

**Hybrid NEMS-CMOS Architectures for Ultra Low Power Smart Systems
Architectures for Ultra Low Power Smart Systems**

Enachescu, Marius

DOI

[10.4233/uuid:58684b58-f0a6-4044-a70e-268d842ad7ec](https://doi.org/10.4233/uuid:58684b58-f0a6-4044-a70e-268d842ad7ec)

Publication date

2016

Document Version

Final published version

Citation (APA)

Enachescu, M. (2016). *Hybrid NEMS-CMOS Architectures for Ultra Low Power Smart Systems: Architectures for Ultra Low Power Smart Systems*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:58684b58-f0a6-4044-a70e-268d842ad7ec>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Marius Enachescu

Hybrid NEMS-CMOS
Architectures for Ultra Low
Power Smart Systems

Hybrid NEMS-CMOS Architectures for Ultra Low Power Smart Systems

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen

op dinsdag 12 april 2016 om 12:30 uur

Door

Marius ENACHESCU

Engineer Degree (equivalent with Bologna Master) in Microelectronics
The Faculty of Electronics, Telecommunications and Information Technology
University Politehnica of Bucharest
geboren te Boekarest, Roemenië

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. K.L.M. Bertels

Copromotor:
Dr. S.D. Cotofana

Samenstelling promotiecommissie:
Rector Magnificus, voorzitter Technische Universiteit Delft, NL
Prof. dr. K.L.M. Bertels, promotor Technische Universiteit Delft, NL
Dr. S.D. Cotofana, copromotor Technische Universiteit Delft, NL

Onafhankelijke leden:
Prof. dr. Y. Lebleblici Ecole Polytechnique Fed. Lausanne, CH
Prof. dr. J. Pineda de Gyvez Technische Universiteit Eindhoven, NL
Prof. dr. ing. A. Rubio Universitat Politecnica de Catalunya, SP
Prof. dr. P.J. French Technische Universiteit Delft, NL
Prof. dr. C. Dan University Politehnica of Bucharest, RO
Prof. dr. ir. G.J.T. Leus, reservelid Technische Universiteit Delft, NL

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Marius ENACHESCU

Hybrid NEMS-CMOS Architectures for Ultra Low Power Smart Systems
Delft: TU Delft, Faculty of Elektrotechniek, Wiskunde en Informatica - III
Thesis Technische Universiteit Delft.

Met samenvatting in het Nederlands.

ISBN 978-94-6186-630-1

Subject headings: nems, nemfet, power management, zero-energy, 3D-SICs.

Copyright © 2016 Marius ENACHESCU

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without permission of the author.

Printed in The Netherlands

*This dissertation is dedicated to my family and my girlfriend, Maria.
Their love and support helped me through all the trying times and have
pushed me to reach within myself to accomplish my goals.*

Summary

The availability of inexpensive and powerful processors provides the means for the computation ecosystem to change its fundamental paradigm towards the Internet of Things (IoT) where ubiquitous nano-systems add intelligence to every object that surrounds us. The new trend for most of those systems is to autonomously operate into a “zero-power” regime, i.e., manage their energy budget in such a way that they can provide the required functionality without any service until they become obsolete. Considering that these systems are most of the time inactive, the static power is the dominant power consumption component, thus the most effective way to fulfill the “zero-power” operation requirement is to diminish the energy consumption into the so called sleep/idle mode. The semiconductor community has been addressing the static power reduction issue at device level, but for the CMOS technology the effectiveness of such approach is limited by the interdependence between static power consumption and device performance. In view of this observation this thesis focuses on improving the energy efficiency of electronic products, battery-powered, and autonomous ones by making use of emerging leakage proof technologies in conjunction with the versatile CMOS counterpart. First, we performed a design space exploration to identify the most promising NEMFET geometries and to evaluate their potential performance in terms of switching delay, current capability, and leakage. Moreover we compared those parameters of interest with the ones offered by traditional transistors utilized in up to date CMOS technologies. Second, we assessed the NEMFET potential when utilized as sleep transistor in circuits featuring 2D cell based power gating, and find out if NEMFETs constitute a viable alternative to $High - V_{TH}$ FETs in sleep mode circuits. Furthermore, we proposed a novel 3D power management approach that attempts to alleviate issues associated with the NEMS utilization as sleep transistor in CMOS power gated integrated circuits. Given the two designs, we evaluated the 2D and 3D NEMFET based power management implementations energy efficiency when embedded into a computation platform executing a bio-medical sensing application. Third, we introduced a NEMFET based logic family tailored to the implementation of ultra-low energy functional units and processors. Fourth, we proposed a memory cell that relies on a NEMFET based inverter designed in such a way that no short circuit current can occur. Finally, we proposed and evaluated the “zero-energy” operation scenario potential of an improved version of the 3D-Stacked NEMS based power management architecture.

Samenvatting

De beschikbaarheid van goedkope en krachtige processoren biedt de mogelijkheid voor een computing ecosysteem om haar fundamentele paradigma te veranderen in de richting van Internet of Things (IoT), waar de alomtegenwoordige nano-systemen intelligentie toe voegen aan alle objecten die ons omringen. De nieuwe trend voor de meeste van deze systemen is het zelfstandig opereren in een “zero-power” regime, d.w.z. het zodanig beheren van hun energiebudget dat ze de vereiste functionaliteit kunnen leveren zonder service, totdat ze verouderen. Gezien het feit dat deze systemen de meeste tijd inactief zijn, is het statische vermogensverbruik het dominante vermogensverbruik, dus is de meest effectieve manier om te voldoen aan de “zero-power” eis het energieverbruik in de zogenaamde slaap/stand-by modus te verminderen. De halfgeleider gemeenschap heeft de kwestie van statische vermogensvermindering op device niveau geaddresserd, maar voor CMOS-technologie is de effectiviteit van een dergelijke aanpak beperkt door de onderlinge samenhang tussen statische stroomverbruik en de device prestaties. Gezien deze waarneming focust dit proefschrift zich op het verbeteren van de energie-efficiëntie van elektronische producten, batterij aangedreven en autonoom, door gebruik te maken van opkomende lekkage-bestendige technologieën in combinatie met de veelzijdige CMOS tegenhanger. Ten eerste hebben we een design space verkenning uitgevoerd om de meest veelbelovende NEMFET geometrieën te identificeren en om hun mogelijke prestaties op het gebied van switching delay, stroom capaciteit en lekstromen te evalueren. Bovendien vergeleken we die parameters die van belang zijn met de degenen die aangeboden werden door traditionele transistors, die gebruikt worden in up-to-date CMOS technologieën. Ten tweede, beoordeelden we het NEMFET potentieel wanneer deze wordt gebruikt als slaap transistor in circuits met op 2D cel-gebaseerde power gating en onderzochten we of NEMFETs een levensvatbare alternatief vormen voor $High - V_{TH}$ FET's circuits in slaapstand. Verder stelden we een nieuwe 3D energiebeheer methode voor, die issues probeert te verlichten geassocieerd aan het NEMS gebruik als slaap transistor in CMOS power gated geïntegreerde schakelingen. Voor de twee ontwerpen evalueerden we de energie-efficiëntie van de op 2D en 3D NEMFET gebaseerde energiebeheer implementaties, wanneer deze zijn ingebed in een computation platform die een bio-medische sensingapplicatie uitvoert. Ten derde, introduceerden we een op NEMFET gebaseerde logica familie gericht op de uitvoering van functionele eenheden en processors met ultra-lage energie. Ten vierde, stelden wij

een geheugencel berust op een NEMFET gebaseerde inverter voor die zodanig is ontworpen dat er geen kortsluitstroom kan optreden. Ten slotte hebben we het potentieel van de “zero-energy” operatie scenario voorgesteld en geëvalueerd voor een verbeterde versie van de op 3D-Stacked NEMS gebaseerde energiebeheer architectuur.

Acknowledgements

I think that for everyone the PhD road is a unique experience. For me it was a long road, curvy, often bumpy, nevertheless exciting. When looking back, I would repeat this experience all over again. This is perhaps also due to the people being together with me in this experience. I was fortunate to be surrounded by extraordinary people at both professional and personal levels. This chapter is all about you, the "team" standing besides me in this journey.

The first person that comes to my mind when thinking about my PhD is Dr. Sorin Cotofana, my co-promoter and daily supervisor. Sorin did not only gave me the opportunity to pursue a PhD, a long time dream, but also guided, encouraged, and challenged me in accomplishing this goal. I would like therefore to start by thanking Sorin for being a role model in my development as a scientist; by showing me how to change an engineering piece of work into a scientific publication; by being critical with myself as well as with others work; by being confident about and stand up for my work; by encouraging me to develop myself; by the inspiring talks; and by sharing his wisdom and philosophy on daily life. Also I would like to thank him for the special and delicious diner events during which we had the chance to chat about gastronomy, life, and religion.

I would like to express my sincere gratitude to Prof. Mircea Bodea at University Politehnica of Bucharest in Romania, for his supportive encouragement before and during my PhD studys. Prof. Bodea was my Engineer Degree thesis supervisor. I thank Alexandru Rusu not only for the valuable discussions on NEMFET model and for facilitating a better collaboration between TU Delft and EPFL during the NEMSIC project, but also for his unconditional friendship.

I would like to thank my promotor, Prof. dr. Koen Bertels, for many fruitful informal 1:1 (coffee breaks) in many difficult moments during the last few years during which I was assured of his unconditional support and encouragement. Also, for all the fun times spend together during so many diverse social events, e.g., BBQ, carting, bowling, soccer, Belgium beer, and spaghetti diner.

I also extend my thanks to the thesis committee for they spent their precious time to review my thesis. I would like also to thank other past and current faculty members in EWI, Arjan van Genderen, Robert Bogdan Staszewski, Anca Molnos, Said Hamdioui, and Zaid Al-Ars, for the interesting talks we had from time to time.

Special thanks are due to the CE secretary Lidwina Tromp for her administrative assistance, generous help, and nice talks that we had from to time. My thanks are also due to Bert, Erik, and Eef, the past and current CE system administrators for their technical support in setting up and maintaining the tools we used to run our simulations.

Over the years I had the opportunity to work with several PhD students. I thank George and Mihai not for the work we did together on developing the platform to simulate hybrid NEMS-CMOS architectures and developing the hybrid NEMS-CMOS memory architecture, but also for their support and friendship. My special thanks for the collaboration and the discussions related to various problems go to Rouzbeh, Saleh, Yao, Motta, Laiq, Seyab, Changlin, Nicoleta, Radu, Vlad, Razvan, Catalin, and to all other CE colleagues, you are to many to be mentioned individually.

I would like to thank Motta for helping me all the time with the Dutch translations, and also guiding me throughout the graduation steps. Also, special thanks to Mihai (again) for designing the elegant thesis cover in a very short time.

I was lucky to make many great Romanian friends during my stay in the Netherlands. Special thanks to the Bogdan, Remus, Razvan, Catalin, Radu, Dragos, Mihai, and Iulia. They helped me settle down in Delft and we shared a lot of happy time together. I would like to extend my special gratitude my good friends Andrei, Stefan, and Bogdan without whom I am sure life in the Netherlands would have been way more boring. I will always remember the good times we spent together (e.g., the trips, the chats, the celebrations, and the delicious meals we enjoyed).

Finally, I would like to thank my family. I am forever indebted to my parents who supported me continuously throughout the years. Many, many thanks to Maria for being an amazing partner. Thank you for being my most supportive friend and confidant!

Marius ENACHESCU

Delft, The Netherlands, April 2016

Contents

Summary	i
Samenvatting	iii
Acknowledgments	v
Table of contents	xi
List of Tables	xiii
List of Figures	xvii
List of Acronyms and Symbols	xvii
1 Introduction	1
1.1 Problem Statement	5
1.1.1 CMOS Delay-Leakage Vicious Circle	7
1.1.2 Is More than Moore providing the solution?	11
1.2 Research Questions	13
1.3 Dissertation Contributions	16
1.4 Dissertation Organization	18
2 Can NEMFET Replace FET In Sleep Circuits?	27
2.1 Introduction	28
2.2 NEMFET Background	28
2.3 Design Space Exploration	30

2.4	NEMFET vs. FET	32
2.5	Conclusions	33
3	NEMFET Based Power Management	37
3.1	Introduction	38
3.2	NEMFET Background	38
3.3	Design Space Exploration	40
3.4	90 nm CMOS 32-bit Adder Analysis	43
3.5	Conclusions	45
4	Advanced NEMS-based Power Management for 3D Stacked ICs	47
4.1	Introduction	48
4.2	Nano-Electro-Mechanical FET	50
4.3	Power Management in 3D Stacked ICs	52
4.4	Experimental setup and results	54
4.5	Conclusions	56
5	Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture for Autonomous Sensors Systems	59
5.1	Introduction	60
5.2	NEMFET based power management architecture	61
5.2.1	NEMFET Background	61
5.2.2	3D-Stacked Hybrid Power Management Architecture	63
5.3	Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture	64
5.3.1	Isolation cells	64
5.3.2	Power Management Controller	66
5.4	Performance evaluation	66
5.4.1	3D-Stacked NEMFET-based Power Management Architecture evaluation	68
5.4.2	Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture evaluation	68

5.5	Conclusions	72
6	Ultra Low Power NEMFET Based Logic	77
6.1	Introduction	78
6.2	NEMFET Background and Compact Modeling	79
6.3	Short Circuit Free NEMFET-based Logic	81
6.4	NEMFET-based Power Management Logic	83
6.4.1	Case Study	86
6.5	Conclusions	88
7	Low-Leakage 3D Stacked Hybrid NEMFET-CMOS Dual Port Memory	91
7.1	Introduction	92
7.2	Background	94
7.2.1	SRAM Energy Consumption	94
7.2.2	NEMFET Background and Basic Operation	95
7.3	NEMFET Inverter as Storage Structure	98
7.3.1	NEMFET Inverter Stability	98
7.3.2	NEMFET Inverter Scalability and Variability	100
7.4	3D-Stacked Hybrid NEMFET-CMOS Memory	107
7.5	3D Hybrid NEMFET-CMOS Dual Port Memory vs. 2D Dual Port SRAM	110
7.5.1	Evaluation Methodology	111
7.5.2	Memory Cell	113
7.5.3	Memory Array	116
7.6	Conclusions	123
8	Is the Road Towards “Zero-Energy” Paved with NEMFET-based Power Management?	131
8.1	Introduction	132
8.2	Nano-Electro-Mechanical Devices as Replacement for MOSFET	134
8.2.1	Nano-Electro-Mechanical Field Effect Transistor . . .	134

8.2.2	NEM Relays	135
8.3	Power budgeting of energy harvesters	136
8.4	Results Evaluation and Discussion	138
8.5	Conclusions	140
9	Conclusions and Future Work	143
9.1	Summary	144
9.2	Future Research Directions	147
	List of Publications	151
	Curriculum Vitae	155

List of Tables

2.1	Optimized NEMFET instances for low switching times and high I_{ON} .	32
2.2	Optimized NEMFET instances.	33
3.1	Optimized NEMFET instances for low switching times and high I_{ON} .	43
4.1	Results	55
4.2	Power switches dimensioning	56
5.1	Power and energy results	69
5.2	Leakage-enhanced architecture area, power and energy results	70
6.1	CMOS vs. NEMFET gates	83
6.2	CMOS vs. NEMFET - variable fan-in NAND	84
6.3	PM Circuitry Power Consumption	88
7.1	NEMFET and CMOS Inverter Static Power and Dynamic Energy Consumption	106
7.2	Dual-port Memory Cell Footprint	114
7.3	Memory Footprint and Area Efficiency	117
8.1	Various energy sources and harvested power densities [11]	137
8.2	Energy budgeting	139

List of Figures

1.1	Active and sleep power consumption for a battery operated sensor node.	4
1.2	Typical transfer characteristics plots (I_D - V_G) for an nMOS. I_D is plotted in (a) linear scale and (b) log scale	7
1.3	a) Typical CMOS inverter schematic; b) transfer characteristics plot (V_{OUT} - V_{IN}); c) CMOS inverter current (I_{OUT}) versus V_{IN}	8
2.1	NEMFET' geometry, transfer characteristic, and equivalent capacitive divider	29
2.2	NEMFET I_{ON} , I_{OFF} , and Propagation Delay Analysis	31
3.1	NEMFET' geometry and transfer characteristic	39
3.2	NEMFET I_{ON} , I_{OFF} , and Propagation Delay Analysis for $t_{gap0} = 20\text{nm}$	42
3.3	NEMFET I_{ON} , I_{OFF} , and Propagation Delay Analysis for $t_{gap0} = 10\text{nm}$	42
4.1	3D Stacked NEMFET Based Power Management Architecture	49
4.2	Nano-Electro-Mechanical FET	51
4.3	Detailed Representation of 3D Stacked NEMFET Based Power Management Architecture	53
5.1	(a) Schematic diagram of NEMFET. (b) Equivalent circuit model for the NEMFET	62
5.2	R_{ON} and I_{OFF} for NEMFET and 65 nm High- V_t CMOS Switch Transistors	63

5.3	Leakage-enhanced 3D Stacked NEMFET Power Management Architecture	65
5.4	Substitution of nMOS Pull-down Isolation cell with NEMFET cell	66
5.5	Simulated waveform of Power Management Controller signals	66
5.6	System on Chip platform for autonomous sensors	67
5.7	Breakdown of leakage power (in nW) in an embedded processor for autonomous sensors	71
5.8	Energy Consumption versus Duty-cycle	72
6.1	Illustrative cross-section of NEMFET. The two states, e.g., pull-out (OFF) and pull-in (ON), are depicted	80
6.2	NEMFET inverter transfer characteristics	81
6.3	NEMFET inverter SC current analysis	82
6.4	Transfer characteristics for NEMFET NOR/NAND	83
6.5	NEMFET-based isolation cells in commercial designs and associated truth table	85
6.6	Heterogeneous state retention cell	86
6.7	State retention signal wave forms	87
7.1	NEMFET Suspended-gate Illustrative Cross-section: the Two States, i.e., Pull-out (OFF) and Pull-in (ON), are Depicted . . .	96
7.2	NEMFET Inverter Schematic and its Hysteretic Transient Behaviour	97
7.3	NEMFET-based Inverter Transfer Characteristics	98
7.4	CMOS vs. NEMFET Inverter Noise Margin	99
7.5	NEMFET Inverter Stability Analysis: $W_{BEAM}=45/65/90$ nm, $H_{BEAM}=10$ nm, and $gap=10/15/20$ nm	100
7.6	A Macro-Model representation of the compact modelling approach for the NEMFET	101
7.7	NEMFET Inverter Stability Analysis: $V_{DD}=1.2V$, $W_{BEAM}=45nm$, $H_{BEAM}=10nm$, and $gap=10/15/20$ nm	104
7.8	Hybrid NEMFET-CMOS Memory Cell Electric Scheme . . .	108

7.9	3D Hybrid NEMFET-CMOS Memory Cell Noise Margin Definition	109
7.10	NEMS-CMOS Memory Cell Stability Analysis: $V_{DD}=1/1.2V$, $Air-gap=15nm$, and $W_{BEAM}=45nm$	110
7.11	Proposed 3D Hybrid NEMFET-CMOS Memory	111
7.12	Two 3D-HtmpMCs Layout: NEMFET Inverters Tier (top) and 45nm CMOS Access Logic Tier (bottom)	112
7.13	Schematic of 10T-DPMC from [19]	113
7.14	3D-HdpMC vs 10T-DPMC Dynamic Energy and Leakage for Different Loads	115
7.15	64-bit Word Width Memories Write Access Time	118
7.16	64-bit Word Width Memories Read Access Time	120
7.17	Static Energy (Leakage) Contribution to Total Energy Consumption (64-bit Memories)	120
7.18	Transient Probability Influence on the Total Energy (8-KB Memory Array)	121
7.19	Activity Factor Impact on the Total Energy 8-KB Memory Sizes	122
7.20	Total Energy for Various Write-Read Ratio and 50% Constant Transition Probability (8-KB Memory Array)	123
8.1	Emerging autonomous hybrid 3D stacked bio-sensor embodiment	133
8.2	Schematic diagram of (a) NEMFET, (b) 3T NEM-Relay, and (c) 4T NEM-Relay	134
8.3	System-level power supply architecture	137

