pages 1705–1708. IEEE, 2011.

[73] Min-Woo Jang, Chia-Ling Chen, Walter E Partlo III, Shruti R Patil, Dongjin Lee, Zhijang Ye, David Lilja, T Andrew Taton, Tianhong Cui, and Stephen A Campbell. A pure single-walled carbon nanotube thin film based three-terminal microelectromechanical switch. *Applied Physics Letters*, 98(7):073502, 2011.

[74] Weon Wi Jang, Jeong Oen Lee, and Jun-Bo Yoon. A dram-like mechanical non-volatile memory. In *TRANSDUCERS 2007-2007 International Solid-State Sensors, Actuators and Microsystems Conference*, pages 2187–2190. IEEE, 2007.

[75] Weon Wi Jang, Jeong Oen Lee, Hyun-Ho Yang, and Jun-Bo Yoon. Mechanically operated random access memory (moram) based on an electrostatic microswitch for nonvolatile memory applications. *IEEE Transactions on Electron Devices*, 55(10):2785–2789, 2008.

[76] Weon Wi Jang, Jeong Oen Lee, Jun-Bo Yoon, Min-Sang Kim, Ji-Myoung Lee, Sung-Min Kim, Keun-Hwi Cho, Dong-Won Kim, Donggun Park, and Won-Seong Lee. Fabrication and characterization of a nanoelectromechanical switch with 15-nm-thick suspension air gap. *Applied Physics Letters*, 92(10):103110, 2008.

[77] Weon Wi Jang, Jun-Bo Yoon, Min-Sang Kim, Ji-Myoung Lee, Sung-Min Kim, Eun-Jung Yoon, Keun Hwi Cho, Sung-Young Lee, In-Hyuk Choi, Dong-Won Kim, et al. Nems switch with 30nm-thick beam and 20nm-thick air-gap for high density non-volatile memory applications. *Solid-State Electronics*, 52(10):1578–1583, 2008.

[78] J. Jeon, V. Pott, H. Kam, R. Nathanael, E. Alon, and T. J. K. Liu. Perfectly complementary relay design for digital logic

applications. *IEEE Electron Device Letters*, 31(4):371–373, April 2010. ISSN 0741-3106. doi: 10.1109/LED.2009.2039916.

[79] J. Jeon, L. Hutin, R. Jevtic, N. Liu, Y. Chen, R. Nathanael, W. Kwon, M. Spencer, E. Alon, B. Nikolic, and T. J. K. Liu. Multiple-input relay design for more compact implementation of digital logic circuits. *IEEE Electron Device Letters*, 33(2):281–283, Feb 2012. ISSN 0741-3106. doi: 10.1109/LED.2011.2177436.

[80] Jaeseok Jeon. *Advanced Relay Design and Technology for Energy-Efficient Electronics*. PhD thesis, EECS Department, University of California, Berkeley, Jul 2011. URL http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-81.html.

[81] Hei Kam, Vincent Pott, Rhesa Nathanael, Jaeseok Jeon, Elad Alon, and Tsu-Jae King Liu. Design and reliability of a micro-relay technology for zero-standby-power digital logic applications. In *2009 IEEE International Electron Devices Meeting (IEDM)*, pages 1–4. IEEE, 2009.

[82] Hei Kam, Tsu-Jae King Liu, and *et al.* 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.

[83] M Kamon, S Maity, D DeReus, Z Zhang, S Cunningham, S Kim, J McKillop, A Morris, G Lorenz, and L Daniel. New simulation and experimental methodology for analyzing pull-in and release in mems switches. In *2013 Transducers & Eurosensors XXVII: The 17th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS & EUROSENSORS XXVII)*, pages 2373–2376. IEEE, 2013.

[84] James Kao, Anantha Chandrakasan, and Dimitri Antoniadis. Transistor sizing issues and tool for multi-threshold cmos technology. In *Proceedings of the 34th annual Design Automation Conference*, pages 409–414. ACM, 1997.