List of Acronyms and Symbols

<i>3D – HdpMC</i>	3D stacked hybrid dual port NEMS-CMOS memory cell
<i>10T – DPMC</i>	Dual port SRAM memory cell
<i>AO</i>	Always on
<i>BB</i>	Body Biasing
<i>CMOS</i>	Complementary Metal-Oxide-Semiconductor
<i>CNTFET</i>	Carbon Nanotube FET
<i>DIBL</i>	Drain Induced Barrier Lowering
<i>DVS</i>	Dynamic Voltage Scaling
<i>E</i>	Young’s Modulus
<i>ECG</i>	Electrocardiogram
<i>FDSOI</i>	Fully Depleted Silicon on Insulator
F_{e1}	Electrical attractive force
F_s	Gate-beam spring resistance
<i>FM</i>	Ferromagnetic
<i>FU</i>	Functional Unit
<i>FEM</i>	Finite Element Modeling
<i>GL</i>	Gate Leakage
<i>IC</i>	Integrated Circuit
<i>IoT</i>	Internet of Things
I_{OFF}	OFF current
I_{ON}	ON current
I_D	Drain current
<i>ISO</i>	Isolation cells
<i>ITRS</i>	International Technology Roadmap for Semiconductors
<i>h</i>	The thickness of the suspended gate
k_{beam}	The lumped linear spring constant of the beam
L_{beam}	The length of the beam
<i>LP</i>	Low Power
<i>MIC</i>	Maximum Instantaneous Current
<i>MIPS</i>	Million Instructions Per Cycle
<i>MOSFET</i>	Metal-Oxide-Semiconductor Field-Effect-Transistor

<i>MPEG</i>	Moving Pictures Experts Group
<i>MTCMOS</i>	Multi-Threshold CMOS
<i>MtM</i>	More than Moore
<i>NEMFET</i>	Nano-Electro-Mechanical Field-Effect-Transistor
<i>NEM – Relay</i>	Nano-Electro-Mechanical Relay
<i>NM_L</i>	Noise Margin Low
<i>M_H</i>	Noise Margin High
<i>PI</i>	Pull-In
<i>PO</i>	Pull-Out
<i>PG</i>	Power Gating
<i>PM</i>	Power Management
<i>PMC</i>	Power Management Controller
<i>R_{ON}</i>	ON state resistance
<i>SCCF</i>	Short-Circuit-Current Free
<i>SET</i>	Single-Electron Tunneling Junctions
<i>SoC</i>	System-on-Chip
<i>SPICE</i>	Simulation Program with Integrated Circuit Emphasis
<i>SR</i>	State Retention
<i>SRAM</i>	Static random-access memory
<i>SS</i>	Subthreshold Slope
<i>ST</i>	Switch (Sleep) Transistors
<i>TFET</i>	Tunnel FET
<i>t_{gap0}</i>	The gap between the oxide and the suspended gate
<i>t_{ox}</i>	The thickness of the gate oxide
<i>TSV</i>	Through Silicon Via
<i>V_G</i>	The gate voltage
<i>V_{KD}</i>	The keep-data voltage
<i>V_t</i>	The threshold voltage
<i>V_{PI}</i>	The pull-in voltage
<i>V_{PO}</i>	the pull-out voltage
<i>VHDL</i>	Very High Scale Integrated Circuits Hardware
<i>W_{beam}</i>	The width of the beam
<i>WF</i>	Gate Workfunction

1

Introduction

The Curta mechanical portable calculator introduced by Curt Herzstark in 1948, was a revolutionary innovation able to perform real-time complex mathematical function evaluations, without the use of electricity or batteries. Starting with early 1960s, such portable mechanical calculators were embedded in rally sport cars for real-time computation of time to checkpoints, distances off-course, and so on. Later on, Curta's utilisation expanded to both commercial and general-aviation flights, because of its ability to precisely calculate airplanes' weight and balance, and power supply free autonomous operation. However, Curta's main drawback has been the computational latency owed by its hand movement limited mechanical handling.

While the mechanical computation era was coming to its end, most digital computers built at the beginning of the 20th century, e.g., Z2 [1], performed calculation with electromechanical devices - electric switches drove mechanical relays. Such computers had a low operation speed (large latency) when compared with the mechanical counterparts and were eventually superseded by much faster all-electric ones, e.g., Z3, Atanasoff-Berry, Colossus, and ENIAC, which were built by hand, using relays and valves (vacuum tubes) [2]. While being faster than (electro)mechanical counterparts they could not have any impact on the portable computation market due to their huge dimensions and power consumption, e.g., the ENIAC machine was weighing 30 tons, containing over 18000 vacuum tubes, 1500 relays, and hundreds of thousands of resistors, capacitors, and inductors while using 200 kilowatts of electric power [2].

However, with the advent of Integrated Circuits (ICs) in 1958 by Jack St. Clair Kilby at Texas Instruments [3], portable electronic calculators emerged to meet the demand of solving larger computational problems with relatively higher speed and reduced footprint. One of the major steps towards this was the

introduction in 1971 of the first Intel single-chip microprocessor, i.e., Intel 4004, build with 2300 transistors on a 12 mm^2 die and able to operate at a maximum clock frequency of 108KHz. Employing a $10\mu\text{m}$ silicon-gate enhancement load pMOS technology, Intel 4004 could execute a total of 0.092 Million Instructions Per Cycle (MIPS), while an 8-bit addition took $850\mu\text{s}$ or 79 cycles [4], for only about 1 Watt power consumption. Hence, ICs created the premises for the realisation of portable electronic computers but could not initially entirely eliminate the mechanical counterparts due to their large size, reduced precision, high price, and relatively high power consumption.

Since 1960s the IC fabrication technology advanced and the planar bulk silicon MOS Field Effect Transistor (MOSFET) became the main IC building block. The MOSFET improvement pace was predicted by Gordon Moore in 1968 by stating that due to transistor scaling, the number of components per IC chip will double each and every two years [5]. Although Moore foresaw that the scaling will last only for a decade, semiconductor technology improvements preserved this trend to this day, which legendary became known as "Moore's Law". Following Moore's exponential growing trend, more components within the same chip area were accommodated and nowadays, the high-end microprocessor Intel Core i7 (2014 edition) is exceeding 1 billion transistors within a die size of 257 mm^2 . It runs at a clock frequency of 4GHz [4] and can execute 20 kMIPS at the expense of about 88W power consumption.

Thus, due to the spectacular fabrication technology evolution that allowed for the realisation of ICs with diminishing size, enhanced precision, and affordable price, portable mechanical computers were eventually superseded by fully-electronic computational systems. Moreover, MOSFET shrinking allowed for the proliferation of microprocessors in many other aspects of our life starting from home appliances to toys. However, for the newly developed battery powered application-specific microprocessors, power/energy requirements tightened. More important, shrinking by itself is creating power density issues, hence power density became the primary constraint [6]. As a response to these market and fabrication technology developments the computer engineering research and design community went through an optimization goal switch from high performance to energy effective computation. In this line of action solutions have been proposed and implemented to: (i) diminish the power density while conserving performance by means of parallelism, and (ii) better manage the available energy.

A notable cornerstone was Intel's decision to follow the paradigm change initiated by IBM's Power 4 and Sun Microsystems' Niagara processors and to

announce in 2005 that its high-performance microprocessors would henceforth rely on multiple processors or cores [7]. Hence, multi-core microprocessor chips (with up to 8 integrated computing cores) became the norm for mobile applications. Subsequently, in 2012 ARM released its most energy effective processor intended for deeply embedded applications that require area and power consumption optimized computation facilities, the ARM Cortex-M0+, which consumes $3\mu\text{W}/\text{MHz}$ [8].

Thus, due to the availability of inexpensive and powerful processors, the computation ecosystem went into a fundamental paradigm change towards the Internet of Things (IoT) where ubiquitous nano-systems add intelligence to almost every object that surrounds us such that sensing and actuating functionalities are "hidden" within the environment. The new trend for most of those systems is to autonomously operate into a "zero-power" regime, i.e., manage their energy budget (from battery and scavenging) in such a way that they can provide the required functionality without any service (battery replacement) until they become obsolete.

We note that most of the IoT nano-systems experience long idle periods, require low energy calculations, are context aware, and able to interact wirelessly with people and with each other. Furthermore, their computation and memory requirements are growing while the available energy resources do not increase [9].

From the power consumption perspective, there are three main power modes in a duty-cycled IoT nano-system, as graphically depicted in Figure 1.1:

- *Operation*: Dynamic power is consumed by the active system;
- *Stand-by*: Static power is consumed by the idle system; and
- *Power-up/Power-down*: State switching power is consumed during on \leftrightarrow off state transitions.

Considering that these systems are most of the time inactive, e.g., a Zigbee sensor node is 99.9% of the time in sleep mode, waking up periodically for a few milliseconds [9], the static power is the dominant power consumption component (as graphically suggested in Figure 1.1), thus the most effective way to fulfil the "zero-power" operation requirement is to diminish the energy consumption into the so called sleep/idle (stand-by) mode [10].

The semiconductor community has been addressing the static power reduction issue at device level, but for the CMOS technology the effectiveness of such

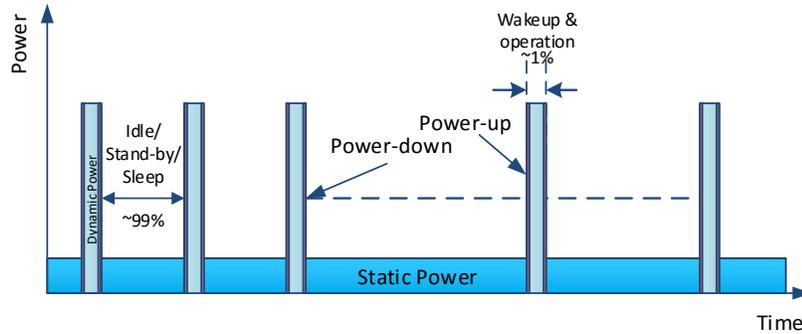


Figure 1.1: Active and sleep power consumption for a battery operated sensor node.

a approach is limited by the interdependence between static power consumption and device performance. As MOSFET feature size is scaling down the supply voltage value (V_{DD}) follows the same trend to reduce the active power consumption, which quadratically depends on V_{DD} . However, supply voltage scaling increases MOSFET's switching delay unless the transistor threshold voltage is also scaled down, which at its turn results in a significant static power consumption increase.

Therefore, for a given application to perform a task within a given amount of time, a clear tradeoff between the static and the active power consumption, i.e., power supply and transistor threshold values, has to be identified [11]. However, as IoT nano-systems evolution enabled the execution of a wide mixture of functions on a single die, it became increasingly difficult to find an optimal point applicable to all circuit blocks on the same die, which suggests that alternative static power reduction/management avenues have to be investigated.

In view of this observation this thesis focuses on improving the energy efficiency of electronic products, especially portable, battery-powered, and autonomous ones by making use of emerging leakage proof technologies in conjunction with the versatile well-established CMOS counterpart.

1.1 Problem Statement

According to the International Technology Roadmap for Semiconductors (ITRS) [12], the following static power consumption contributing factors are projected to exponentially increase with MOSFET feature size scaling:

- **subthreshold leakage**, a weak inversion current across the device; and
- **gate leakage**, a tunneling current through the gate oxide insulation.

By scaling threshold voltage as a natural effect of MOSFET feature size scaling, the subthreshold leakage current increases exponentially, i.e., for 100mV threshold voltage decrease, the subthreshold leakage current increases $10\times$. Moreover, with temperature, the subthreshold leakage increases about $10\times/100^\circ\text{C}$ and the gate leakage increases about $2\times/100^\circ\text{C}$ [13].

In an attempt to diminish the severity of these device scaling related phenomena, the static power consumption reduction issue has been addressed at circuit level by techniques such as Dynamic Voltage Scaling (DVS) [14], Power Gating (PG) [15, 16], Body Biasing (BB) [17], forced transistor stacking [18], and multi-threshold voltage designs [19].

The most effective techniques, i.e., PG and DVS, use power supply voltage, V_{DD} , as the primary knob for reducing leakage currents, by gating a circuit from its power supply, or by lowering V_{DD} to reduce leakage, respectively. Specifically, PG relies on placing Switch (Sleep) Transistors (STs) between power/ground rails and the to be isolated Functional Unit (FU). When STs are active, power is supplied through them to the FU providing normal operation conditions, and, respectively, when the STs are open the power supply is cut-off. One can observe from Figure 1.1 that the main PG consequence is that the FU (power gated block) leakage power is now determined by the STs leakage and PG may result in substantial power savings if low leakage devices are utilised as STs. Low-Leakage *High- V_{TH}* Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs) were initially utilized as STs, while later on, various enhancements based on Multi-Threshold CMOS (MTCMOS) technologies have been proposed to further diminish the *High- V_{TH}* ST leakage current [20]. By applying such enhanced PG techniques to several benchmark circuits up to 90% leakage power savings have been reported in [21].

Dynamic Voltage Scaling is a supply voltage adaptation method, which relies on a voltage regulator to dynamically scale the supply voltage for various digital circuits as their workload varies with time. It allows the application

to dynamically change FUs processing performance such that the lowest possible power consumption is achieved while maintaining a given application execution scenario specific throughput. At the expense of additional power management circuitry DVS can provide up to $4.5\times$ energy reduction for less compute-intensive AUDIO benchmarks executed on a general purpose 8-bit microprocessor while for computing-intensive MPEG application it delivers only a 11% energy consumption reduction [14].

Body biasing is a V_{TH} regulation method, which relies on biasing the substrate/wells on a die by something else than GND (in case of n-channel MOS) or V_{DD} (in case of p-channel MOS). Hence by dynamically adjusting the V_{TH} value one can boost performance when the system is active and reduce transistor subthreshold leakage thus the static power, when the system is idle. However, the reverse body biasing approach worsens short channel effects like Drain Induced Barrier Lowering (DIBL), and increases V_{TH} variation across a die, which makes it less effective with technology scaling [19].

The leakage current flowing through series connected transistors decreases with the number of "OFF" transistors increase. For example, when considering a 2-input NAND logic gate with already build in 2 stacked (series) nMOS transistors, the leakage when both nMOS transistors are "OFF" is 1 order of magnitude lower when compared with the situation when only one nMOS transistor is "OFF" [22, 23]. However, for digital circuits, i.e., 32-bit Kogge Stone adder, the standby leakage varies with 30%-40% [23] based on the number of transistor stacks in the design with more than one "OFF" device. The number of series "OFF" transistors depend on the input vectors values. Hence, transistor stacking method emerged as a static power reduction technique for digital circuits by storing a predefined input vector sequence and generating it during standby so that the number of nMOS/pMOS stacks is maximized to more than one OFF device. This approach has a low impact on the area overhead due to the presence of stack transistors in a large number of logic gates, for example pMOS stack in NOR logic gates and nMOS stack in NAND logic gates. However, the transistor stack method requires additional active power consumption to switch to the desired low leakage state and for more complex digital circuits the leakage reduction may become trivial [19].

In view of the previous discussion we can conclude that the efficiency of all the above leakage reduction techniques is determined by the MOSFET' leakage and the designer ability to control it by means of, e.g., BB, with negligible circuit performance reduction. Hence, in an attempt to determine a power management efficiency upper bound for current and future CMOS technologies we

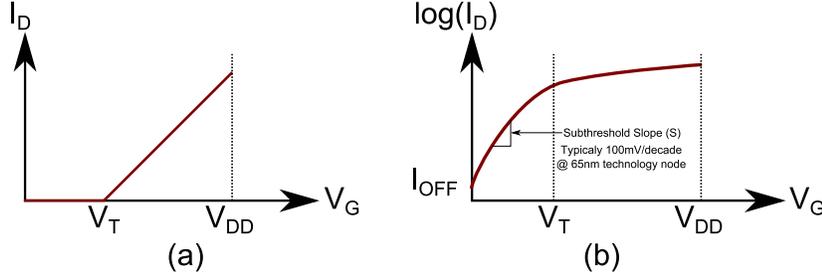


Figure 1.2: Typical transfer characteristics plots (I_D - V_G) for an nMOS. I_D is plotted in (a) linear scale and (b) log scale

analyze in the next subsection the MOSFET capability to effectively reduce the overall energy consumption.

1.1.1 CMOS Delay-Leakage Vicious Circle

A MOSFET is essentially speaking an electronic switch whose *ON/OFF* state is controlled by the potential difference between its gate and source (V_{GS}). If we increase V_{GS} as depicted in Figure 1.2a, at some point, i.e., the threshold voltage V_{TH} , the source potential barrier height becomes insignificant such that the carriers can easily diffuse into the channel region. This happens when the transistor is towards inversion (*ON* state). On the other hand, if we decrease V_{GS} , at the same point (V_{TH}) the mobile positively charged holes are attracted to the region beneath the gate. This happens when the device leaves the depleted region towards accumulation (*OFF* state) [24]. The MOSFET drain current (I_D) is plotted on a logarithmic scale against V_{GS} in Figure 1.2b to show the mild transition between the two operation regions. Hence, the *ON/OFF* state current (I_{ON}/I_{OFF}) of a MOSFET can be approximated by the following equations:

$$I_{ON} \propto \mu C_{OX} \frac{W}{L} (V_{DD} - V_{TH})^2 \quad (1.1a)$$

$$I_{OFF} \propto 10^{-V_{TH} / \overbrace{\ln 10 \frac{kT}{q} \left(1 + \frac{C_{DEP}}{C_{OX}}\right)}^{\text{Subthreshold Slope (S)}}}, \quad (1.1b)$$

where μ is the carrier mobility, C_{OX} the gate capacitance, C_{DEP} the depletion capacitance, $\frac{kT}{q}$ the thermal voltage (26mV at room temperature), and W and L are the transistor gate width and length, respectively.

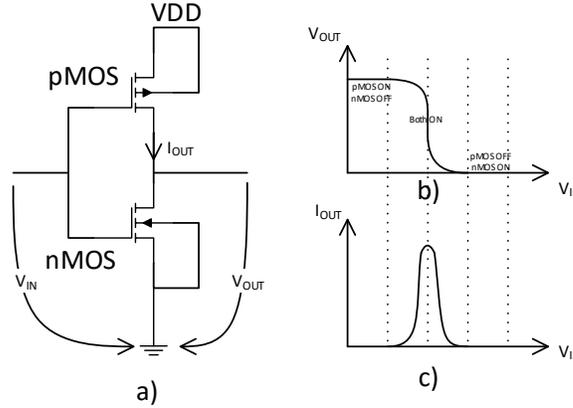


Figure 1.3: a) Typical CMOS inverter schematic; b) transfer characteristics plot (V_{OUT} - V_{IN}); c) CMOS inverter current (I_{OUT}) versus V_{IN}

According to ITRS [25], V_{TH} is predicted to scale well below 300mV in the next few years, which results in the reduction of the MOSFET's V_{GS} margin required to mitigate the I_D value when the transistor is turned *OFF*. From Figure 1.2b, it can be noticed that by decreasing V_{GS} below V_{TH} , I_D decreases exponentially with respect to V_{GS} , hence by mitigating the inverse slope of the $\log I_D$ - V_G curve, i.e., the Subthreshold Slope (SS), the I_{ON}/I_{OFF} ratio increases. Therefore, considering I_{ON} constant, I_{OFF} reduces with respect to the SS , according to (1.1b). However, ideally, $C_{OX} \gg C_{DEP}$, hence the minimum achievable SS becomes $\ln 10 \frac{kT}{q}$, i.e., 60mV/decade at room temperature, leading to a maximum I_{ON}/I_{OFF} ratio of only 10^5 . Regrettably, in real life conditions, e.g., for a commercial 65nm technology, SS becomes 100mV/decade, leading to a significant I_{OFF} increase to about 0.1% of I_{ON} . This is very unfortunate as, e.g., given that in a general purpose microprocessor, only about 10% of the devices are active during normal operation [26], the energy consumed by the idle devices becomes equal with the energy consumed by the active ones, which are producing the useful calculation, thus for the above considered 65nm technology, 50% of the energy is lost due to leakage.

Apart of the due to leakage energy lost current ICs waste energy also in the active devices. This relates to the fact that MOSFET based Boolean logic gates rely on n -type and p -type transistors to achieve complementary switching be-

havior, i.e., only one device is turned on at a time when the gate voltage is high (V_{DD}) or low (GND), which is the modus operandi that made Complementary MOS (CMOS) the most spread digital IC design style and technology. If we analyze the inverter depicted in Figure 1.3a, which is the most popular CMOS digital circuit, we observe that the p -channel MOSFET turns on when its gate voltage is low to help "pull-up" the output node to V_{DD} , while the n -channel MOSFET turns on when its gate voltage is high to help "pull-down" the output node to ground (GND), as illustrated in Figure 1.3b. Consequently, during the inverter output switching process a direct unwanted current path from V_{DD} to GND is formed which leads to a short-circuit current that does not help calculations, i.e., it does not charge inverter's output, as Figure 1.3b indicates. A theoretical solution to overcome the inverter non-ideal switching behaviour can be found by examining Figure 1.2b: if the SS can be made steeper around V_{TH} , below 60mV/decade, to improve the I_{ON}/I_{OFF} ratio, the CMOS inverter would experience a lower I_{OFF} , making further energy efficiency improvements possible. Fortunately, when the inverter is not switching, the direct current path is not forming, thus the static power dissipation is limited to the leakage component [27].

Generally speaking, the energy dissipated by a system over a time interval T (see Figure 1.1) corresponds to the integral of the instantaneous consumed power as follows:

$$E = \int_0^T P(t) dt \quad (1.2)$$

For the particular case of a CMOS circuit the total energy consumption relates to two major components as indicated in (1.3): (i) The Dynamic Energy (E_{ON}) consisting of Switching Energy (E_{SW}) spent for load capacitance charging, and Short-Circuit Energy (E_{SC}) dissipated when both nMOS and pMOS transistors are partially ON , and (ii) Static Energy (E_{OFF}) consisting of subthreshold leakage (I_S), gate leakage (I_G), junction leakage (I_J), and contention current (I_{CT}):

$$E = \underbrace{\alpha f L_D C V_{DD}^2}_{E_{SW}} + \underbrace{\alpha f L_D I_{SC} V_{DD}}_{E_{SC}} + \underbrace{f L_D (I_S + I_G + I_J + I_{CT})}_{I_{OFF}} V_{DD} \tau_D, \quad (1.3)$$

where α is the activity factor, L_D the logic depth, f the average fan-out of the logic gates through which the signal needs to travel from source to destination, C is circuit intrinsic capacitance, and τ_D the system's latency per operation.

Until recently, according to ITRS [28], in parallel with declining fabrication cost per device, digital logic device size, delay, and supply voltage have been

also reduced. In relation with this trend we can conclude from (1.3) that V_{DD} scaling is effective for diminishing E_{SW} (quadratic dependency of V_{DD}) but has limited impact on E_{SC} and E_{OFF} (linear dependency of V_{DD}).

We note that when designing a digital circuit expected to experience long idle periods, e.g., IoT devices, a parameter that can extend or reduce its duty cycle is the per operation delay [27] defined as:

$$\tau_D = \frac{L_D f C V_{DD}}{2 I_{ON}} \quad (1.4)$$

According to (1.1a), (1.4), and Figure 1.2b, when V_{DD} scales down, I_{ON} reduces and τ_D increases, which results in a system performance degradation, thus in a longer duty cycle and by implication on a higher energy consumption. The circuit's performance could in turn be maintained if V_{TH} is reduced while $V_{DD} - V_{TH}$ is maintained constant, however, this is not an effective solution as it results into a MOSFET I_D - V_G characteristic left shift, i.e., in an I_{OFF} increase, which is closing the CMOS delay-leakage vicious circle.

Another way to maintain the circuit performance is by scaling MOSFET's channel length. However, CMOS technology is approaching its physical limits due to an onslaught of new challenges, i.e., lithography limitations for printing sub-wavelength features, short-channel effects, increasing static leakage with threshold voltage reduction, increasing transistor performance variability, etc. Furthermore, due to the fact that the thermal voltage $k_B * T/q$ does not scale with transistor dimensions, the threshold voltage (V_{TH}) reaches a scalability frontier, which is enforcing a saturated supply voltage value [29]. In practice, for a 65nm n-channel MOSFET at 70°C, $SS=100\text{mV/decade}$, and $V_{TH} \approx 300\text{mV}$ [24], hence, only ≈ 3 decades separate I_{ON} from I_{OFF} . According to ITRS [25], with the MOSFET feature size dropping to 22nm, $\approx 100\text{mV}$ V_{TH} scalability frontier was reached. Hence I_{ON}/I_{OFF} ratio became ≈ 1 decade, which is by no means the best fit solution for portable, battery-powered, autonomous systems implementation.

To sum up, CMOS can certainly deliver the required performance for current and future IoT nodes, moreover, scaling helps on this matter. Oppositely, I_{ON}/I_{OFF} ratio cannot be improved beyond $10^{V_{DD}/(2 \times 0.1)}$, which limits the effectiveness of any kind of MOSFET based power management technique. Given that IoT nodes tend to be more and more idle and the active power consumption diminished by means of architectural choices and technology scaling, I_{OFF} is becoming more and more the dominant energy consumption contributor. Thus we need to investigate hybrid solutions which rely on CMOS logic for calculation and another alternative device with a lot larger I_{ON}/I_{OFF}

ratio for power management.

In the next section we briefly analyze state of the art emerging devices in an attempt to identify potential candidates that are better fitted for portable, battery-powered, autonomous systems implementation.

1.1.2 Is More than Moore providing the solution?

As an alternative, emerging More than Moore (MtM) technologies have been proposed and investigated to overcome the basic CMOS device I_{ON}/I_{OFF} ratio, e.g., Carbon Nanotube FETs (CNTFETs), spin transistors, ferromagnetic nanodots, tunnel FETs, quantum-dot cellular automata, Single-Electron Tunneling Junctions (SETs), Nano-Electro-Mechanical Relays (NEM-Relays), and NEMFETs [12]. If one candidate technology is to partially replace the MOSFET in the context of portable, battery-powered, autonomous systems implementation, it must satisfy as many of the following requirements as possible: (i) given the anticipated device density, the candidate technology should have extremely low I_{OFF} and switching power consumption, and (ii) the device switching should occur sufficiently fast, such that it can compete with or even outperform "classic" CMOS-based computation in terms of active energy consumption. In the following we briefly discuss the mentioned devices by focussing on the following aspects: (i) I_{OFF} , (ii) if possible I_{ON}/I_{OFF} ratio, and (iii) other aspects that facilitate or preclude their utilization as ultra-low I_{OFF} device, in an attempt to identify the most promising one in view of the previous discussion.

CNTFETs could potentially increase the switching speed and minimize SS (i.e., minimize the short channel effects) by a surround gate geometry due to the charge carriers high mobility [30], when compared with 32nm CMOS. Lin et al. [31] reported I_{ON}/I_{OFF} ratios from 5×10^3 to 5×10^5 with a minimum I_{OFF} of about 100pA. On the other hand, there are multiple challenges one has to face in order to achieve such ratios, including: (i) ability to control the bandgap, (ii) growth of the nanotubes in required locations and directions, (iii) control of charge carrier type and concentration, (iv) deposition of a gate dielectric, and (v) formation of a low resistance electrical contact [12].

Spin transistors exhibit transistor behavior by making use of magnetoresistive devices. Spin transistor's primary feature is the ability to control its output via spin or magnetization [32] [33]. I_{ON}/I_{OFF} ratios of 10^4 were reported in [34], however, remaining core issues that require addressing are the injection of a high percentage of spin-polarized electrons from a half-metal source into the

channel, and the interconnection of such devices [12].

Ferromagnetic (FM) logic devices operate on a completely different paradigm than spin devices as they rely on the individual spin dynamics of one or a few charge carriers [35]. FM devices store the computational state by means of the local magnetization orientation of a domain of a ferromagnetic material. Moreover, they have the potential of being non-volatile and radiation hard, which is derived from the properties of the ferromagnetic materials themselves. I_{ON}/I_{OFF} ratios larger than 10^4 for hole and 10^2 for electron conduction regimes were reported in [36]. However, a very large SS of about 11V per decade was observed, which is equivalent with a SS reported in other 2D semiconductor devices with 285nm SiO_2 dielectric. This large SS is attributed to the thicker SiO_2 backgate dielectric and the possible presence of interface states [36].

A SET junction can be viewed as two (metal) conductors, separated by a thin layer of insulating material (see [37] [38] [39] for overview papers). The operating principle of SET circuits is based on the controlled transport of charge through tunnel junctions. Specifically, SETs switch ON/OFF tunnel currents are conveying electrons that are being transported one by one from source to drain through a small island [12, 40]. SET junctions can be potentially utilized to build general purpose Boolean logic since they deliver high device density and power efficiency at good speed [41] [42]. Hajjam et al. [43] reported I_{ON}/I_{OFF} ratios of only about 10^2 with minimum I_{ON} of about 2nA. However, before their potential use in commercial circuits, the issues of the large V_{TH} variation and the low current drivability should be addressed. Moreover, the required significant circuit and architecture changes constitute an additional overhead for the industry [12].

Tunnel FETs (TFETs) are gated reverse-biased p-i-n junctions that are expected to potentially have low standby leakage current, and OFF-ON switching transitions much more abrupt than conventional MOSFETs, whose 60-mV/decade SS limit is set by the thermal injection of carriers from the source to the channel [44–46]. Guo et al. [47] reported I_{ON}/I_{OFF} ratios of 5×10^4 with ultra-low I_{OFF} below 1nA. However, key challenges for achieving an experimental TFET device include engineering the source tunneling region (junction abruptness, band-gap, carrier effective mass) and enhancement of gate control on the internal electric field. Moreover, for the design of TFET based integrated circuits the development of TFETs compact models are needed [12].

NEM-Switches essentially are electrostatic switches having three terminals. They rely on a nano-size mechanical beam to electrically open or close a con-

tact, thus to connect or disconnect a circuit path [48]. NEM-Relays feature the following desirable properties for logic computation which lack in MOSFETs: (i) abrupt switching due to the electromechanical instability at a certain threshold voltage (ideally SS is 0mV/decade), (ii) "zero" I_{OFF} due to the air gap that separates the source and drain electrodes in the *OFF* state, (iii) hysteresis, and (iv) stiction induced by the surface forces [49] [27]. The former two features enable a high I_{ON}/I_{OFF} ratio thus NEM-Relays potential to reduce both dynamic and static energy consumption, while the latter two features make them good candidates for (non)volatile memory applications [50, 51]. However, the NEM-Relays *ON* state resistance is increasing from $k\Omega$ to tens of $k\Omega$ after only 10^4 actuation, which makes them applicable only to the implementation of extremely low activity circuits. Moreover they require large control voltage which negatively impact the dynamic energy consumption.

NEMFETs are basically FETs with the surface potential at the oxide-semiconductor interface being controlled by a mechanically moving beam, i.e., a suspended gate [52]. Hence the current flow through the formed FET channel, not through the beam as is the case for NEM-Relays. While NEMFETs have been initially meant to be utilized in sensing [53], it has been suggested that they could be viable candidates for Sleep Transistor implementation due to their extremely low I_{OFF} [54, 55]. However, their main weakness is the large *ON* \leftrightarrow *OFF* switching delay, which cannot be reduced under 1ns due to the required beam movement [12].

To conclude, none of the above emerging devices can completely replace MOSFETs. If we focus on the I_{ON}/I_{OFF} ratio, the best in bread are NEM-Relays. However, given their reliability issues and the required high voltage operation, they are not fitted for portable, battery-powered, autonomous systems implementation. The next best candidate, which is also potentially co-integrable with CMOS, is NEMFET with an I_{ON}/I_{OFF} ratio larger than 10^6 (3 order of magnitude larger than MOSFET). In view of this and given that NEMFETs: (i) do not require high voltage operation and (ii) have other interesting properties (e.g., abrupt switching and hysteresis) that one can take advantage in building slow but ultra low power logic, we decided to pursue our investigations on the hybrid CMOS-NEMFET design avenue.

1.2 Research Questions

In view of the above discussion, we propose and asses in this thesis the potential practical impact and feasibility of hybrid CMOS-NEMFET circuits and

systems. The main goal is to demonstrate that CMOS-NEMFET device synergy enabled by 3D stacked integration can deliver substantial energy savings for power managed ICs as well as foster alternative implementations of Boolean functions and Static-RAM (SRAM) memory arrays. In view of this we address a number of research questions as follows:

What is the potential Nano-Electro-Mechanical Field Effect Transistor (NEMFET) device level benefit when compared with the "classical" MOSFET?

Starting with a conventional power gating technique, to reduce the idle energy consumption, the possibilities to replace High- V_{TH} MOSFETs with NEMFETs have to be identified. Due to NEMFET abrupt switching and ultra low I_{OFF} , it appears that it has the potential to replace traditional FETs in sleep mode circuits. However, the performance, e.g., switching delay, I_{ON} capability, and energy consumption, of different NEMFET geometries might be different. Thus, we have to investigate which NEMFET's geometry is the more appropriate for potential utilization as Sleep Transistor (ST) in energy efficient electronic products, especially portable, battery-powered, and autonomous ones.

To which extend NEMFET-based power gating can take advantage of the emerging 3D-Stacked integration?

Although fabrication feasibility of simple 2D hybrid CMOS-NEMS circuits has been reported [56], further technological enhancements are needed for such an approach to become a potentially viable industrial solution. Therefore, relocating the NEMFET STs on a different tier within a 3D-Stacked integrated structure could simplify the fabrication process complexity. In this context, it is of interest to evaluate to which extent is the STs relocation affecting the IC footprint when compared with the 2D counterpart. Moreover, a performance comparison between 2D and 3D stacked implementations of a generic execution unit, e.g., a real life processor, equipped with a power gating mechanism is of interest to determine the potential benefits the 3D-Stacked technology may provide.

Can we build NEMFET based Boolean gates? How do they compare with CMOS counterparts?

Given that both nNEMFETs and pNEMFETs can be realized we can start with a conventional MOSFET based logic family and replace n/p-FETs with n/p NEMFETs, respectively. However it is of interest to investigate to which extent we can take advantage of the NEMFET specific behavior, i.e., abrupt switching

and hysteresis, to obtain a slow but ultra-low power logic family?

Can NEMFETs be utilised in conjunction with FETs in the implementation of energy effective SRAM memory cells?

After proving the NEMFET logic benefits, the next obvious step is to design an SRAM cell in NEMFET technology. In this context we want to investigate low cost energy effective hybrid memory cells that can take advantage of the NEMFET based inverter properties. Moreover, 3D stacked memory cell structure and its potential low energy efficiency should be considered to combine the appealing NEMFET properties, with the CMOS technology versatility. Last but not least it is of interest to evaluate the potential performance of such hybrid memory arrays for different capacities and memory utilisation scenarios, as applications may have read or write dominated memory access traces.

To which extent can hybrid NEMFET-CMOS 3D stacked power managed computation platforms diminish the energy consumption of low activity embedded applications?

To evaluate the practical implications of the proposed 3D stacked hybrid NEMFET-CMOS power managed computation platform, we should assess the potential energy savings it enables when a real life application is executing on it. Hence, a 3D embodiment of an embedded low complexity 16-bit processor based SoC platform running, for example, a bio-medical sensing application for heart rate detection should be considered. Furthermore, it is of interest to acknowledge the effects of the 3D hybrid architecture on sensitive metrics used in power gating designs, e.g., delay degradation, power-up and power-down behavior, and overall energy consumption. Last but not least, we should investigate the energy consumption of the always-on circuitry, e.g., isolation (ISO) cells and the Power Management (PM) controller, and evaluate the performance of their redesign with NEMFET based Boolean gates.

Can hybrid CMOS-NEMFET 3D stacked platforms enable "zero-power" operated applications?

To investigate the capability of a hybrid CMOS-NEMFET 3D stacked platform to operate according to the "zero-power" paradigm, e.g., perform its function until it becomes obsolete without power supply and/or battery change, we rely on a full NEMFET power management approach and evaluate the feasibility of a multi tier stack that: (i) conserves the 3D IC footprint, and (ii) includes

energy scavenging tire(s). Given that several types of scavengers have emerged lately, a thorough analysis of their performance at circuit level is of interest. Based on all of the above investigations, conclusions and guidance should be formulated in order to help defining the most energy efficient hybrid NEMS-CMOS 3D embedded platform.

1.3 Dissertation Contributions

In this section, we highlight the main contributions of the research work described in this dissertation, as follows:

- We present a preliminary assessment of the NEMFET potential if utilized as Sleep Transistor (ST) in real life circuits, e.g., microprocessors. We first evaluate various NEMFET instances in terms of switching delay, current capability, and leakage. Subsequently, we compare these figures with the ones of traditional switch transistors utilized in CMOS technologies. According to our simulation results, NEMFET based sleep transistors enable substantial leakage reductions due to their extremely low *OFF* currents (4 orders of magnitude lower than FET) at the expense of a $4\times$ larger active area for the same I_{ON} capability. Finally, we evaluate the potential implications of the utilization of NEMFETs as sleep transistors in a 90 nm CMOS technology 32-bit Adder. Our simulations indicate that the leakage is mitigated, while the active area of the sleep transistor is increasing with 130%.
- We introduce a novel power management architecture which relies on the synergy of two new technological developments as follows: (i) Nano-Electro-Mechanical (NEM) devices, i.e., the NEM Field Effect Transistor (NEMFET), as sleep transistors, and (ii) 3D stacking, which allows for placing the sleep transistors (the entire power management infrastructure) on a dedicated tier of the 3D-stacked hybrid platform. As a test case, we consider the 3D embodiment of an embedded *openMSP430* processor based SoC platform running a bio-medical sensing application for heart rate detection and measure the 3D hybrid architecture consequences on sensitive metrics used in power gating designs, e.g., delay degradation, power-up and power-down behavior, and overall energy consumption. Our experiments indicate that, the system idle energy is decreased by $2.74\times$ for the same footprint, as the STs $4\times$ area

overhead is relocated on the NEMS tier. The energy-delay product of the *openMSP430* processor based SoC executing the heart rate detection application is reduced by 9%, with a potential reduction of up to 60% for applications with lower activity, e.g., wireless sensor networks. Last but not least the 3D stacked architecture prevents clock period degradation phenomena, since the IR Drop is reduced with a factor of 4 when compared with the 2D embodiment.

- We introduce a Short-Circuit-Current Free (SCCF) NEMFET based logic family tailored to the implementation of low speed and ultra low energy functional units and processors. We analyse and compare basic Boolean gates implemented with NEMFETs against equivalent CMOS realisations. Our simulations suggest that the proposed SCCF NEMFET gates are between 10 to 20× slower, but provide up to 10× dynamic energy reduction and up to 2 orders of magnitude less leakage, when compared with CMOS counterparts. We also analyse the fan-in influence on gate performance and observe that the NEMFET gates energy advantage increases with fan-in.
- We introduce a dual port 3D stacked hybrid SRAM memory that combines the ultra-low power NEMFET capabilities with the MOSFET technology versatility. The proposed memory relies on NEMFET based SCCF inverters to store data, and on adjacent CMOS based logic to allow for read and write operations, and data preservation. By utilising only one inverter per memory cell, instead of a cross coupled pair, a low write energy is achieved, as only one bitline is required. Furthermore, the static energy is drastically reduced due to NEMFET's extremely low I_{OFF} . The proposed dual port 3D NEMFET-CMOS hybrid memory relies on a memory cell with 140% and 30% footprint increase, for 2-die and 3-die implementations, respectively. However, by placing the memory array column and row circuitry within the memory cells, the total footprint of an 8-KB memory increases with only 60% for a 2-die implementation and decreases with 20% for a 3-die implementation. The access time, when compared with a state of the art CMOS based dual port memory, is equivalent for read operations, while for write operations it is approximately 4× higher, as it is dominated by the mechanical movement of the NEMFET's suspended gate. However we propose a solution to achieve a similar write access time by adding extra write ports, which, due to the available CMOS tier real estate, has no negative impact on footprint and read access time. We also compared the energy

consumption of standard and hybrid memory arrays as follows: (i) for small memory sizes of 2-KB and 8-KB our proposal results in about 10% and 30% write energy reduction, and a read energy reduction of 10% and 13%, respectively, and (ii) for large memory sizes, e.g., 128-KB, we obtain an energy reduction of 58%, regardless of the access type, as in this case the static energy is predominant. We have further considered different memory utilisation scenarios for an 8-KB memory, case in which our proposal results in up to 22% and up to 35% energy reduction for read and write dominated memory access traces, respectively.