[85] Kimihiko Kato, Vladimir Stojanovic, and Tsu-Jae King Liu. Non-volatile nano-electro-mechanical memory for energy-efficient data searching. *IEEE Electron Device Letters*, 37(1):31–34, 2016.

[86] Takao Kawano, Naoya Onizawa, Atsushi Matsumoto, and Takahiro Hanyu. Adjacent-state monitoring based fine-grained power-gating scheme for a low-power asynchronous pipelined system. In *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 2067–2070. IEEE, 2011.

[87] Michael Keating, David Flynn, Rob Aitken, Alan Gibbons, and Kaijian Shi. *Low power methodology manual: for system-on-chip design*. Springer Publishing Company, Incorporated, 2007.

[88] Min-Wu Kim, Yong-Ha Song, Seung-Deok Ko, Sang-Joon Ahn, and Jun-Bo Yoon. Ultra-low voltage mems switch using a folded hinge structure. *Micro and Nano Systems Letters*, 2(1):1–5, 2014.

[89] Nam Sung Kim, Todd Austin, David Baauw, Trevor Mudge, Krisztián Flautner, Jie S Hu, Mary Jane Irwin, Mahmut Kandemir, and Vijaykrishnan Narayanan. Leakage current: Moore's law meets static power. *computer*, 36(12):68–75, 2003.

[90] Nam Sung Kim, Krisztian Flautner, David Blaauw, and Trevor Mudge. Circuit and microarchitectural techniques for reducing cache leakage power. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(2):167–184, 2004.

[91] Sung Hwan Kim, Hei Kam, and *et al.* Germanium-source tunnel field effect transistors with record high ion/ioff. In *VLSI Technology, 2009 Symposium on*, pages 178–179, 2009.

[92] Ezekiel JJ Kruglick and Kristofer SJ Pister. Lateral mems microcontact considerations. *Journal of Microelectromechanical Systems*, 8(3):264–271, 1999.

[93] Ken Kundert and Olaf Zinke. *The designers guide to Verilog-AMS*. Springer Science & Business Media, 2006.

[94] Lawrence Lagebrake, Shinzo Onishi, Scott Samson, and Raj Popuri. Mems dc to dc switching converter, January 3 2012. US Patent 8,088,638.

[95] D Lee, WS Lee, J Provine, J-O Lee, J-B Yoon, RT Howe, S Mitra, and H-SP Wong. Titanium nitride sidewall stringer process for lateral nanoelectromechanical relays. In *Micro Electro Mechanical Systems (MEMS), 2010 IEEE 23rd International Conference on*, pages 456–459. IEEE, 2010.

[96] D. Lee, W. S. Lee, C. Chen, F. Fallah, J. Provine, S. Chong, J. Watkins, R. T. Howe, H. S. P. Wong, and S. Mitra. Combinational logic design using six-terminal nem relays. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(5):653–666, May 2013. ISSN 0278-0070. doi: 10.1109/TCAD.2012.2232707.

[97] Jeong-Oen Lee, Min-Wu Kim, Seung-Deok Ko, Hee-Oh Kang, Woo-Ho Bae, Min-Ho Kang, Ki-Nam Kim, Dong-Eun Yoo, and Jun-Bo Yoon. 3-terminal nanoelectromechanical switching device in insulating liquid media for low voltage operation and reliability improvement. In *Proc. Intl. Electron Dev. Meeting*, pages 1–4, 2009.

[98] Jeong Oen Lee, Yong-Ha Song, Min-Wu Kim, Min-Ho Kang, Jae-Sub Oh, Hyun-Ho Yang, and Jun-Bo Yoon. A sub-1-volt

nanoelectromechanical switching device. *Nature nanotechnology*, 8(1):36–40, 2013.

[99] Jin Woo Lee et al. Squeeze-film damping of flexible microcantilevers at low ambient pressures: theory and experiment. *Journal of Micromechanics and Microengineering*, 19(10):105029, 2009. URL http://stacks.iop.org/0960-1317/19/i=10/a=105029.