- We propose and evaluate the "zero-power" operation paradigm potential of an improved version of the 3D-Stacked NEMS, i.e., NEMFET and NEM-Relay, based power management architecture when executing a heart beat detection application. The platform builds upon the following three embodiments: (i) an *openMSP430* ultra low power processing core appropriate for wireless sensor nodes, (ii) NEMS devices for the implementation of Sleep Transistors and of the additional power management low frequency circuitry necessary for power gating, and (iii) energy harvesters to provide enough energy for the processing core when executing a low duty cycle application. Our investigations indicate that the hybrid NEMFET-oriented approach, which relies on sleep transistors and associated management logic implemented on a dedicated NEMFET die, is the most promising in terms of energy consumption and reliability. Moreover, when combined with a thermal energy harvester, of 0.23 cm^2 potentially implementable on the same die, it can enable the road towards energy autonomous computing.

1.4 Dissertation Organization

This dissertation is organised as a selection of papers as follows:

In Chapter 2 we introduce NEMFET background and in Chapter 3 we assess NEMFET potential to be utilized as sleep transistor in real life circuits, e.g., microprocessors.

In Chapter 4 we propose a novel power management architecture that relies on CMOS and NEMFET synergetic utilization within the framework of 3D Through Silicon Vias based integration.

In Chapter 5 we describe and evaluate a 3D hybrid power management architecture which makes use of NEMFETs as power switches that cut-off the

power supply of inactive blocks. Furthermore, the relocation of isolation cells, and components for power management controller design, on the NEMFET tier is also investigated.

In Chapter 6, we introduce and analyze a NEMFET based logic family tailored to the implementation of low speed and ultra low energy functional units and processors. We also analyse the fan-in influence on gate performance and observe that the NEMFET gate energy advantage is increasing with its fan-in. Finally, we consider a 3D-Stacked hybrid NEMFET-CMOS computation platform running a heartbeat rate monitor application and demonstrate that NEMFET based logic is an enabling factor for the implementation of "zero-energy" operated systems.

In Chapter 7, we propose and evaluate a dual port 3D stacked hybrid memory that combines NEMFET abrupt switching and hysteresis with CMOS technology versatility.

In Chapter 8, we investigate the capability of a hybrid CMOS-NEM 3D staked platform to operate according to the "zero-power" paradigm, e.g., perform its function until it becomes obsolete without power supply and/or battery change. We evaluate 3D platforms equipped with NEMFETs and NEM-Relays based power management in combination with various efficient energy harvesters while having in mind the tight energy budgets of "zero-power" operating autonomous sensor systems.

Finally, Chapter 9 wraps up the dissertation, by presenting our conclusions and indicating significant and promising follow-up research directions.

Bibliography

- [1] Z. Horst, “Part 4:Konrad Zuse’s Z1 and Z3 Computers,” *The Life and Work of Konrad Zuse*.
- [2] “Computer history museum,” 2015. [Online]. Available: <http://www.computerhistory.org/>
- [3] “Jack st clair kilby biography by texas instruments.” [Online]. Available: <http://www.ti.com/corp/docs/kilbyctr/jackstclair.shtml>
- [4] “Intel processors,” 2014. [Online]. Available: <http://ark.intel.com/>
- [5] G. E. Moore, “Cramming more components onto integrated circuits,” *Electronics*, vol. 38, no. 8, April 1965.
- [6] T. Kuroda, “Low-power, high-speed cmos vlsi design,” in *Computer Design: VLSI in Computers and Processors, 2002. Proceedings. 2002 IEEE International Conference on*, 2002, pp. 310 – 315.
- [7] K. Asanovic, R. Bodik, B. C. Catanzaro, J. J. Gebis, K. Keutzer, D. A. Patterson, W. L. Plishker, J. Shalf, S. W. Williams, K. A. Yelick, M. J. Demmel, W. Plishker, J. Shalf, S. Williams, and K. Yelick, “The landscape of parallel computing research: A view from Berkeley,” TECHNICAL REPORT, UC BERKELEY, Tech. Rep., 2006.
- [8] “Arm cortex m+,” 2012. [Online]. Available: <http://low-powerdesign.com/sleibson/2012/03/16/how-low-can-you-go-arm-does-the-limbo-with-cortex-m0-processor-core-tiny-ultra-low-power>
- [9] V. Sharma, S. Cosemans, M. Ashouei, J. Huisken, F. Catthoor, and W. Dehaene, “Ultra low energy SRAM design for smart ubiquitous sensors,” *IEEE Micro*, pp. 1–1, 2012.
- [10] J. Hu, W. Liu, and M. Ismail, “Sleep-mode ready, area efficient capacitor-free low-dropout regulator with input current-differencing,” *Analog Integrated Circuits and Signal Processing*, vol. 63, no. 1, pp. 107–112, 2010.
- [11] R. Gonzalez, B. Gordon, and M. Horowitz, “Supply and threshold voltage scaling for low power cmos,” *Solid-State Circuits, IEEE Journal of*, vol. 32, no. 8, pp. 1210–1216, Aug 1997.
- [12] ITRS, “Emerging Research Devices,” 2009. [Online]. Available: <http://www.itrs.net/>

- [13] F. Fallah and M. Pedram, "Standby and active leakage current control and minimization in cmos vlsi circuits," *IEICE Trans. Electron. (Special Section on Low-Power LSI and Low-Power IP)*, vol. E88-C, no. 4, pp. 509–519, 2005.
- [14] T. Burd, T. Pering, A. Stratakos, and R. Brodersen, "A dynamic voltage scaled microprocessor system," in *Solid-State Circuits Conference, 2000. Digest of Technical Papers. ISSCC. 2000 IEEE International*, Feb 2000, pp. 294–295.
- [15] H. Singh, K. Agarwal, D. Sylvester, and K. Nowka, "Enhanced leakage reduction techniques using intermediate strength power gating," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 15, no. 11, pp. 1215–1224, Nov 2007.
- [16] J. Tschanz, S. Narendra, Y. Ye, B. Bloechel, S. Borkar, and V. De, "Dynamic sleep transistor and body bias for active leakage power control of microprocessors," *Solid-State Circuits, IEEE Journal of*, vol. 38, no. 11, pp. 1838–1845, nov. 2003.
- [17] K. von Arnim, P. Seegebrecht, R. Thewes, and C. Pacha, "A low-leakage 2.5ghz skewed cmos 32b adder for nanometer cmos technologies," in *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, feb. 2005, pp. 380–605 Vol. 1.
- [18] Y. Ye, S. Borkar, and V. De, "A new technique for standby leakage reduction in high-performance circuits," in *VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on*, jun 1998, pp. 40–41.
- [19] J. Kao and A. Chandrakasan, "Dual-threshold voltage techniques for low-power digital circuits," *Solid-State Circuits, IEEE Journal of*, vol. 35, no. 7, pp. 1009–1018, July 2000.
- [20] *Low Power Methodology Manual*. Boston, MA: Springer US, 2007. [Online]. Available: <http://www.springerlink.com/index/10.1007/978-0-387-71819-4>
- [21] M. Anis, S. Areibi, M. Mahmoud, and M. Elmasry, "Dynamic and leakage power reduction in mtcmos circuits using an automated efficient gate clustering technique," in *Design Automation Conference, 2002. Proceedings. 39th*, 2002, pp. 480–485.

- [22] S. Mukhopadhyay, C. Neau, R. Cakici, A. Agarwal, C. Kim, and K. Roy, "Gate leakage reduction for scaled devices using transistor stacking," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 11, no. 4, pp. 716–730, Aug 2003.
- [23] Y. Ye, S. Borkar, and V. De, "A new technique for standby leakage reduction in high-performance circuits," in *VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on*, June 1998, pp. 40–41.
- [24] N. H. E. Weste and D. M. Harris, *CMOS VLSI Design A Circuits and Systems Perspective*. Addison Wesley, 2011.
- [25] H. Iwai, "Roadmap for 22nm and beyond," *Microelectronic Engineering*, vol. 86, no. 7-9, pp. 1520–1528, Jul. 2009.
- [26] J. Hu and M. Ismail, *CMOS High Efficiency On-chip Power Management*. New York: Springer, 2011.
- [27] N. Rhesa, "Phd dissertation: Nano-electro-mechanical (nem) relay devices and technology for ultra-low energy digital integrated circuits," Ph.D. dissertation, EECS Department, University of California, Berkeley, 2012. [Online]. Available: <http://http://www.eecs.berkeley.edu/~tking/theses/rhesa.pdf>
- [28] "International technology roadmap for semiconductors," ITRS, Tech. Rep., 2011. [Online]. Available: <http://www.itrs.net/>
- [29] R. Nathanael, "Nano-electro-mechanical (NEM) relay devices and technology for ultra-low energy digital integrated circuits," Ph.D. dissertation, University of California, Berkeley, 2012. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5617293
- [30] A. Le Louarn, F. Kapche, J.-M. Bethoux, H. Happy, G. Dambrine, V. Derycke, P. Chenevier, N. Izard, M. F. Goffman, and J.-P. Bourgoin, "Intrinsic current gain cutoff frequency of 30ghz with carbon nanotube transistors," *Applied Physics Letters*, vol. 90, no. 23, pp. –, 2007.
- [31] A. Lin, N. Patil, K. Ryu, A. Badmaev, L. De Arco, C. Zhou, S. Mitra, and H.-S. Wong, "Threshold voltage and ratio tuning for multiple-tube carbon nanotube fets," *Nanotechnology, IEEE Transactions on*, vol. 8, no. 1, pp. 4–9, Jan 2009.

- [32] S. Fan, I. Appelbaum, and J. D. Joannopoulos, "Near-field scanning optical microscopy as a simultaneous probe of fields and band structure of photonic crystals: A computational study," *Applied Physics Letters*, vol. 75, no. 22, 1999.
- [33] Y. Saito, M. Ishikawa, T. Inokuchi, H. Sugiyama, T. Tanamoto, K. Hamaya, and N. Tezuka, "Spin-based mosfets for logic and memory applications and spin accumulation signals in coFe/tunnel barrier/soi devices," *Magnetics, IEEE Transactions on*, vol. 48, no. 11, pp. 2739–2745, Nov 2012.
- [34] D. Harame, *SiGe, Ge, and Related Compounds 3: Materials, Processing, and Devices*, ser. ECS transactions. Electrochemical Society, 2008, no. no. 10. [Online]. Available: <https://books.google.ro/books?id=BmvWuzENr7kC>
- [35] S. Matsunaga, J. Hayakawa, S. Ikeda, K. Miura, H. Hasegawa, T. Endoh, H. Ohno, and T. Hanyu, "Fabrication of a nonvolatile full adder based on logic-in-memory architecture using magnetic tunnel junctions," *Applied Physics Express*, vol. 1, no. 9, p. 091301, 2008.
- [36] V. Kamalakar, Madhushankar, A. Dankert, and S. P. Dash., *Engineering schottky barrier in black phosphorus field effect devices for spintronic applications*.
- [37] A. N. KOROTKOV, "Single-electron logic and memory devices," *International Journal of Electronics*, vol. 86, no. 5, pp. 511–547, 1999.
- [38] K. Likharev, "Single-electron devices and their applications," *Proceedings of the IEEE*, vol. 87, no. 4, pp. 606–632, Apr 1999.
- [39] C. Lageweg, S. Cotofana, and S. Vassiliadis, "A linear threshold gate implementation in single electron technology," in *VLSI, 2001. Proceedings. IEEE Computer Society Workshop on*, May 2001, pp. 93–98.
- [40] R. Waser, *Nanoelectronics and Information Technology*. Wiley, 2012. [Online]. Available: <https://books.google.nl/books?id=1PgYS7zDCM8C>
- [41] S. Cotofana, C. Lageweg, and S. Vassiliadis, "On computing addition related arithmetic operations via controlled transport of charge," in *Computer Arithmetic, 2003. Proceedings. 16th IEEE Symposium on*, June 2003, pp. 245–252.

- [42] C. Lageweg, S. Cotofana, and S. Vassiliadis, "Evaluation methodology for single electron encoded threshold logic gates," in *VLSI-SOC: From Systems to Chips*, ser. IFIP International Federation for Information Processing, M. Glesner, R. Reis, L. Indrusiak, V. Mooney, and H. Eveking, Eds. Springer US, 2006, vol. 200, pp. 247–262. [Online]. Available: http://dx.doi.org/10.1007/0-387-33403-3_16
- [43] K. El Hajjam, M. Bounouar, N. Baboux, S. Ecoffey, M. Guilmain, E. Puyoo, L. Francis, A. Souifi, D. Drouin, and F. Calmon, "Tunnel junction engineering for optimized metallic single-electron transistor," *Electron Devices, IEEE Transactions on*, vol. 62, no. 9, pp. 2998–3003, Sept 2015.
- [44] J. Quinn, G. Kawamoto, and B. McCombe, "Subband spectroscopy by surface channel tunneling," *Surface Science*, vol. 73, no. 0, pp. 190 – 196, 1978.
- [45] T. Baba, "Proposal for surface tunnel transistors," *Japanese Journal of Applied Physics*, vol. 31, no. 4B, p. L455, 1992.
- [46] Q. Zhang, W. Zhao, and A. Seabaugh, "Low-subthreshold-swing tunnel transistors," *Electron Device Letters, IEEE*, vol. 27, no. 4, pp. 297–300, April 2006.
- [47] A. Guo, P. Matheu, and T.-J. K. Liu, "Soi tfet i_{on}/i_{off} enhancement via back biasing," *Electron Devices, IEEE Transactions on*, vol. 58, no. 10, pp. 3283–3285, Oct 2011.
- [48] J. Hutchby, "Maturity evaluation for selected beyond cmos emerging technologies," in *Workshop and ERD/ERM Working Group Meeting*, July 2008, p. 3.
- [49] R. Nathanael, V. Pott, H. Kam, J. Jeon, and T. Liu, "4-terminal relay technology for complementary logic," in *IEEE International Electron Devices Meeting*, 2009, pp. 1–4.
- [50] H. Kam, "MOSFET replacement devices for energy-efficient digital integrated circuits," DTIC Document, Tech. Rep., 2009.
- [51] H. Kam, V. Pott, R. Nathanael, J. Jeon, E. Alon, and T. Liu, "Design and reliability of a micro-relay technology for zero-standby-power digital logic applications," in *IEEE International Electron Devices Meeting*, 2009, pp. 1–4.

-
- [52] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, "Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture," in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [53] X. M. H. Huang, M. Manolidis, S. C. Jun, and J. Hone, "Nanomechanical hydrogen sensing," *Applied Physics Letters*, vol. 86, no. 14, pp. 143 104–143 104–3, apr 2005.
- [54] H. F. Dadgour and K. Banerjee, "Design and analysis of hybrid NEMSCMOS circuits for ultra low-power applications," in *Proceedings of the 44th annual Design Automation Conference*, 2007, pp. 306–311.
- [55] D. Tsamados, Y. Chauhan, C. Eggimann, K. Akarvardar, H. Wong, and A. Ionescu, "Numerical and analytical simulations of suspended gate - fet for ultra-low power inverters," in *Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European*, sept. 2007, pp. 167–170.
- [56] H. Dadgour and K. Banerjee, "Hybrid NEMSCMOS integrated circuits: a novel strategy for energy-efficient designs," *IET Computers & Digital Techniques*, vol. 3, no. 6, p. 593, 2009.

2

Can NEMFET Replace FET in Sleep Mode Circuits?¹

Abstract: The Nano-Electro-Mechanical Field Effect Transistor (NEMFET) appears to have the potential to replace traditional FETs in sleep mode circuits, due to its abrupt switching enabled by electromechanical instability at a certain threshold voltage and its ultra low OFF current (I_{OFF}). This paper presents a preliminary assessment of the NEMFET potential if utilized as sleep transistor in real applications, e.g., microprocessors. We first evaluate various NEMFET instances in terms of switching delay, current capability, and leakage. Subsequently, we compare these figures with the ones offered by traditional switch transistors utilized in CMOS technologies. Our simulation results indicate that NEMFET based sleep mode circuits are potentially interesting as they clearly enable substantial leakage reductions due to their extremely low OFF currents (4 orders of magnitude lower than FET) at the expense of a $4\times$ larger active area for the same capability to drive current.

Copyright © 2009 Springer. Personal use of this material is permitted. However, permission to use this material for any other purpose must be obtained from Springer by sending an email to permissions.dordrecht@springer.com.

¹This chapter is based on the research article published as "Can SG-FET Replace FET In Sleep Mode Circuits?" by M. Enachescu, S. D. Cotofana, A. van Genderen, D. Tsamados, and A. M. Ionescu, *ACM 4th International ICST Conference on Nano-Networks (NANO-NET)*, pp. 99-104, Luzern, Switzerland, October, 2009.

2.1 Introduction

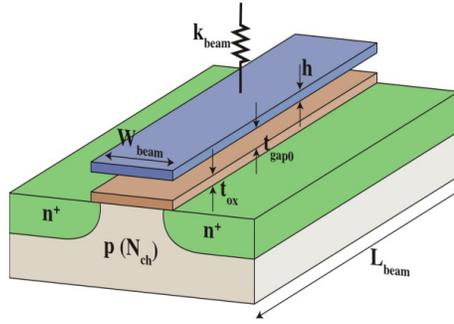
The Nano-Electro-Mechanical Field Effect Transistor (NEMFET) appears to have the potential to replace traditional FETs in sleep mode circuits, due to its abrupt switching enabled by electromechanical instability at a certain threshold voltage and its ultra low OFF current (I_{OFF}). The purpose of this paper is to assess the NEMFET potential if utilized as sleep transistor in real applications, e.g., micro (processors), and to find out if NEMFET constitutes a promising alternative to normal FET in sleep mode circuits. In this line of reasoning we need to evaluate the NEMFET performance in terms of switching delay, current capability, and leakage and compare those with the ones offered by traditional switch transistors utilized in up to date CMOS technologies. To achieve our goal we go through the following steps. We first perform a design space exploration in order to identify the most promising NEMFET geometries and to evaluate their potential performance. Subsequently, we compare the performance of an N-channel NEMFET with the one of an N-channel normal FET, having the same active area, in 90 nm CMOS technology.

This paper is organized as follows: in Section 2.2 brief introduction is provided on NEMFET including its basic operation and modeling. Section 2.3 describes the design space exploration for NEMFET model parameters. In Section 2.4 we compare nNEMFET with nFET by means of I_{ON} , I_{OFF} , and switching delay and finally concluding remarks are made in Section 3.5.

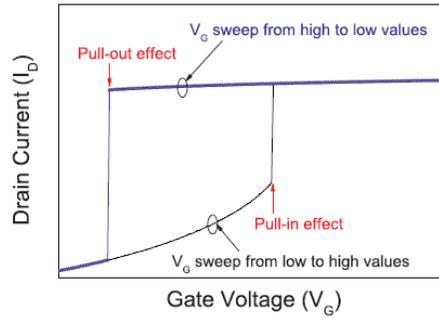
2.2 NEMFET Background

The NEMFET described in [1] and [2] is a rather complex device with a 3D geometry as presented in Figure 3.1(a), where: (i) t_{ox} - the thickness of the gate oxide, (ii) h - the thickness of the suspended gate, (iii) W_{beam} - the width of the beam, (iv) L_{beam} - the length of the beam, (v) t_{gap0} - the gap between the oxide and the suspended gate, (vi) k_{beam} - the lumped linear spring constant of the beam.