[100] Sang Wook Lee, Dong Su Lee, Raluca E Morjan, Sung Ho Jhang, Martin Sveningsson, OA Nerushev, Yung Woo Park, and Eleanor EB Campbell. A three-terminal carbon nanorelay. *Nano letters*, 4(10):2027–2030, 2004.

[101] Te-Hao Lee et al. Electromechanical computing at 500 c with silicon carbide. *Science*, 329(5997):1316–1318, 2010.

[102] Tong Lin, Kwen-Siong Chong, Bah-Hwee Gwee, and Joseph S Chang. Fine-grained power gating for leakage and short-circuit power reduction by using asynchronous-logic. In *2009 IEEE International Symposium on Circuits and Systems*, pages 3162–3165. IEEE, 2009.

[103] Jian Lu, Yi Zhang, T. Itoh, and R. Maeda. Design, fabrication, and integration of piezoelectric mems devices for applications in wireless sensor network. In *Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP), 2011 Symposium on*, pages 217–221, May 2011.

[104] K. Ma, N. Chandramoorthy, X. Li, S. K. Gupta, J. Sampson, Y. Xie, and V. Narayanan. Using multiple-input nems for parallel a/d conversion and image processing. In *2015 IEEE Computer Society Annual Symposium on VLSI*, pages 339–344, July 2015. doi: 10.1109/ISVLSI.2015.114.

[105] Kai Ma and Xiaorui Wang. Pgcapping: exploiting power gating for power capping and core lifetime balancing in cmps. In *Proceedings of the 21st international conference on Parallel architectures and compilation techniques*, pages 13–22. ACM, 2012.

[106] S.K. Manohar, R. Venkatasubramanian, and P.T. Balsara. Hybrid nems-cmos dc-dc converter for improved area and power efficiency. In *VLSI Design (VLSID), 2012 25th International Conference on*, pages 221–226, 2012. doi: 10.1109/VLSID.2012.74.

[107] Sujan K Manohar, Ramakrishnan Venkatasubramanian, and Poras T Balsara. Heterogeneous nems-cmos dcm buck regulator for improved area and enhanced power efficiency. *IEEE Transactions on Nanotechnology*, 14(1):140–151, 2015.

[108] J. N. Mistry et al. Sub-clock power-gating technique for minimising leakage power during active mode. In *2011 DATE*, pages 1–6, March 2011. doi: 10.1109/DATE.2011.5763026.

[109] Andrey Mokhov, Danil Sokolov, and Alex Yakovlev. Adapting asynchronous circuits to operating conditions by logic parametrisation. In *2012 IEEE 18th International Symposium on Asynchronous Circuits and Systems*, pages 17–24. IEEE, 2012.

[110] Andrey Mokhov, Maxim Rykunov, Danil Sokolov, and Alex Yakovlev. Design of processors with reconfigurable microarchitecture. *Journal of Low Power Electronics and Applications*, 4(1): 26–43, 2014.

[111] K. Mori, Z. Tran, G.T. Dao, and M.E. Ramon. Method of forming an electromechanical power switch for controlling power to integrated circuit devices and related devices, July 22 2014. URL https://www.google.co.uk/patents/US8786130. US Patent 8,786,130.

[112] E. Musoll. Hardware-based load balancing for massive multicore architectures implementing power gating. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 29(3): 493–497, March 2010. ISSN 0278-0070. doi: 10.1109/TCAD.2009.2018863.

[113] Shin'ichiro Mutoh, Takakuni Douseki, Yasuyuki Matsuya, Takahiro Aoki, Satoshi Shigematsu, and Junzo Yamada. 1-v power supply high-speed digital circuit technology with multithreshold-voltage cmos. *IEEE Journal of Solid-state circuits*, 30(8):847–854, 1995.