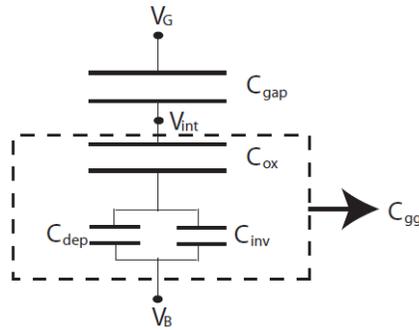
Figure 3.1(b) presents the typical I_D - V_G characteristics of NEMFET. As V_G starts increasing, the beam starts moving down due to electrostatic attraction and I_D increases. During this phase, the gate-oxide capacitance is in series with the air-gap capacitance resulting in low electrostatic coupling of the gate to the channel and I_D is very small. At a specific gate bias, the electrostatic force cannot be compensated by the mechanical restoring force anymore, and the beam collapses on the oxide. This is called pull-in effect as depicted in



(a) Basic 3D NEMFET Geometry



(b) $I_D - V_G$ characteristic



(c) NEMFET equivalent capacitive divider

Figure 2.1: (a) NEMFET geometry, (b) NEMFET transfer characteristic, and (c) NEMFET equivalent capacitive divider.

Figure 3.1(b). After pull-in, increase in I_D with V_G is similar to the standard MOSFET. If V_G is decreased from some high value, then I_D starts decreasing. At certain value of V_G , the system becomes unstable due to combined electro-

mechanical force and beam is pulled-out. This causes sudden decrease in I_D due to large decrease in capacitance (Figure 3.1(c)). This effect is called pull-out effect as indicated in Figure 3.1(b). NEMFET features a dynamic threshold voltage: (i) high in the up-state, and (ii) low in the down-state. This property is not always beneficial, especially in (micro)processors domain, where the supply voltage should be as low as possible.

2.3 Design Space Exploration

To carry on a thorough analysis of the NEMFET potential capabilities we need to generate a large set of feasible NEMFET geometries and to evaluate them by means of simulations. To characterize the various NEMFET device instances we utilize the NEMFET Verilog-A model introduced in [3] in combination with Cadence Spectre circuit simulator [4].

Given the complexity of the design space we have to restrict the dynamic range for the device parameters for (micro)processors. The supply voltage for processor applications in 90 nm technology is 1.1 V, according to the 2007 ITRS roadmap [5]. This low supply voltage, assuming the lithography constraints (minimum $W_{beam} = 350$ nm, minimum $t_{ox} = 3$ nm, minimum $t_{gap0} = 20$ nm) is not sufficient for such an NEMFET device to properly function. In view of that, we focused the current investigation on finding NEMFET geometries with a pull-in voltage of 3 V, which can be of interest for applications with two supply voltages (3.3 V and 1.1 V). To find the device that is best suited for the considered application, we investigate a wide range of geometrical shapes as follows: (i) we vary h from 70 nm to 100 nm with a step increment of 10 nm, and (ii) we vary t_{gap0} from 10 nm to 25 nm, with a step increment of 5 nm. Other parameters that influence the performance of NEMFET are the gate work-function (WF) and the quality factor (Q). Every vibrating structure is subject to some energy loss, which translates in a reduction of vibration amplitude over time. The long settling times associated with those large Q_s are however detrimental for rapidly switching devices such as the NEMFET [6]. We note here that in our preliminary study we only simulate one switching cycle (pulse), due to large amount of simulation data (many samples), thus the effect of the quality factor is not fully exposed. The gate work function mainly influences transistors characteristics by shifting them with respect to the applied gate bias [7]. In our experiments we assumed the following values:

- WF of 4.4eV, 4.6eV, 4.8eV, and 5eV, and

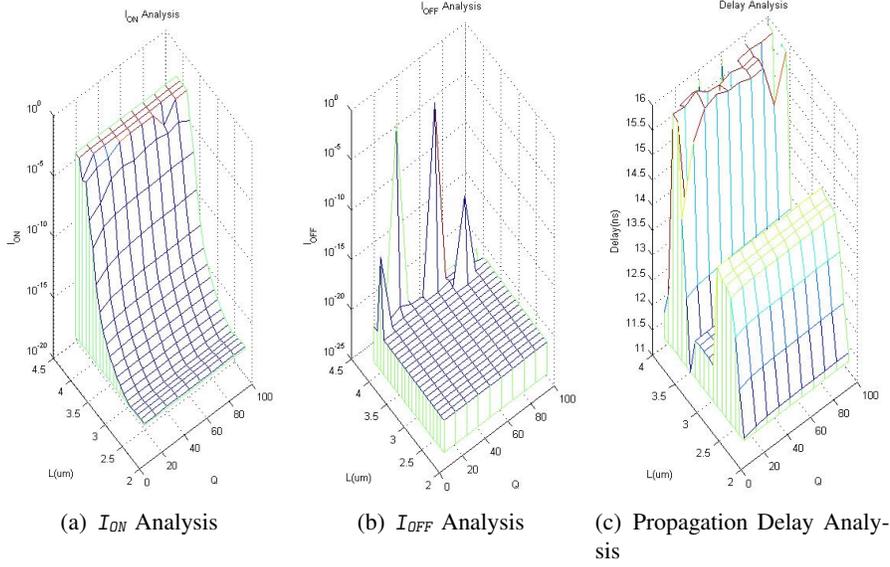


Figure 2.2: (a) I_{ON} Analysis, (b) I_{OFF} Analysis, and (c) Propagation Delay Analysis, for $WF = 5\text{eV}$, $h = 100\text{ nm}$, $\tau_{gap0} = 20\text{ nm}$.

- we varied Q from 10 to 100, with a step increment of 10.

The parameters of interest are determined as follows:

- I_{ON} is 90% of the maximum drain current produced as result of an input step signal,
- I_{OFF} is the drain current after the pull-out event, and
- The switching delay is the time required for the device to reach 50% of its maximum drain current, when the gate voltage is larger than the pull-in voltage (V_{PI})

Examples of the I_{ON} , I_{OFF} , and switching delay, we deduced via SPICE simulations are depicted in Figure 3.3 and Figure 3.3.

The results of our simulations suggest the following:

- switching delay $\propto \tau_{gap0}$, h , $1/L_{BEAM}$ (area)
- $I_{ON} \propto L_{BEAM}$, $1/WF$, $1/h$, $1/\tau_{gap0}$, and

- $I_{OFF} \propto L_{BEAM}$

Moreover we observe that while there are clear relations between the various device parameters and the NEMFET performance there is no absolute best in breed geometry and various tradeoffs are possible.

Table 3.1 presents three NEMFET configurations that we deduced from our extensive simulations results. The first set of parameters was selected as optimal for low switching time and high I_{ON} , with respect to pull-in and pull-out effects, when (i) $WF = 5eV$, (ii) $t_{gap0} = 20$ nm, 25 nm and (iii) $h = 70$ nm, 80 nm, 90 nm, 100 nm. The second set of parameters was selected as optimal for low switching time and high I_{ON} , with respect to pull-in and pull-out effects, when $WF = 4.4eV, 4.6eV, 4.8eV, 5eV$, selected as optimal for low switching time and high I_{ON} , with respect to pull-in and pull-out effects, when $WF = 4.4eV, 4.6eV, 4.8eV, 5eV, t_{gap0} = 10$ nm, 15 nm, 20 nm, 25 nm, and $h = 70$ nm, 80 nm, 90 nm, 100 nm.

Table 2.1: Optimized NEMFET instances for low switching times and high I_{ON} .

	h (nm)	t_{gap0} (nm)	WF (eV)	L (nm)	W (nm)	Q	I_{ON} (mA)	I_{OFF} (Leakage floor)(fA)	Delay (ns)
I	100	20	5	4.2	350	100	2.2	5e-7(12)	8.67
II	100	20	4.4	4	350	100	3	2e-4(4.5e3)	7.41
III	100	10	5	4.5	350	100	1E-06	5e-7(5e-4)	3.03

Table 3.1 indicates that the best compromise between high I_{ON} and low delay, with respect to pull-in and pull-out effects, for $t_{gap0} = 20$ nm, 25 nm and $WF = 5eV$, can be reached for the set of NEMFET parameters Table 3.1(I).

2.4 NEMFET vs. FET

In this section we compare the performance of an N-channel NEMFET with the one of an N-channel FET in 90 nm CMOS technology, assuming the same active area. To do that we utilize the best performance NEMFET instance still lithographically feasible having the parameters in Table 3.1(I): $W_{BEAM}=350$ nm, $L_{BEAM}= 4.2 \mu\text{m}$, $h=100$ nm, $t_{gap0}=20$ nm, $WF=5eV$, $t_{ox}=3$ nm, $N_A=5 \times 10^{17} \text{ cm}^{-3}$.

For a fair comparison we assume as counterpart an N-FET with the width equal to L_{BEAM} of the NEMFET and the length equal to W_{BEAM} , in order to have the same active area for both transistors. The results of our simulations are

presented in Table 2.2, which includes the key performance data for the normal nFET transistors and the nNEMFET devices for $V_D=1.2$ V and $V_G=3$ V.

It is clear from Table 2.2 that the main NEMFET advantage is its extremely small I_{OFF} , and leakage floor, which are 10 and 4, orders of magnitude smaller, respectively, while I_{ON} even though smaller it is comparable with the I_{ON} of normal FET. The NEMFET however is about $100\times$ slower than the normal FET and the active area is $4\times$ larger for the same capability to drive current.

Table 2.2: Optimized NEMFET instances.

	I_{ON} (mA/ μ m)	I_{OFF} (pA/ μ m)	Leakage Floor (pA/ μ m)	Delay (ps)
nFET	0.8	13	200	20
nNEMFET	0.52	2.5E-10	1.2E-02	8670

To conclude, Table 2.2 suggests that NEMFET is a viable alternative to FET as sleep transistor due to its extremely low I_{OFF} and leakage floor. However, due to its relatively large switching delay, this device appears not to be suited for applications where the switching between active mode and sleep mode occurs too often. Fortunately, for processors, this is not the case in practice. For example, as indicated in [8], the wake-up time for a mobile application is about 2μ s.

2.5 Conclusions

In this paper we presented the results of the preliminary evaluation we carried on to estimate the NEMFET potential if utilized as sleep transistor in (micro)processors, to find out if NEMFET constitutes a promising alternative to normal FET. For this we evaluated various NEMFET geometries in terms of switching delay, current capability, and leakage and compared those with the ones offered by traditional switch transistors utilized in up to date CMOS technology. Our results indicate that NEMFETs can be potentially used as sleep transistors, due to their very low leakage floor and I_{OFF} , which are with 4 and 10 orders of magnitude smaller than the one of the normal FETs, respectively. However, for the current fabrication technology limitations, we could not obtain pull-in effects for gate voltages smaller than 3 V and this implies some design overhead due to the utilization of an additional power supply. Moreover, the NEMFET requires a larger area when compared with a "normal" FET for the same capability to drive current.

Bibliography

- [1] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, “Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture,” in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [2] B. Pruvost, K. Uchida, H. Mizuta, and S. Oda, “Design optimization of NEMS switches for suspended-gate single-electron transistor applications,” *IEEE Transactions on Nanotechnology*, vol. 8, no. 2, pp. 174–184, Mar. 2009.
- [3] M. Enachescu, A. van Genderen, S. D. Cotofana, D. Tsamados, and A. Ionescu, “Benchmark suite and specifications for nem switches for power management,” NEMSIC Deliverable 2.1, 2009. [Online]. Available: <http://www.nemsic.org/members/pfn.v2/documents.html>
- [4] “Spectre, Cadence Design Systems,” 2009. [Online]. Available: <http://www.cadence.com/us/pages/default.aspx>
- [5] “International technology roadmap for semiconductors,” ITRS, Design, 2007. [Online]. Available: <http://www.itrs.net/>
- [6] Gabriel M. Rebeiz and Jeremy B. Muldavin, *RF MEMS*. John Wiley & Sons, Inc., 2004, no. 9780471225287.
- [7] D. Tsamados, Y. Singh Chauhan, C. Eggimann, K. Akarvardar, H. S. Philip Wong, and A. Mihai Ionescu, “Finite element analysis and analytical simulations of suspended gate-FET for ultra-low power inverters,” *Solid-State Electronics*, vol. 52, no. 9, p. 13741381, 2008.
- [8] K. Fukuoka, O. Ozawa, R. Mori, Y. Igarashi, T. Sasaki, T. Kuraishi, Y. Yasu, and K. Ishibashi, “A 1.92us wake-up time thick-gate-oxide power switch technique for ultra low-power single-chip mobile processors,” in *VLSI Circuits, 2007 IEEE Symposium on*, june 2007, pp. 128 –129.

3

Nano-Electro-Mechanical Field Effect Transistor Based Power Management - A 32-Bit Adder Case Study¹

Abstract: Recent investigations suggest that the Nano-Electro-Mechanical Field Effect Transistor (NEMFET) appears to have the potential to replace traditional high-V_t FETs, utilized as sleep transistors in power management circuits, due to its abrupt switching enabled by electromechanical instability at a certain threshold voltage and its ultra low OFF current (I_{OFF}). This paper presents a preliminary assessment of the NEMFET potential if utilized as sleep transistor in circuits featuring cell based power gating. We first evaluate various NEMFET instances in terms of switching delay, current capability, and leakage. Subsequently, we compare these figures with the once offered by traditional switch transistors utilized in CMOS technologies. Finally, we evaluate the potential implications of the utilization of NEMFETs as sleep transistors in a 90 nm CMOS technology 32-bit Adder. Our simulations indicate that I_{OFF} is reducing by 10 orders of magnitude, while the active area of the sleep transistor is increasing with 130%.

Copyright © 2009 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purpose must be obtained from IEEE by sending an email to *pubs – permissions@ieee.org*.

¹This chapter is based on the research article published as "Suspended Gate Field Effect Transistor based Power Management - A 32-bit Adder Case Study" by M. Enachescu, A. van Genderen, and S. D. Cotofana, *IEEE 32nd International Semiconductor Conference (CAS)*, pp. 561-564, Sinaia, Romania, October, 2009.

3.1 Introduction

According to International Technology Roadmap for Semiconductors (ITRS) [1], leakage in nanometer CMOS technologies can exceed active power consumption for VLSI circuits.

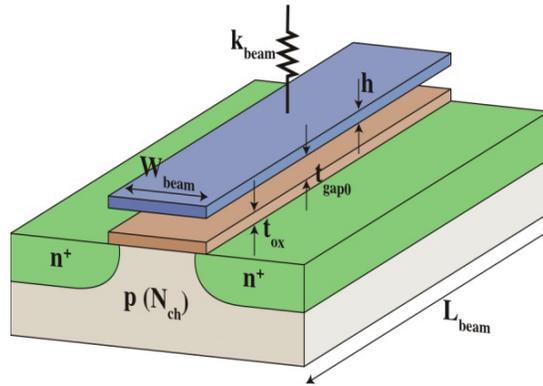
The Nano-Electro-Mechanical Field Effect Transistor (NEMFET) appears to have the potential to replace traditional FETs in sleep mode circuits, due to: (i) its abrupt switching enabled by electromechanical instability at a certain threshold voltage that can lower the dynamic leakage, and (ii) its ultra low OFF current (I_{OFF}) that can lower the static leakage [2].

The purpose of this paper is twofold: (i) asses the NEMFET potential when utilized as sleep transistor in circuits featuring cell based power gating, and (ii) find out if NEMFETs constitute a viable alternative to high-Vt FETs in sleep mode circuits. In this line of reasoning we need to evaluate the NEMFET performance in terms of switching delay, current capability, and leakage and compare those with the ones offered by traditional switch transistors utilized in up to date CMOS technologies.

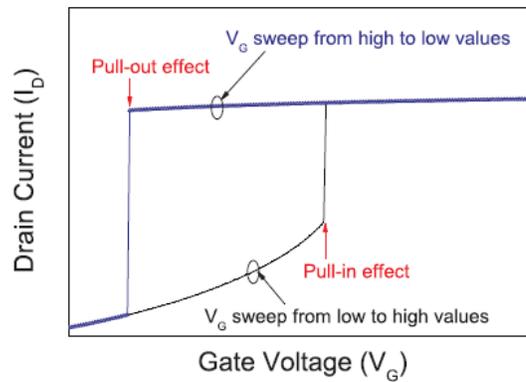
To achieve our goal we go through the following steps. We first focus on NEMFET background and its advantages. Subsequently, we perform a design space exploration in order to identify the most promising NEMFET geometries and to evaluate their potential performance. Finally, we compare the performance of an N-channel NEMFET with the one of an N-channel high-Vt FET in 90 nm CMOS technology having the same active area. This paper is organized as follows: in Section 3.2 a brief introduction is provided on NEMFET including its basic operation and modeling. Section 3.3 describes the NEMFET design space exploration process. In Section 3.4 we evaluate the potential implications of the utilization of NEMFETs as sleep transistors in a 90 nm CMOS 32-bit Adder [3] and finally concluding remarks are made in Section 3.5.

3.2 NEMFET Background

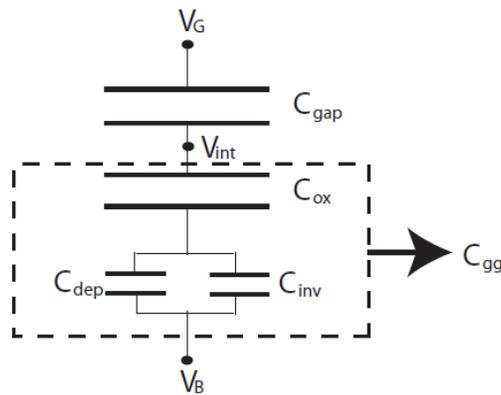
The NEMFET described in [4] and [5] is a rather complex device with a 3D geometry as presented in Figure 3.1(a), where: (i) t_{ox} - the thickness of the gate oxide, (ii) h - the thickness of the suspended gate, (iii) W_{beam} - the width of the beam, (iv) L_{beam} - the length of the beam, (v) t_{gap0} - the gap between the oxide and the suspended gate, (vi) k_{beam} - the lumped linear spring constant of the beam.



(a) Basic 3D NEMFET Geometry



(b) $I_D - V_G$ characteristic



(c) NEMFET equivalent capacitive divider. C_{gap} , C_{ox} , C_{dep} , C_{inv} , and C_{gg} represent gap, oxide, depletion, inversion, and gate-to-gate capacitances, respectively.

Figure 3.1: NEMFET (a) geometry, (b) transfer characteristic, and (c) equivalent capacitive divider.

Figure 3.1(b) presents the typical I_D - V_G characteristics of NEMFET. As V_G starts increasing, the beam starts moving down due to electrostatic attraction and I_D increases. During this phase, the gate-oxide capacitance is in series with the air-gap capacitance resulting in low electrostatic coupling of the gate to the channel and I_D is very small (Figure 3.1(c)). At a specific gate bias, the electrostatic force cannot be compensated by the mechanical restoring force anymore, and the beam collapses on the oxide. This is called pull-in effect as depicted in Figure 3.1(b). After pull-in, increase in I_D with V_G is similar to the standard MOSFET. If V_G is decreased from some high value, then I_D starts decreasing. At certain value of V_G , the system becomes unstable due to combined electro-mechanical force and beam is pulled-out. This causes sudden decrease in I_D due to large decrease in capacitance (Figure 3.1(c)). This effect is called pull-out effect as indicated in Figure 3.1(b).

NEMFET features a dynamic threshold voltage: (i) high in the up-state, and (ii) low in the down-state. This property is not always beneficial, especially in (micro)processors domain, where the supply voltage should be as low as possible.

3.3 Design Space Exploration

To carry on a thorough analysis of the NEMFET potential capabilities we need to generate a large set of feasible NEMFET geometries and to evaluate them by means of simulations. To characterize the various NEMFET device instances we utilize the NEMFET Verilog-A model introduced in [6] in combination with Cadence Spectre circuit simulator [7].