[114] N. M. Nair and R. M. Kuruvithadam. Nem relay based multinary ram a memory with optimized power, area and speed. In *2015 International Conference on Control Communication Computing India (ICCC)*, pages 563–568, Nov 2015. doi: 10.1109/ICCC.2015.7432961.

[115] R. Nathanael, V. Pott, and *et al.* 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.

[116] Rhesa Nathanael. *Nano-Electro-Mechanical (NEM) Relay Devices and Technology for Ultra-Low Energy Digital Integrated Circuits*. PhD thesis, EECS Department, University of California, Berkeley, May 2013. URL http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-45.html.

[117] L. Nazhandali, M. Minuth, and T. Austin. Sensebench: toward an accurate evaluation of sensor network processors. In *IEEE International. 2005 Proceedings of the IEEE Workload Characterization Symposium, 2005.*, pages 197–203, Oct 2005. doi: 10.1109/IISWC.2005.1526017.

[118] D Niarchos. Magnetic mems: key issues and some applications. *Sensors and Actuators A: Physical*, 109(1):166–173, 2003.

[119] N.S. Nise. *Control Systems Engineering*. John Wiley & Sons, Incorporated, 2011. ISBN 9780470646120. URL http://books.google.co.uk/books?id=7sVfbwAACAAJ.

[120] K. J. Nowka, G. D. Carpenter, E. W. MacDonald, H. C. Ngo, B. C. Brock, K. I. Ishii, T. Y. Nguyen, and J. L. Burns. A 32-bit powerpc system-on-a-chip with support for dynamic voltage scaling and dynamic frequency scaling. *IEEE Journal of Solid-State Circuits*, 37(11):1441–1447, Nov 2002. ISSN 0018-9200. doi: 10.1109/JSSC.2002.803941.

[121] Austin Ogweno, Alex Yakovlev, and Patrick Degenaar. Power gating in asynchronous micropiplines for low power data driven computing. In *Ph. D. Research in Microelectronics and Electronics (PRIME), 2015 11th Conference on*, pages 342–345. IEEE, 2015.

[122] Carlos Ortega, Jonathan Tse, and Rajit Manohar. Static power reduction techniques for asynchronous circuits. In *Asynchronous Circuits and Systems (ASYNC), 2010 IEEE Symposium on*, pages 52–61. IEEE, 2010.

[123] G. Palumbo and D. Pappalardo. Charge pump circuits: An overview on design strategies and topologies. *Circuits and Systems Magazine, IEEE*, 10(1):31–45, First 2010. ISSN 1531-636X. doi: 10.1109/MCAS.2009.935695.

[124] G Palumbo, D Pappalardo, and M Gaibotti. Charge pump with adaptive stages for non-volatile memories. *IEE proceedings. Circuits, devices and systems*, 153(2):136–142, 2006.

[125] R Parsa, K Akarvardar, J Provine, D Lee, D Elata, S Mitra, H Wong, and R Howe. Composite polysilicon-platinum lateral nanoelectromechanical relays. In *Proceedings of Hilton Head Workshop: A Solid-State Sensors, Actuators and Microsystems Workshop*, pages 7–10, 2010.

[126] Alexis Peschot, Chuang Qian, and Tsu-Jae King Liu. Nano-electromechanical switches for low-power digital computing. *Micromachines*, 6(8):1046–1065, 2015.

[127] Khem C Pokhrel. Physical and silicon measures of low power clock gating success: An apple to apple case study. *Synopsys Users Group (SNUG)*, 2007.

[128] V. Pott, Hei Kam, and *et al.* 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.

[129] R. M. Proie, R. G. Polcawich, J. S. Pulskamp, T. Ivanov, and M. E. Zaghloul. Development of a pzt mems switch architecture for low-power digital applications. *Journal of Microelectromechanical Systems*, 20(4):1032–1042, Aug 2011. ISSN 1057-7157. doi: 10.1109/JMEMS.2011.2148160.