Given the complexity of the design space we have to restrict the dynamic range for the device parameters for (micro)processors. The supply voltage for processor applications in 90 nm technology is 1.1 V, according to the 2007 ITRS roadmap [1]. This low supply voltage, assuming the lithography constraints (minimum $W_{beam} = 350$ nm, minimum $t_{ox} = 3$ nm, minimum $t_{gap0} = 20$ nm) is not sufficient for such an NEMFET device to properly function. In view of that, we focused the current investigation on finding NEMFET geometries with a pull-in voltage of 3 V, which can be of interest for applications with two supply voltages (3.3 V and 1.1 V). To find the device that is best suited for the considered application, we investigate a wide range of geometrical shapes as follows: (i) we vary h from 70 nm to 100 nm with a step increment of 10 nm, and (ii) we vary t_{gap0} from 10 nm to 25 nm, with a step increment of 5 nm. Other parameters that influence the performance of NEMFET are the

gate work-function (WF) and the quality factor (Q). Every vibrating structure is subject to some energy loss, which translates in a reduction of vibration amplitude over time. The long settling times associated with those large Q_s are however detrimental for rapidly switching devices such as the NEMFET [8]. We note here that in our preliminary study we only simulate one switching cycle (pulse), due to large amount of simulation data (many samples), thus the effect of the quality factor is not fully exposed. The gate work function mainly influences transistors characteristics by shifting them with respect to the applied gate bias [2]. In our experiments we assumed the following values:

- WF of 4.4eV, 4.6eV, 4.8eV, and 5eV, and
- we varied Q from 10 to 100, with a step increment of 10.

The parameters of interest are determined as follows:

- I_{ON} is 90% of the maximum drain current produced as result of an input step signal,
- I_{OFF} is the drain current after the pull-out event, and
- The switching delay is the time required for the device to reach 50% of its maximum drain current, when the gate voltage is larger than the pull-in voltage (V_{PI})

Examples of the I_{ON} , I_{OFF} , and switching delay, we deduced via SPICE simulations are depicted in Figure 3.3 and Figure 3.3.

The results of our simulations suggest the following:

- switching delay $\propto t_{gap0}, h, 1/L_{BEAM}$ (area)
- $I_{ON} \propto L_{BEAM}, 1/WF, 1/h, 1/t_{gap0}$, and
- $I_{OFF} \propto L_{BEAM}$

Moreover we observe that while there are clear relations between the various device parameters and the NEMFET performance there is no absolute best in breed geometry and various tradeoffs are possible.

Table 3.1 presents three NEMFET configurations that we deduced from our extensive simulations results. The first set of parameters was selected as optimal for low switching time and high I_{ON} , with respect to pull-in and pull-out

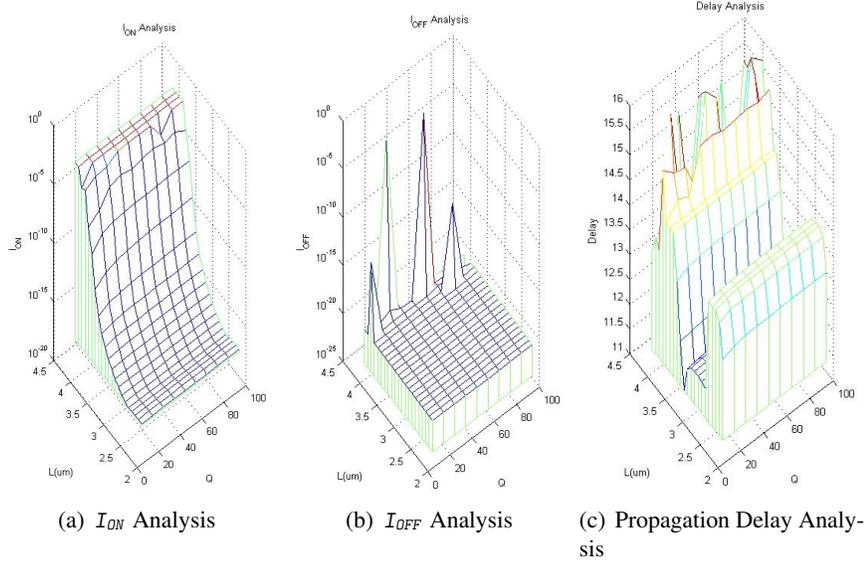


Figure 3.2: (a) I_{ON} Analysis, (b) I_{OFF} Analysis, and (c) Propagation Delay Analysis, for for $WF = 5eV$, $h = 100$ nm, $\tau_{gap0} = 20$ nm.

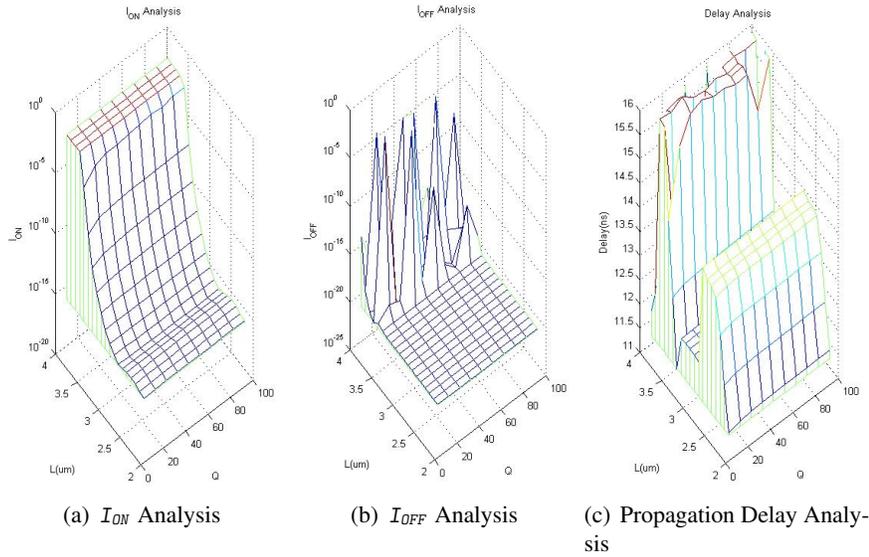


Figure 3.3: (a) I_{ON} Analysis, (b) I_{OFF} Analysis, and (c) Propagation Delay Analysis, for for $WF = 5eV$, $h = 100$ nm, $\tau_{gap0} = 10$ nm.

effects, when (i) $WF = 5\text{eV}$, (ii) $t_{gap0} = 20\text{ nm}, 25\text{ nm}$ and (iii) $h = 70\text{ nm}, 80\text{ nm}, 90\text{ nm}, 100\text{ nm}$. The second set of parameters was selected as optimal for low switching time and high I_{ON} , with respect to pull-in and pull-out effects, when $WF = 4.4\text{eV}, 4.6\text{eV}, 4.8\text{eV}, 5\text{eV}$, selected as optimal for low switching time and high I_{ON} , with respect to pull-in and pull-out effects, when $WF = 4.4\text{eV}, 4.6\text{eV}, 4.8\text{eV}, 5\text{eV}$, $t_{gap0} = 10\text{ nm}, 15\text{ nm}, 20\text{ nm}, 25\text{ nm}$, and $h = 70\text{ nm}, 80\text{ nm}, 90\text{ nm}, 100\text{ nm}$.

Table 3.1: Optimized NEMFET instances for low switching times and high I_{ON} .

	h (nm)	t_{gap0} (nm)	WF (eV)	L (nm)	W (nm)	Q	I_{ON} (mA)	I_{OFF} (Leakage floor)(fA)	Delay (ns)
I	100	20	5	4.2	350	100	2.2	5e-7(12)	8.67
II	100	20	4.4	4	350	100	3	2e-4(4.5e3)	7.41
III	100	10	5	4.5	350	100	1E-06	5e-7(5e-4)	3.03

Table 3.1 indicates that the best compromise between high I_{ON} and low delay, with respect to pull-in and pull-out effects, for $t_{gap0} = 20\text{ nm}, 25\text{ nm}$ and $WF = 5\text{eV}$, can be reached for the set of NEMFET parameters Table 3.1(I).

3.4 90 nm CMOS 32-bit Adder Analysis

To evaluate the potential implications of the utilization of NEMFETs as sleep transistors in a real application we consider a 32-bit adder equipped with a sleep mode circuit. The main purpose of this analysis is to demonstrate that NEMFET can be a better candidate for sleep transistor than FET for this kind of applications. Before doing that, we would like to mention that when properly designing a sleep transistor in a real digital circuit, one must take in consideration the followings parameters:

- leakage current,
- ON current
- area, and
- propagation delay.

Furthermore, the tolerance accepted for the degradation of the gate performance is in the order of at most 5%. We carry on our analysis over the 2.5

GHz CMOS 32-bit adder presented in [3]. For this circuit, in 90 nm technology, with high-Vt devices, $T_{ox}=2.2$ nm, the sleep circuit (sleep transistor together with the sleep circuit that triggers it) has the following characteristics:

- the total leakage is 13 nA,
- the I_{ON} is 10 mA,
- the active area of the sleep transistor is 31.5 μm^2 ,
- the switching delay of the sleep circuit is 10 μs .

If we want to replace the FET sleep transistor in this design with an NEMFET we have to be sure that this one can drive the same active current of 10 mA. To achieve this we need to use twenty-five NEMFET devices in parallel having the following parameters: (i) $W_{BEAM} = 350$ nm, (ii) $L_{BEAM} = 4.2$ μm , (iii) $h = 70$ nm, (iv) $gap = 90$ nm, (v) $N_A = 5 \times 10^{17}$, (vi) $WF = 5$ eV, and (vii) $Q = 100$. These parameters were selected from Table 3.1 as being the best suited for this application. Such a NEMFET structure has the following characteristics: (i) $I_{OFF} = 50 \times 0.4 \times 10^{-21} = 0.02$ aA (sub-threshold drain current), (ii) $I_{ON} = 50 \times 0.2 \times 10^{-3} = 10$ mA, (iii) active area = 73.5 μm^2 , and (iv) propagation delay = 10 ns.

One can observe that this replacement is very much reducing (makes it negligible) the sleep transistor off current since I_{OFF} of NEMFET is much smaller (10 orders of magnitude) than I_{OFF} of normal FET. This suggests that the leakage savings due to the simple replacement of the FET with an NEMFET are very significant. Further research is required however in order to evaluate all the leakage components and to take full advantage of the NEMFET technology by designing special NEMFET tailored sleep circuits. The price that we have to pay for the sub-thresholds reduction is the increase of the active area of the sleep transistor by 130%, which can be quite significant in some circumstances.

It is hard to put into the right perspective the delay aspect. This is twofold: (i) The delay for the original design is covering the entire go to sleep/weak up process thus we do not have information about the actual delay of the FET sleep transistor, and (ii) in this evaluation we assumed that we just replace the FET sleep transistor with an SG-FET without changing anything in the rest of the sleep circuit.

In this evaluation we assumed that we just replace the FET sleep transistor with an NEMFET without changing anything in the rest of the sleep circuit.

As suggested before this may not be the case in practice as this replacement may requires changes in the sleep circuit as well.

3.5 Conclusions

NEMFET based sleep mode circuits are potentially interesting as they clearly enable substantial leakage reductions due to their extremely low OFF currents (10 orders of magnitude lower than FET). This holds true for the adder we considered and by implication for computing dominated applications. There is a clear area overhead associated with the utilization of NEMFET sleep transistors. We note here that our area estimates are not very accurate as they just include the active area of the NEMFET transistors and do not take into consideration extra area overhead that may result from potential changes in the rest of the sleep mode circuit and NEMFET specifics layout design rules. However, given that in nanotechnology context area is not the main issue any longer, as it was replaced by energy consumption and reliability, area overhead may not be perceived as a severe drawback.

Bibliography

- [1] “International technology roadmap for semiconductors,” ITRS, Design, 2007. [Online]. Available: <http://www.itrs.net/>
- [2] D. Tsamados, Y. Singh Chauhan, C. Eggimann, K. Akarvardar, H. S. Philip Wong, and A. Mihai Ionescu, “Finite element analysis and analytical simulations of suspended gate-FET for ultra-low power inverters,” *Solid-State Electronics*, vol. 52, no. 9, p. 13741381, 2008.
- [3] K. von Arnim, P. Seegebrecht, R. Thewes, and C. Pacha, “A low-leakage 2.5ghz skewed cmos 32b adder for nanometer cmos technologies,” in *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, feb. 2005, pp. 380–605 Vol. 1.
- [4] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, “Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture,” in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [5] B. Pruvost, K. Uchida, H. Mizuta, and S. Oda, “Design optimization of NEMS switches for suspended-gate single-electron transistor applications,” *IEEE Transactions on Nanotechnology*, vol. 8, no. 2, pp. 174–184, Mar. 2009.
- [6] M. Enachescu, A. van Genderen, S. D. Cotofana, D. Tsamados, and A. Ionescu, “Benchmark suite and specifications for nem switches for power management,” NEMSIC Deliverable 2.1, 2009. [Online]. Available: http://www.nemsic.org/members/pfn_v2/documents.html
- [7] “Spectre, Cadence Design Systems,” 2009. [Online]. Available: <http://www.cadence.com/us/pages/default.aspx>
- [8] Gabriel M. Rebeiz and Jeremy B. Muldavin, *RF MEMS*. John Wiley & Sons, Inc., 2004, no. 9780471225287.

4

Advanced NEMS-based Power Management for 3D Stacked Integrated Circuits¹

Abstract: In this paper we introduce a novel power management architecture for 3D Through Silicon Vias based integration technology. Our approach relies on the synergy of two new technological developments as follows: (i) we utilize a NanoElectroMechanical (NEM) device, the NEM Field Effect Transistor (NEMFET), as sleep transistor; and (ii) we make use of the 3D potential by placing the sleep transistor (the entire power management infrastructure) on a dedicated tier of the 3D stacked Integrated Circuit. Due to the extreme low leakage current of the NEMFET our proposal results in 2 orders of magnitude static power reduction, when compared with equivalent counterparts based on traditional CMOS devices. The NEMFET power switch requires about 4x more area when compared to bulk CMOS, however, due to the 3D integration which allows for heterogeneous dies to be stacked, the power gating devices can be placed to a low cost dedicated layer, which also results in a substantial IR-drop reduction with minimum impact on leakage.

Copyright © 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purpose must be obtained from IEEE by sending an email to *pubs – permissions@ieee.org*.

¹This chapter is based on the research article published as "Advanced NEMS-based Power Management for 3D Stacked Integrated Circuits" by M. Enachescu, G. Voicu, and S. D. Coto-fana, *IEEE International Conference on Energy Aware Computing (ICEAC)*, pp. 1-4, Cairo, Egypt, December, 2010.

4.1 Introduction

In the nano era context power consumption has become one of the most important issues in Integrated Circuit (IC) design altogether with variability/reliability related phenomena like electro-migration, IR-drop, crosstalk, and gate delay variations [1]. Leakage power, that was irrelevant in technologies above 130nm, has emerged as an important constraint factor as technology node shrinks. To mitigate the leakage power component in circuit design, many low power techniques, e.g., power gating, body bias, multi threshold voltage devices, frequency scaling [2], have been proposed and implemented in practical designs.

A straightforward and successful way to reduce static power consumption associated with inactive hardware is to apply a power shut-off technique (Power Gating) to inactive (sleeping) Functional Units (FUs) in a design implemented in Multi-Threshold CMOS (MTCMOS) technology. To this end Switch (Sleep) Transistors (ST) are placed between the power/ground rails and the FU. When the STs are active power is supplied through them to the FU providing normal operation conditions, and, respectively, when the STs are open the power supply is cut-off to the gated functional unit, hence the leakage power of the entire FU (gated block) is reduced to the leakage power of the STs. The drawbacks of the solution are: (i) area and propagation delay penalty incurred by the addition to the gated block of STs, isolation and state retention cells; and (ii) undesired IR drop voltage on the STs. Thus, along with power gating benefits a number of design issues and power-performance-area trade-offs need to be addressed [3].

For certain specific low activity battery-operated embedded systems used in applications, e.g., environment monitoring sensors, biomedical implants, the sleep mode power, though having a low value, adds up to an energy figure that represents a significant fraction from the total energy consumed by the chip [4, 5]. For this reason new ways to further reduce the leakage power while mitigating the area and delay penalty are needed. Various enhancements for MTCMOS technologies have been proposed to suppress the sleep transistor leakage current [6]. Moreover it has been suggested that nano-electro-mechanical FETs (NEM-FETs) are viable candidates to replace the High- V_t FET based sleep transistor [7, 8]. However, although fabrication feasibility of simple hybrid CMOS-NEMS circuits has been reported in [9], further technological enhancements are needed for such an approach to become a potentially viable industrial solution.

In this paper we propose a novel 3D power management approach that attempts to alleviate some issues associated with the NEMS utilization as sleep transistor in CMOS power gated integrated circuits. Our proposal, graphically depicted in Figure 4.1, relies on a 3D arrangement of the power gated functional units and power management circuits. In this way we allocate a dedicated tier in the stack for the sleep transistors and the afferent control logic. The structure in Figure 4.1 can, in principle, be extended such that more levels of active logic and power management tiers are intermixed. While this structure can be also utilized in conjunction with MTCMOS technology we propose to use Suspended Gate FETs (NEMFET) as sleep transistors due to their extremely low off currents (leakage). The main implications of this approach are as follows: (i) the fabrication process complexity is simplified as we do not make use of a hybrid CMOS and NEMFET technology; (ii) the static power consumption is reduced with up to 2 orders of magnitude due to the extreme low leakage current of the SG-FETs; and (iii) the IR-drop is reduced when compared with the 2D implementation, which has a positive effect on timing.

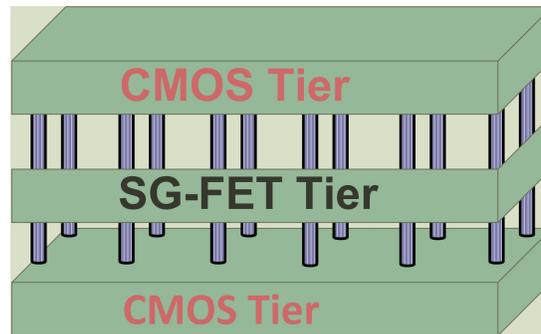


Figure 4.1: 3D Stacked NEMFET Based Power Management Architecture

We note inhere that CMOS-based power gating in 3D stacked circuits has been previously addressed in [10] by studying the influence of TSV placement on the power/ground network noise. Our paper however discusses the feasibility of using NEMFET based power gating in 3D stacked circuits and investigates power management architectures and techniques possible in 3D integration, along with potential design trade-offs. To the best of our knowledge, this paper is the first one to take advantage of the heterogeneous nature of vertical stacked ICs in order to provide a power management architecture based on NEMFET.

To evaluate the practical implications of our proposal we analyze area, IR-drop and leakage for a simplified OpenSPARC T1 processor Execution Unit implemented with power gating using classic CMOS and NEMFET based Sleep

Transistors (STs) in 2D and 3D stacked approaches. Our experiments indicate that the NEMFET based implementation offers up to 2 orders of magnitude reduction of the static power consumption, compared with equivalent counterparts based on traditional CMOS devices, at about 4x more area. Additionally due to 3D integration the NEMFET based solution results in 50% IR-drop reduction with minimum impact on leakage.

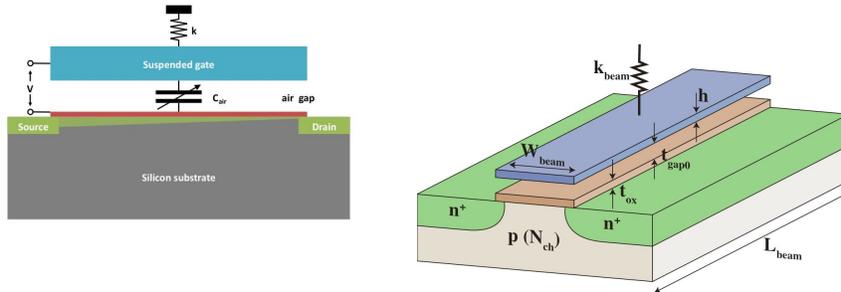
The remainder of the paper is organized as follows: In Section 4.2 a brief introduction is provided on NEMFET including its basic operation. Section 4.3 describes the proposed 3D power management architecture and Section 4.4 presents the results of an evaluation case-study and its associated methodology. Finally, concluding remarks are made in Section 7.6.

4.2 Nano-Electro-Mechanical FET

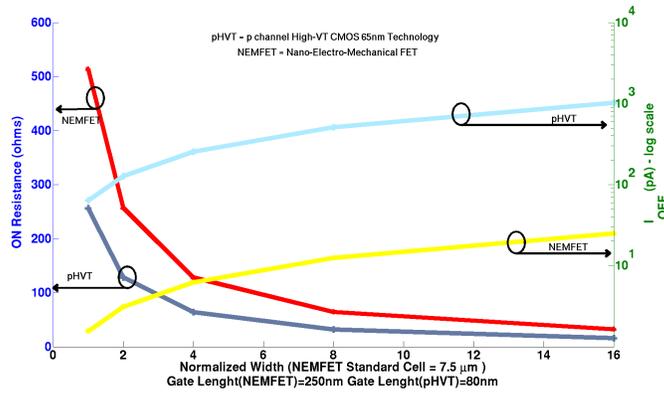
The NEMFET described in [11] is a rather complex device with a 3D geometry as presented in Figure 4.2(b). Essentially speaking the device behaves like an electromechanical switch which responds to gate bias changes as follows. When the gate voltage (V_G) is low the gate-oxide capacitance is in series with the air-gap capacitance (Figure 4.2(a)) resulting in low electrostatic coupling of the gate to the channel thus in a negligible drain current (I_D). If V_G increases the situation remains unchanged until it reaches a certain "on" voltage case in which the electrostatic force cannot be compensated anymore by the mechanical restoring force and the suspended gate (beam) snaps onto the gate oxide, thus turning on the device. This is called *pull-in effect* and corresponds to a sudden I_D increase. After the pull-in, the I_D increase with V_G is comparable with the one of a standard MOSFET. On the other way around when V_G is decreased from some high value I_D starts decreasing until at a certain V_G value the system becomes unstable due to combined electro-mechanical force and the beam is pulled-out. This causes an abrupt I_D decrease due to a large decrease in capacitance and it is called *pull-out effect*.

As indicated in [8] NEMFET devices can potentially replace High- V_t sleep transistors due to their ultra low leakage characteristics. Note that for power gated designs two major features are desirable for the sleep transistors: (i) low sub-threshold leakage current to minimize the static power consumption; and (ii) low "on" state resistance to minimize the voltage difference between the virtual and the real power supply nodes. A comparison between the "on" state resistance R_{ON} and "off" state leakage current I_{OFF} of NEMFET and 90 nm High- V_t CMOS based Sleep Transistors (STs) is presented in Figure 5.2 for

different area in terms of standard cells. The R_{ON} values are computed for an IR-drop of 10 mV over the ST, and the I_{OFF} values consider a 1.2 V power supply.



(a) Schematic representation of a NEMFET and the associated air-gap capacitance C_{air} . V is the potential difference between the gate and the upper surface of the gate dielectric and k is the effective spring constant of the gate. (b) 3D schematic: t_{ox} - the thickness of the gate oxide, h - the thickness of the suspended gate, W_{beam} - the width of the beam, L_{beam} - the length of the beam, t_{gap0} - the gap between the oxide and the suspended gate, k_{beam} - the lumped linear spring constant of the beam.



(c) R_{ON} and I_{OFF} for NEMFET and 90 nm High- V_t CMOS Sleep Transistors.

Figure 4.2: Nano-Electro-Mechanical FET

One can observe that when the sleep transistor area is increasing the NEMFET ST R_{ON} tends to become equal with the CMOS ST R_{ON} , while NEMFET ST I_{OFF} (leakage) is about 2 orders of magnitude smaller than the one of the CMOS counterpart.

4.3 Power Management in 3D Stacked ICs

One of the emergent solutions to achieve tight chip integration is to use 3D stacking technology with Through Silicon Vias (TSVs) as interconnects between the stacked dies [12]. TSVs are relative large metal vias ($<10\ \mu\text{m}$ diameter, $<100\ \mu\text{m}$ length) that are passing completely through the silicon die. The immediate advantage of vertically stacking dies is the interconnect latency reduction, which lately became the dominant latency factor [13]. Moreover, dies implemented in different technologies can be part of the same stack, which opens novel system integration avenues and tradeoffs. The side effect of this performance improvement is the increase in power density in the dies placed in the middle of the stack.

Current power gating architectures place the STs as either a ring surrounding the gated block or as columns throughout the gated block [14]. In this paper we introduce a power management architecture that capitalizes on the heterogeneity of 3D stacked systems. The proposed architecture makes use of NEMS technology dies containing NEMFET STs placed between dies containing the actual power gated circuits, as illustrated in Figure 4.3 for a 2-die stack. By moving the bulky STs to a different die, precious area surrounding/inside the gated blocks previously allocated to them in the planar case can be reclaimed. Furthermore, interconnect length inside/between gated blocks is reduced, thereby potentially resulting in increased performance. Moreover this approach can make use of the extreme low NEMFET leakage without requiring an expensive and currently not available NEMFET CMOS hybrid fabrication technology.

One of the key issue in the proposed approach relates to the way the TSV electrical characteristics [15] influence the power gating efficiency and granularity. The "on" state voltage drop-out across the TSV and their behavior at power-on need to be addressed. Copper TSVs exhibit low resistivity, approximately $0.2\ \Omega$ for a $5\ \mu\text{m}$ diameter and $20\ \mu\text{m}$ length, hence negligible when compared with the power switch R_{ON} . The TSV pillar is isolated through an oxide layer from the silicon substrate thus the formed parasitic capacitance, typically $40\ \text{fF}$, has the predominant effect on the TSV propagation delay, and potentially limits the power-on time. This capacitance adds up to the gated circuit capacitance, therefore the TSV has a direct impact on the granularity at which the power gating technique can be optimally applied. This capacitance combined with the TSV minimum footprint constraint make the scaling down of the gating granularity being limited by the TSV characteristic instead of the power switch ones.

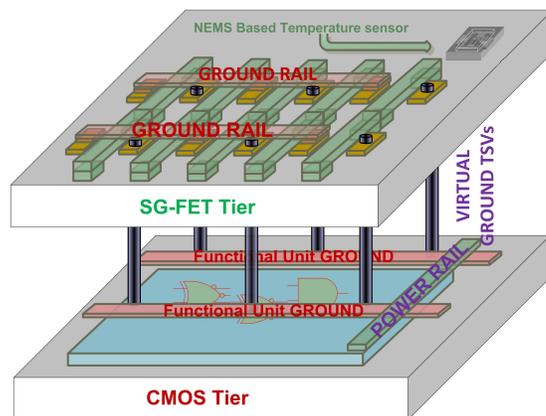


Figure 4.3: Detailed Representation of 3D Stacked NEMFET Based Power Management Architecture

The power management controller and the always-on cells (for isolation and state-retention) can be placed either on the NEMS die (NEMFET based logic gates were successfully simulated in [16]), either on the logic die containing the gated blocks. As indicated in Table 4.1, the power switches area is lower than the one of the associated gated logic, thus it is advantageous from the silicon area point of view to place the NEMFET die between two computing dies and/or to unify all the power management related blocks on a dedicated die. However, the latest solution infers using a hybrid CMOS NEMFET die processing technology, which can affect the yield. Alternatively, having a single process NEMFET die with a greater yield decreases the manufacturing cost. This cost can also be reduced by designing a generic power gating die consisting of a regular array of NEMFET power switches and CMOS always-on cells that can be reused for any logic designs just by connecting the TSVs to the stacked dies. The power management controller, which is usually chip-dependent, can be placed on the logic die.

To increase the reliability, we suggest to also include a thermal management mechanism by placing NEMS temperature sensors on the same die with the NEMFET sleep transistors. In this way the power management controller can determine when a processing core becomes too hot and signals the operating system to migrate the threads running on that core to other cooler cores. Once the migration is complete the power controller can turn off the core and allow it to cool down.

4.4 Experimental setup and results

To preliminarily evaluate the practical impact our proposal we performed a case study on the 2D and 3D implementations of a simplified OpenSPARC T1 processor [17] Execution Unit containing the ALU and the Shifter. We implemented the design in a commercial 90 nm Low Power MTCMOS technology using Cadence Encounter Digital Implementation 9.1 Low Power Flow. The design is successfully signed-off at 333 mHz. To evaluate the power savings we first implemented a reference design with no power gating mechanisms (the entire design is always powered-on), then we placed the Shifter in a switchable power domain, with 90 nm High- V_t CMOS power switches and NEMFET switches, respectively. Four layouts are investigated: the reference one without power gating (REF), 2D layouts with power gating with classic High- V_t CMOS (2DHVT) and NEMFET (2DSGF) and a two-tier 3D layout with one logic die and a NEMS die with NEMFET power switches (3DSGF).

To dimension the sleep transistors we first had to determine the Maximum Instantaneous Current (MIC) of the Shifter in the routed design which is 4.73 mA. Using the Finite Element Modeling characterization of NEMFET in [16], we sized the power switches in both 2D cases according to [18], in order to achieve a target IR-drop under 20 mV on them for the measured MIC. The total number of required NEMFET/High- V_t CMOS standard cells are presented in Table 4.2. Our 3D stacked floorplan takes advantage of the extra area available and increases the area of NEMFET based ST up to the Shifter unit one, in order to equally distribute the MIC on the STs and to mitigate the IR-drop effect.

Accurate dynamic power and rail analysis with the switchable power domain functioning and turned off were performed using Cadence VoltageStorm to evaluate the power gating efficiency. The switching activity was set to 0.2. Since a dynamic characterization of the NEMFET device was not available at that time, we generated a power grid library containing only the static port view model of it, based on the R_{ON} , $I_{D_{sat}}$ and I_{OFF} values from the FEM analysis of the device. The static port power-grid view of the High- V_t CMOS power switch was used for consistency. For the High- V_t CMOS power switch we ran a multi-mode multi-corner analysis in typical ($V_{DD} = 1.2$ V, $T = 25$ °C), best ($V_{DD} = 1.32$ V, $T = 0$ °C), and worst case ($V_{DD} = 1.08$ V, $T = 125$ °C) conditions; the NEMFET analysis was done only in the typical case.

Table 4.1 shows that when using 2D High- V_t CMOS based implementation the Shifter leakage current is reduced by 34x with a cost of 19 mV IR-drop and 13.25% area increase, and negligible total power penalty. By using NEMFET

Table 4.1: Results

Implementation Type	Shifter	Area (μm^2)		Power switch parameters								Total power (mW)	
		Power switches	Power switch overhead (%)	ON - average IR-drop (mV)		OFF - I_{OFF} (nA)				ON	OFF		
				BC	TC	WC	BC	TC	WC				
REF	4487	N/A	N/A	N/A	N/A	N/A	1.08E3	2.93E2	2.84E3	6.731	N/A		
2DHVT	5081	594.72	13.25	14.26	19.09	25.82	1.25	8.62	2.22E2	6.733	5.9		
2DSGF	6626	2139	47.67	N/A	17.12	N/A	N/A	3.01E-2	N/A	6.73	5.867		
3DSGF	ST die Logic die	4482 N/A	99.95, but different die	N/A	8.61	N/A	N/A	6.0E2	N/A	6.73	5.93		

Table 4.2: Power switches dimensioning

Power gating standard cell	Channel width (W) (μm)	I_{ON} at $V_{DS} = 20 \text{ mV}$ (mA)	I_{OFF} at $V_{DS} = 1.2 \text{ V}$ (A)	Number of cells
High- V_t based	52	513	8.58E-10	10
NEMFET based	7.5	51	4.47E-13	93

based ST we further reduce by 286x the Shifter leakage, with a further addition in area cost of 34.43%, for roughly the same IR-drop. Furthermore, by placing the NEMFET based STs on a separate die and increase its area according to the Shifter footprint, the IR-drop is halved with a minimum effect on leakage.

4.5 Conclusions

In this paper we introduced a novel NEMFET based power management architecture for 3D Through Silicon Vias based integration technology. Our proposal demonstrates that 3D integration is an effective way to make use of the NEMFET characteristics, i.e., abrupt switching and extreme low leakage, in effective NEMS-based power/energy/thermal management.

Bibliography

- [1] “International technology roadmap for semiconductors,” ITRS, Design, 2007. [Online]. Available: <http://www.itrs.net/>
- [2] N. H. E. Weste and D. M. Harris, *CMOS VLSI Design A Circuits and Systems Perspective*. Addison Wesley, 2011.
- [3] H. Jiang, M. Marek-Sadowska, and S. R. Nassif, “Benefits and costs of power-gating technique,” in *ICCD 2005, IEEE International Conference on Computer Design: VLSI in Computers and Processors*, 2005, pp. 559–566.
- [4] H. S. Won, K. S. Kim, K. O. Jeong, K. T. Park, K. M. Choi, and J. T. Kong, “An MTCMOS design methodology and its application to mobile computing,” in *Proceedings of the 2003 international symposium on Low power electronics and design*, 2003, p. 115.
- [5] G. Panic, Z. Stamenkovic, and R. Kraemer, “Power gating in wireless sensor networks,” in *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, 2008, pp. 499–503.
- [6] *Low Power Methodology Manual*. Boston, MA: Springer US, 2007. [Online]. Available: <http://www.springerlink.com/index/10.1007/978-0-387-71819-4>
- [7] H. F. Dadgour and K. Banerjee, “Design and analysis of hybrid NEMS-CMOS circuits for ultra low-power applications,” in *Proceedings of the 44th annual Design Automation Conference*, 2007, pp. 306–311.
- [8] M. Enachescu, S. Cotofana, A. Genderen, D. Tsamados, and A. Ionescu, “Can SG-FET replace FET in sleep mode circuits?” *Nano-Net*, pp. 99–104, 2009.
- [9] N. Abele, R. Fritschi, K. Boucart, F. Casset, P. Ancey, and A. M. Ionescu, “Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor,” in *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 2006, pp. 479–481.
- [10] Y. Xu, W. Liu, Y. Wang, J. Xu, X. Chen, and H. Yang, “On-line MP-SoC scheduling considering power gating induced Power/Ground noise,” in *2009 IEEE Computer Society Annual Symposium on VLSI*, Tampa, Florida, USA, May 2009, pp. 109–114.

- [11] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, "Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture," in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [12] W. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. Sule, M. Steer, and P. Franzon, "Demystifying 3D ICs: the pros and cons of going vertical," *IEEE Design and Test of Computers*, vol. 22, no. 6, pp. 498–510, Jun. 2005.
- [13] J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl, "Interconnect limit on gigascale integration in the 21st century," *Proceedings of the IEEE*, vol. 89, no. 3, pp. 305–324, 2002.
- [14] "TSMC reference flow 7.0," 2010. [Online]. Available: <http://www.tsmc.com>
- [15] G. Katti, A. Mercha, J. Van Olmen, C. Huyghebaert, A. Jourdain, M. Stucchi, M. Rakowski, I. Debusschere, P. Soussan, W. Dehaene *et al.*, "3D stacked ICs using cu TSVs and die to wafer hybrid collective bonding," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2010, pp. 1–4.
- [16] D. Tsamados, Y. Singh Chauhan, C. Eggimann, K. Akarvardar, H. S. Philip Wong, and A. Mihai Ionescu, "Finite element analysis and analytical simulations of suspended gate-FET for ultra-low power inverters," *Solid-State Electronics*, vol. 52, no. 9, p. 13741381, 2008.
- [17] "OpenSPARC t1 processor," 2010. [Online]. Available: <http://www.opensparc.net/opensparc-t1/index.html>
- [18] M. Anis, S. Areibi, and M. Elmasry, "Design and optimization of multithreshold cmos (mtcmos) circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 10, pp. 1324–1342, Oct. 2003.

5

Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture for Autonomous Sensors Systems¹

Abstract: With the technology moving into the deep sub-100 nm region, the increase of leakage power consumption necessitates more aggressive power reduction techniques using emerging devices. Power gating with Nano-Electro-Mechanical Field Effect Transistors (NEMFET) is a promising avenue to reduce energy consumption of embedded autonomous sensor systems. Our research emphasizes that 3D Stacked hybrid circuits with NEMFET sleep transistors can be further enhanced to reduce leakage power by redesigning the entire power management circuitry with NEMFETs. To evaluate the practical implications of such an approach we implement NEMFET based power gating, which makes use of NEMFETs as sleep transistors, isolation cells, and components for power management controller design, on an embedded SoC platform running a bio-medical sensing application. Preliminary energy evaluations with Cadence EDI flow indicate that the enhanced architecture provides a reduction of 7% over the 3D Hybrid architecture at the expense of 4.7% area increase and of about 15% energy reduction with respect to the "classic" 2D CMOS counterpart. Furthermore, for applications with lower activity, the potential energy improvement the enhanced architecture could provide can reach up to 90% with respect to the 2D CMOS reference design.

Copyright © 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purpose must be obtained from IEEE by sending an email to pubs-permissions@ieee.org.

¹This chapter is based on the research article published as "Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture for Autonomous Sensors Systems" by M. Enachescu, G. R. Voicu, and S. D. Cotofana, *IEEE 15th International Conference on System Theory, Control and Computing (ICSTCC 2011)*, pp. 224-229, Sinaia, Romania, October, 2011. *Best PhD student paper award.*

5.1 Introduction

CMOS technology is approaching its physical limits; hence emerging technologies have been investigated and proposed to supersede the basic CMOS device, i.e., the MOSFET. Examples of such devices are Nano-Electro-Mechanical FETs (NEMFETs), carbon nano tubes, quantum-dot cellular automata, single-electron devices [1–4]. However, at this time, only low complexity circuits were successfully designed with the above technologies.

On the other hand, scaling CMOS technology could still provide 30% improvement in performance at 28 nm node with respect to 40 nm node according to [5]. The price to be paid reflects also into the tremendously increase of power dissipation of digital CMOS. Hence, the life of the battery operated embedded systems with low activity, e.g., autonomous sensor systems, which is mostly determined by the standby leakage power dissipated by the circuit in sleep mode, is directly affected. Thus, the use of power gating technique in portable devices is vital in achieving long life, low energy implementation solutions.

Power gating works on the principle of using Sleep Transistors (STs) to cut off supply voltage to the circuit in standby mode thus cutting off leakage current path to ground. Various enhancements based on Multi-Threshold CMOS (MTCMOS) technologies have been proposed to suppress the sleep transistor leakage current [6], however the NEMFET based ST proved to be more effective when compared with High- V_T counterpart [7].

One solution that partially address the leakage issue by exclusively minimizing the leakage of the ST is a hybrid CMOS-NEMS 3D stacked chip architecture. Benefitting from the NEMS-based power management, the architecture proposed in [8], emphasizes the outstanding characteristics of NEMFET devices, i.e., "abrupt" switching and ultra low leakage. To validate this proposal and evaluate its performance in a real-life scenario a careful assessment of the implications of this hybrid power management architecture on the rest of the system was described in [9].

Since power gated circuits spend very long times in standby mode, one can essentially infer that the overall energy can be further substantially reduced when the overall OFF state power consumption mostly generated by the Always-On (AO) cells is mitigated. In this paper we propose a novel 3D power management approach that attempts to alleviate some leakage overhead associated with the use of CMOS devices as AO cells, in isolation (ISO) cells and in the

Power Management (PM) controller. Our proposal relies on: (i) moving the ISO cells and the PM controller on the NEMS die, and (ii) redesigning them in the NEMS technology to take advantage of the NEMFET ultra low leakage power. Our goal is to explore these implications and evaluate the performance and the energy efficacy of this power management architecture in a real-life scenario.

Our experiments indicate that, due to the extreme low leakage and the "abrupt" switching of the NEMFET, the system idle energy is decreased by 6.35x at the expanse of a 4.75% area overhead on the NEMS tier. Moreover, due to the leakage-enhanced 3D hybrid approach the energy-delay product of the embedded SoC platform is reduced by 7% with a potential improvement of up to 90% with respect to the 2D CMOS reference design.

Even though the NEMFET based AO cells add overhead to the total area, this does not affect the chip footprint since it is displaced to another tier (which contains NEMFET sleep transistors). Moreover, since the AO cells are not on the critical path, their replacement with the NEMFET-based cells, which for the time being have a limited operation frequency of about 100 MHz, do not change the overall frequency of our approach. Exploiting 3D stacked hybrid integration, our approach can be applied directly on an industrial low power design which supports NEMS and CMOS technology.