[130] Mateja Putic, Liang Di, Benton H Calhoun, and John Lach. Panoptic dvs: A fine-grained dynamic voltage scaling framework for energy scalable cmos design. In *Computer Design, 2009. ICCD 2009. IEEE International Conference on*, pages 491–497. IEEE, 2009.

[131] Tian Qin, Sunil Rana, and Dinesh Pamunuwa. Design methodologies, models and tools for very-large-scale integration of nem relay-based circuits. In *Computer-Aided Design (ICCAD), 2015 IEEE/ACM International Conference on*, pages 641–648. IEEE, 2015.

[132] Furong Qu, Mengyan Nie, Yu Liu, and Yi Ou. Analysis and design of a mems dc/dc converter. *IEEJ Transactions on Electrical and Electronic Engineering*, 10(3):340–344, 2015.

[133] J. Qui, Jeffrey H. Lang, AH. Slocum, and Ralf Strumpler. A high-current electrothermal bistable mems relay. In *Micro Electro Mechanical Systems, 2003. MEMS-03 Kyoto. IEEE The Sixteenth Annual International Conference on*, pages 64–67, Jan 2003. doi: 10.1109/MEMSYS.2003.1189688.

[134] Jan Rabaey. *Low power design essentials*. Springer Science & Business Media, 2009.

[135] Jan M. Rabaey, Anantha Chandrakasan, and Borivoje Nikolic. *Digital integrated circuits- A design perspective*. Prentice Hall, 2ed edition, 2004.

[136] S. Rana, T. Qin, D. Grogg, M. Despont, Y. Pu, C. Hagleitner, and D. Pamunuwa. Modelling nem relays for digital circuit applications. In *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, pages 805–808, May 2013. doi: 10.1109/ISCAS.2013.6571969.

[137] S. Rana, Qin Tian, A. Bazigos, D. Grogg, M. Despont, C.L. Ayala, C. Hagleitner, A.M. Ionescu, R. Canegallo, and D. Pamunuwa.

Energy and latency optimization in nem relay-based digital circuits. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 61(8):2348–2359, Aug 2014. ISSN 1549-8328. doi: 10.1109/TCSI.2014.2309752.

[138] A. Raychowdhury, Jeong Il Kim, D. Peroulis, and K. Roy. Integrated mems switches for leakage control of battery operated systems. In *Custom Integrated Circuits Conference, 2006. CICC '06. IEEE*, pages 457–460, Sept 2006. doi: 10.1109/CICC.2006.320821.

[139] M. Riverola, G. Vidal-Alvarez, F. Torres, and N. Barniol. 3-terminal tungsten cmos-nem relay. In *2014 10th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, pages 1–4, June 2014. doi: 10.1109/PRIME.2014.6872734.

[140] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits. *Proceedings of the IEEE*, 91(2):305–327, Feb 2003. ISSN 0018-9219. doi: 10.1109/JPROC.2002.808156.

[141] Kaushik Roy, Hamid Mahmoodi-Meimand, and Saibal Mukhopadhyay. Leakage control for deep-submicron circuits. In *Microtechnologies for the New Millennium 2003*, pages 135–146. International Society for Optics and Photonics, 2003.

[142] M Ruan, J Shen, and CB Wheeler. Latching microelectromagnetic relays. *Sensors and Actuators A: Physical*, 91(3):346–350, 2001.

[143] Stefan Rusu, Simon Tam, Harry Muljono, David Ayers, Jonathan Chang, Brian Cherkauer, Jason Stinson, John Benoit, Raj Varada, Justin Leung, et al. A 65-nm dual-core multithreaded xeon® processor with 16-mb l3 cache. *IEEE Journal of Solid-State Circuits*, 42(1):17–25, 2007.

[144] J.I Seeger and S.B. Crary. Stabilization of electrostatically actuated mechanical devices. In *Solid State Sensors and Actuators, 1997. TRANSDUCERS '97 Chicago., 1997 International Conference on*, volume 2, pages 1133–1136 vol.2, Jun 1997. doi: 10.1109/SENSOR.1997.635402.

[145] Dipanjan Sengupta and Resve A Saleh. Application-driven voltage-island partitioning for low-power system-on-chip design. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(3):316–326, 2009.

[146] Mingoo Seok, S. Hanson, D. Sylvester, and D. Blaauw. Analysis and optimization of sleep modes in subthreshold circuit design. In *Design Automation Conference, 2007. DAC '07. 44th ACM/IEEE*, pages 694–699, June 2007.

[147] Azam Seyedi, Vasileios Karakostas, Stefan Cosemans, Adrian Cristal, Mario Nemirovsky, and Osman Unsal. Nemscam: A novel cam cell based on nano-electro-mechanical switch and cmos for energy efficient tlbs. In *Proceedings of the 2015 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH¿¿ 15)*, pages 51–56. IEEE, 2015.

[148] Rishad A Shafik, Sheng Yang, Anup Das, Luis A Maeda-Nunez, Geoff V Merrett, and Bashir M Al-Hashimi. Learning transfer-based adaptive energy minimization in embedded systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 35(6):877–890, 2016.

[149] M. Shavezipur, K. Harrison, W. S. Lee, S. Mitra, H. S. P. Wong, and R. T. Howe. Partitioning electrostatic and mechanical domains in nanoelectromechanical relays. *Journal of Microelectromechanical Systems*, 24(3):592–598, June 2015. ISSN 1057-7157. doi: 10.1109/JMEMS.2014.2335157.

[150] Xiaoying Shen, Soogine Chong, and *et al.* 2d analytical model for the study of nem relay device scaling. In *Simulation of Semiconductor Processes and Devices (SISPAD), 2011 International Conference on*, pages 243–246, 2011. doi: 10.1109/SISPAD.2011.6035070.

[151] H. Shobak, M. Ghoneim, N. El Boghdady, S. Halawa, S. Iskander, and M. Anis. Power gating of vlsi circuits using mems switches in low power applications. In *Microelectronics (ICM), 2011 International Conference on*, pages 1–5, Dec 2011. doi: 10.1109/ICM.2011.6177407.

[152] Harmander Singh, Kanak Agarwal, Dennis Sylvester, and Kevin J Nowka. Enhanced leakage reduction techniques using intermediate strength power gating. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 15(11):1215–1224, 2007.

[153] Nipun Sinha, Timothy S Jones, Zhijun Guo, and Gianluca Piazza. Body-biased complementary logic implemented using aln piezoelectric mems switches. *Journal of Microelectromechanical Systems*, 21(2):484–496, 2012.

[154] Vijay K Sirigir, Khawla Alzoubi, Daniel G Saab, Fatih Kocan, and Massood Tabib-Azar. Ultra-low-power ultra-fast hybrid cnems-cmos fpga. In *2010 International Conference on Field Programmable Logic and Applications*, pages 368–373. IEEE, 2010.

[155] Sam Skalicky et al. A parallelizing matlab compiler framework and run time for heterogeneous systems. In *HPCC-CSS-ICESS*, pages 232–237. IEEE, 2015.

[156] M. Spencer, F. Chen, and *et al.* 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.

[157] Matthew Edmund Spencer. *Design Considerations for Nano-Electromechanical Relay Circuits*. PhD thesis, EECS Department, University of California, Berkeley, Jul 2015. URL http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-81.html.

[158] Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

[159] Ivan E. Sutherland. Micropipelines. *Communications of the ACM*, 32(6):720–738, 1989.

[160] Benjamin Z Tang, Stephen Longfield Jr, Sunil A Bhave, and Rajit Manohar. A low power asynchronous gps baseband processor. In *Asynchronous Circuits and Systems (ASYNC), 2012 18th IEEE International Symposium on*, pages 33–40. IEEE, 2012.

[161] B.Z. Tang, S.A. Bhave, and R. Manohar. Low power asynchronous vlsi with nem relays. In *Asynchronous Circuits and Systems (ASYNC), 2014 20th IEEE International Symposium on*, pages 85–92, May 2014. doi: 10.1109/ASYNC.2014.19.

[162] Min Tang, Ebin Liao, Cheng Kuo Cheng, Dayong Lee, R. Kumar, Yong Hean Lee, R. Shankar, O. Le Neel, G. Noviello, and F. Italia. A magnetostatic mems switch designed for portable applications. *Microelectromechanical Systems, Journal of*, 19(5):1131–1139, Oct 2010. ISSN 1057-7157. doi: 10.1109/JMEMS.2010.2067434.

[163] WH Teh, JK Luo, MR Graham, A Pavlov, and CG Smith. Near-zero curvature fabrication of miniaturized micromechanical ni switches using electron beam cross-linked pmma. *Journal of Micromechanics and Microengineering*, 13(5):591, 2003.

[164] Harrie AC Tilmans, E Fullin, H Ziad, Myriam DJ Van de Peer, J Kesters, E Van Geffen, J Bergqvist, M Pantus, Eric Beyne, Kris Baert, et al. A fully-packaged electromagnetic microrelay. In *Micro Electro Mechanical Systems, 1999. MEMS'99. Twelfth IEEE International Conference on*, pages 25–30. IEEE, 1999.

[165] R. Venkatasubramanian, S.K. Manohar, and P.T. Balsara. Nem relay-based sequential logic circuits for low-power design. *Nanotechnology, IEEE Transactions on*, 12(3):386–398, 2013. ISSN 1536-125X. doi: 10.1109/TNANO.2013.2252923.

[166] HDL Verilog. Verilog-ams language reference manual.

[167] Xiaohang Wang, Baoxin Zhao, Ling Wang, Terrence Mak, Mei Yang, Yingtao Jiang, and Masoud Daneshtalab. A pareto-optimal runtime power budgeting scheme for many-core systems. *Microprocessors and Microsystems*, 46:136–148, 2016.

[168] Huaizhi Wu, Martin DF Wong, and I-Min Liu. Timing-constrained and voltage-island-aware voltage assignment. In *Proceedings of the 43rd annual Design Automation Conference*, pages 429–432. ACM, 2006.

[169] Fei Xia, Andrey Mokhov, Yu Zhou, Yuan Chen, Isi Mitrani, Delong Shang, Danil Sokolov, and Alex Yakovlev. Towards power-elastic systems through concurrency management. *Computers & Digital Techniques, IET*, 6(1):33–42, 2012.

[170] Nuo Xu, Jeff Sun, I-Ru Chen, Louis Hutin, Yenhao Chen, Jun Fujiki, Chuang Qian, and Tsu-Jae King Liu. Hybrid cmos/beol-nems technology for ultra-low-power ic applications. In *2014 IEEE International Electron Devices Meeting*, pages 28–8. IEEE, 2014.

[171] Alex Yakovlev. On balancing regularity and burstiness in architecting new computer systems. Technical report, Newcastle University, 2012.

[172] Sheng Yang et al. Adaptive energy minimization of embedded heterogeneous systems using regression-based learning. In *Power and Timing Modeling, Optimization and Simulation (PATMOS), 2015 25th International Workshop on*, pages 103–110, Sept 2015. doi: 10.1109/PATMOS.2015.7347594.

[173] Pei Zhang, Christopher M Sadler, Stephen A Lyon, and Margaret Martonosi. Hardware design experiences in zebranet. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 227–238. ACM, 2004.

[174] Shifang Zhou, Xi-Qing Sun, and William N Carr. A micro variable inductor chip using mems relays. In *Solid State Sensors and Actuators, 1997. TRANSDUCERS'97 Chicago., 1997 International Conference on*, volume 2, pages 1137–1140. IEEE, 1997.