The rest of the paper is organized as follows. First, in Section 5.2, we give an overview of the 3D hybrid power management architecture and some relevant NEMFET background. Section 5.3 describes the enhanced-leakage architecture, which makes use of NEMFETs for the implementation of the entire power gating related infrastructure, i.e., sleep transistors, isolation cells, and power management controller. Section 8.4 discusses the experimental results and finally, Section 8.5 presents the conclusions.

5.2 NEMFET based power management architecture

5.2.1 NEMFET Background

The Nano-Electro-Mechanical FET (NEMFET) described in [10] is a rather complex device with a 3D geometry and cross-section as presented in Figure 5.1a.

Essentially speaking the device behaves like an electromechanical switch

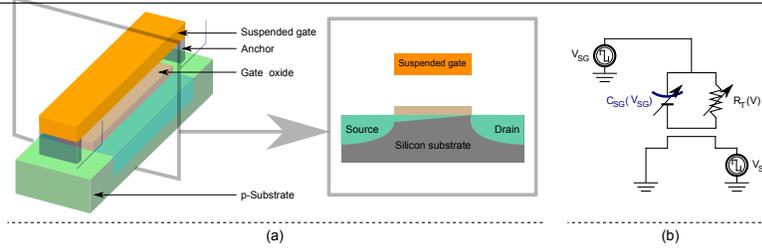


Figure 5.1: (a) Schematic diagram of NEMFET. (b) Equivalent circuit model for the NEMFET

which responds to gate bias changes as follows. When the gate voltage V_G is low the gate-oxide capacitance is in series with the air-gap capacitance Figure 5.1b resulting in low electrostatic coupling of the gate to the channel thus in a negligible drain current I_D . If V_G increases the situation remains unchanged until it reaches a certain "ON" voltage case in which the electrostatic force cannot be compensated anymore by the mechanical restoring force and the suspended gate (beam) snaps onto the gate oxide, thus turning on the device. This is called pull-in effect and corresponds to a sudden I_D increase. After the pull-in, the I_D increase with V_G is comparable with the one of a standard MOSFET. On the other way around when V_G is decreased from some high value I_D starts decreasing until at a certain V_G value the system becomes unstable due to combined electro-mechanical force and the beam is pulled-out. This causes an abrupt I_D decrease due to a large decrease in capacitance and it is called pull-out effect.

As indicated by [7] NEMFET devices can potentially replace High- V_T sleep transistors due to their ultra low leakage characteristics. Note that for power gated designs two major features are desirable for the sleep transistors: (i) low sub-threshold leakage current to minimize the static power consumption, and (ii) low "ON" state resistance to minimize the voltage difference between the virtual and the real power supply nodes.

A comparison between the "ON" state resistance R_{ON} and "OFF" state leakage current I_{OFF} of NEMFET and 65 nm High- V_T CMOS based Sleep Transistors (STs) is presented in Figure 5.2 for different area in terms of standard cells. The R_{ON} values are computed for an IR-drop of 10 mV over the ST, and the I_{OFF} values consider a 1.2 V power supply. One can observe that when the ST area is increasing the NEMFET ST R_{ON} tends to become equal with the CMOS ST R_{ON} , while NEMFET ST I_{OFF} (leakage) is about 2 orders of magnitude

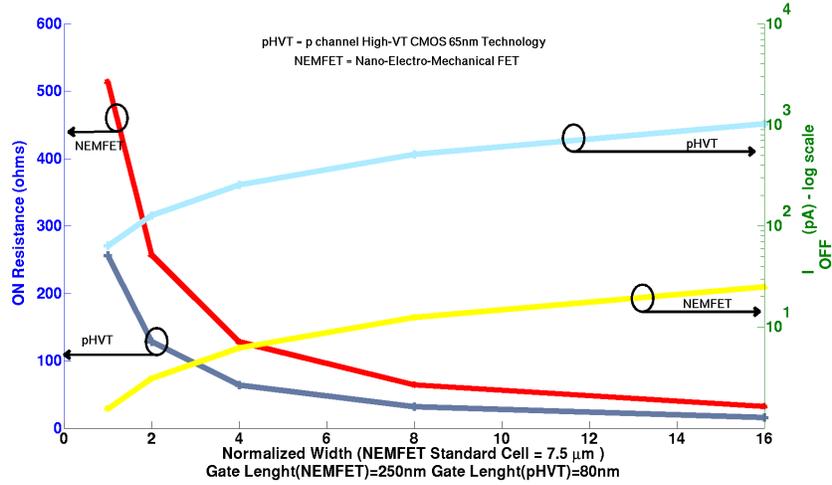


Figure 5.2: R_{ON} and I_{OFF} for NEMFET and 65 nm High- V_t CMOS Switch Transistors

smaller than the one of the CMOS counterpart.

5.2.2 3D-Stacked Hybrid Power Management Architecture

In [9] the integration impact that a recent proposed 3D stacked power management architecture based on Nano-Electro-Mechanical FET power switches may have on a real-life embedded SoC design was described. The architecture relies on NEMS technology dies containing NEMFET STs placed on top of the die containing the actual power gated circuits, as in the two-tier stack test vehicle from Figure 5.3. The bottom CMOS tier comprises the active logic circuit, while the top tier incorporates the NEMFET STs. The dies are stacked using Through-Silicon Vias (TSVs), allowing for large density, high speed, and low power interconnect [11].

The approach has a number of advantages as follows: (i) it combines the appealing extremely low leakage currents of the NEMFETs with the versatility of CMOS technology by allowing for the power switches to be fabricated on a separate die, (ii) it simplifies the floorplanning in general, and can increase the computation platform performance because of extra area cleared by the switches on the CMOS tier, and (iii) it leverages the integration of other NEMS/MEMS devices, e.g., energy harvesters, sensors, on the same tier with the power management circuitry.

The test vehicle has three power domains (PD Core, PD DMem and PD PMem), which can be independently powered-off by the corresponding group of sleep transistor on top of the power domain. The Always-On (AO) power gating related blocks, contained in the PCM power domain are part of the bottom die. The next section presents a way to reduce the leakage power by the relocation of the PCM power domain, as displayed by the light blue arrow in Figure 5.3.

5.3 Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture

We propose to enhance the above presented PM architecture by moving the Always-On (AO) power domain location to the NEMS die. This entails re-designing the isolation (ISO) cells and the PM controller in the NEMS technology. This translation alleviates part of the leakage overhead associated with the use of CMOS devices as AO cells by taking advantage of the NEMFET ultra low leakage power. The bottom CMOS tier comprises the active logic circuit, while the top tier incorporates the STs, ISO cells, and PM controller implemented with NEMFETs.

The Always ON cells include the following elements:

1. Isolation cells which controls the outputs of powered down blocks and clamp the outputs to a specific, legal value;
2. The Power Management Controller, which generates the control signals for the STs and the ISO cells in the appropriate sequence for shut-down and power-up transitions.

5.3.1 Isolation cells

Floating outputs of the power-down blocks altogether with crowbar currents during powering down may result in spurious behavior in the power-up blocks inputs [6]. Isolation cells are practically basic NAND/NOR gates, or even single pull-up/pull-down transistors with the gate input acting as the isolation signal generated by the PM controller that can hold the output to logic high/low.

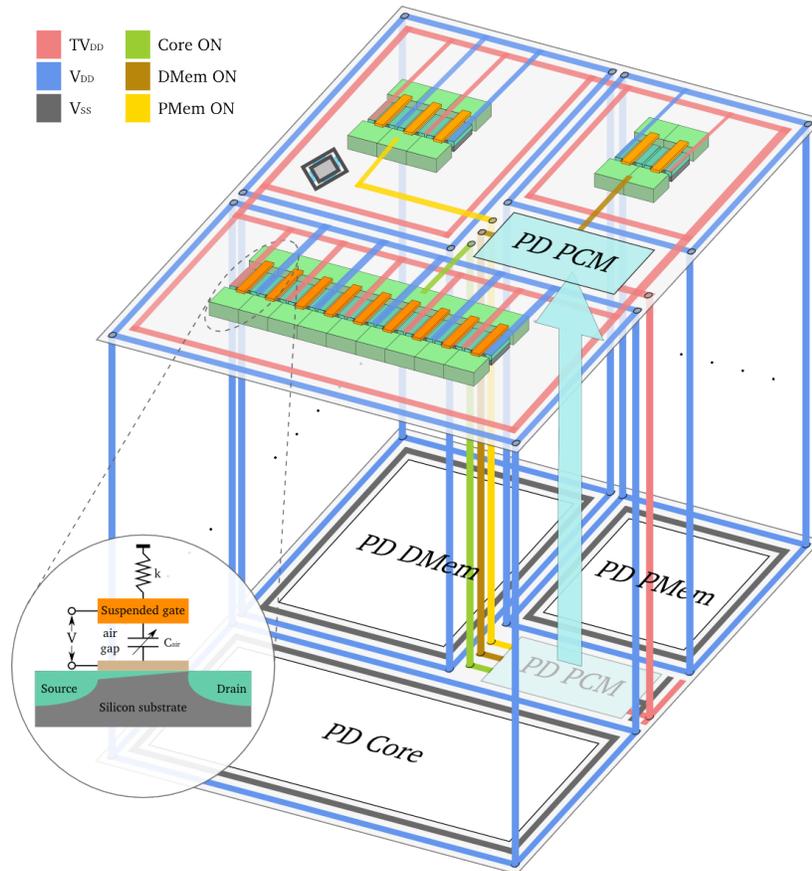


Figure 5.3: Leakage-enhanced 3D Stacked NEMFET Power Management Architecture

By replacing the isolation cells with NEMFET pull-up/pull-down transistors as indicated in Figure 5.4 we take advantage of the ultra-low leakage characteristic of the NEMFET to reduce with approximately two orders of magnitude the additional energy cost incurred by isolation. The twofold trade-offs are area and delay penalties. Taking into account that the operating frequency of a NEMFET based logic is currently limited to about 100 MHz and the parasitic TSV interconnect capacitance we can expect that the isolation sequence will increase from one to two clock cycles.

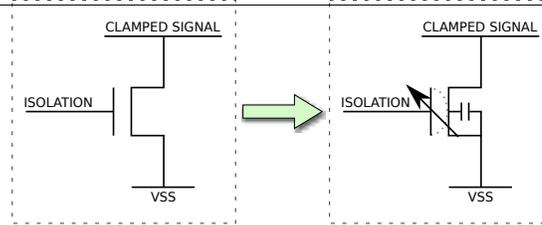


Figure 5.4: Substitution of nMOS Pull-down Isolation cell with NEMFET cell

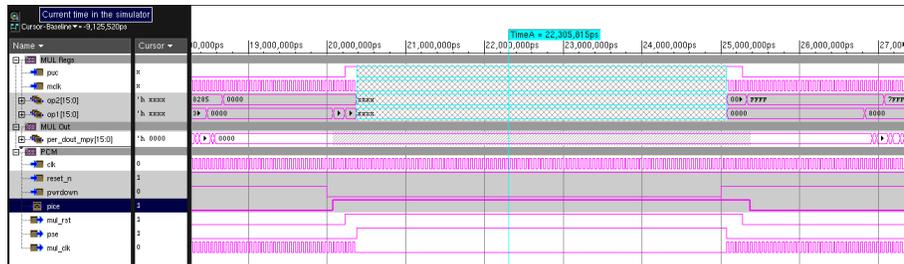


Figure 5.5: Simulated waveform of Power Management Controller signals

5.3.2 Power Management Controller

This module performs the following sequence of operations to power-down:

1. Stop the clock to minimize leakage into the power-gated region;
2. Assert the isolation control signal to park all outputs in a safe condition;
3. Assert reset to the block, so that it powers up in the reset condition;
4. Assert the power gating control signal to power down the block.

To restore power the same sequence is performed in the reversed order. The PM control signals for a complete power-down/power-up cycle simulated in Cadence LP-NCSim are depicted in Figure 5.5.

5.4 Performance evaluation

To evaluate the energy savings of the 3D stacked NEMFET based Power Management (PM) architecture an energy analysis was first performed on the Reference single-tier design with classic High- V_T sleep transistors. Subsequently,

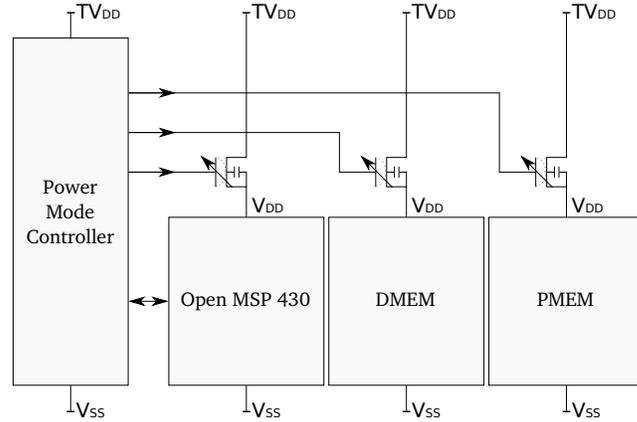


Figure 5.6: System on Chip platform for autonomous sensors

by moving the bulky STs to a different die, the Stacked design, precious area surrounding/inside the gated blocks previously allocated to them in the planar case can be reclaimed. Finally, by substituting the High- V_T STs with the equivalent area of NEMFET STs, the Hybrid implementation was analyzed.

Figure 5.6 presents the considered experimental platform. It consists of a typical SoC for low power embedded devices, based on the open-source 16-bit synthesizable processor core openMSP430 [12], a clone of the commercial Texas Instruments MSP430 microcontroller. The program memory (PMEM) and data memory (DMEM) sizes are 4-KB and 2-KB, respectively. Although not used by the reference application, all the peripherals of the openMSP430 core (16-bit multiplier, timer, 6 8-bit input/output ports) are part of the platform in order to account for their leakage power.

The reference application is a heart beat rate monitor which detects the QRS complex in a digitized electrocardiogram (ECG) signal. The QRS event corresponds to the depolarization of the ventricles, comprising a series of three deflections seen on a typical ECG signal. The middle one, i.e. the R wave, is the well-known “peak” of the ECG. The application program identifies the R peaks in the ECG signal and measures the interval between two consecutive R peaks. The algorithm we use is based on the open source arrhythmia detection software from EP Limited [13], which uses the filtering based Pan-Tomkins [14] method for R peak detection. For this method the input data are channeled through a five steps filtering process consisting of a low pass filter, a high pass filter, a derivative, an absolute value, and an integrator function, in this order. The utilized filter equations are valid for an ECG sample frequency

5.4.1 3D-Stacked NEMFET-based Power Management Architecture evaluation

The reference platform was implemented using Cadence Encounter Digital Implementation [15] in a 65nm commercial Low Power CMOS technology. A high operating frequency of the SoC means that it has to be synthesized with tighter timing constraints, which results in a design with faster and/or bigger cells. As a consequence leakage and dynamic power increase. For the presented SoC this effect limits the operating frequency to 150MHz without causing a significant increase in the leakage power.

The obtained energy consumption values for the three designs mentioned at the beginning of this section are presented in Table 8.2. One can observe that in spite of the reduced OFF leakage power the total improvement on the energy-delay product figure is only 9%. This happens because of two reasons: (i) the test application has a too high duty cycle, and (ii) the ON power term is dominant in the total energy equation.

The power consumption could be further reduced in two ways:

- addressing the OFF state power through implementing the PM controller and the isolation cells with NEMFET devices as discussed in Section III,
- optimizing the design in terms of ON state power and the software application to optimally utilize the hardware resources, hence reducing the run time.

5.4.2 Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture evaluation

The proposed enhanced 3D power management hybrid approach makes use of the first option and redesigns the entire extra power gating logic using low-power NEMFET technology and place it on the different tier, taking advantage of the 3D stacking technology. We compare the enhanced architecture with the hybrid one and evaluate the overall energy gain from the embedded system running the bio-medical application perspective. Using the same power

Table 5.1: Power and energy results

Implementation	ON power	OFF power [nW]			Energy
	[mW]	STs	Always-on cells*	Total	[μ J]
<i>Reference</i>	5.0340	56.28	26.02	81.28	9.05
<i>Stacked</i>	5.0343	900.44	26.02	925.44	9.46
<i>Hybrid</i>	5.0339	4.52	26.02	30.54	8.61

* Always-on cells = Power management controller and isolation cells

analysis methodology we extracted the power and area values presented in Table 5.2.

Even though the total area overhead due to logic implementation using NEMFETs is 4.75% of the design size, this does not affect the chip footprint since it is displaced to another tier (which contains NEMFET switch transistors). The prominent reduction of about two orders of magnitude in the OFF state power is due to the practical "zero-leakage" characteristic of NEMFET. Furthermore, the abrupt switching effect of the NEMFET causes a reduction in the dynamic power of PM controller and ISO cells, reducing the total ON state power. The changes in active and idle power are directly reflected in the energy figure, improving the energy consumption by 7% with respect to the hybrid design. Compared with the classic one-tier power-gated CMOS implementation Table 8.2 the overall energy saving could be up to 15%.

Table 5.2: Leakage-enhanced architecture area, power and energy results

Implementation type	Area [% of die size]		ON power [mW]	STs	OFF power [nW]		Total energy* [μ J]
	ISO cells	PM Controller			ISO cells	PM Controller	
<i>Hybrid [9]</i>	0.15	0.02	5.0339	4.52	18.01	8.01	8.61
<i>Leakage-enhanced</i>	3.75	1.17	4.7	4.52	0.289	0.075	8.00

* Active time=1.73 ms, Idle time=993.35 ms, Transition time=5.32 ms @ 150 MHz.

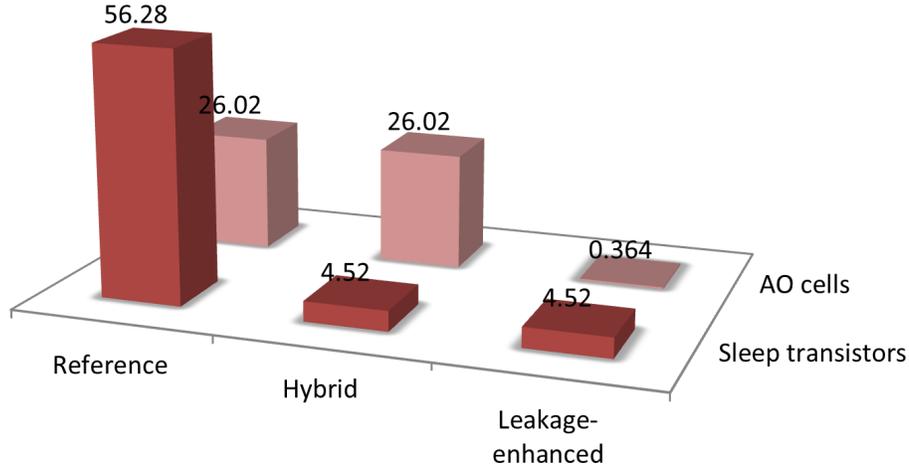


Figure 5.7: Breakdown of leakage power (in nW) in an embedded processor for autonomous sensors

Figure 5.7 presents a comparison of the system idle power for the classic High- V_T Reference, i.e., the 3D Hybrid, and the proposed Leakage-enhanced architectures. The idle power is equal with the leakage power of the only active components in the idle state, the STs and the AO cells. A two-step approach to reduce the leakage power can be observed. Hybrid design implements in the first step only the STs in the NEMFET die, which results in a substantial reduction of the ST leakage but doesn't affect the AO cell figure of merit. In the second step, the Leakage-enhanced design contains all STs, AO cells and the complete PM circuit on the top die, implemented with NEMFETs. In total, by using the Leakage-enhanced architecture, we reduce the leakage power in the idle state with 92%, from 82.3 nW in the 2D Reference design to 5.884 nW.

The ON state power consumption can be further reduced by optimizing the design in terms of power and the software application to optimally utilize the hardware resources, thus reducing the run time. To characterize the energy consumption efficiency versus various applications, we plot in Figure 5.8 the platform energy consumption for different operating duty cycles. We assume the same ON and OFF state power figures as for the bio-medical application. The duty-cycle is defined as: $T_{ON} / (T_{ON} + T_{OFF})$.

One can observe in the figure that the energy requirement has a steep decrease until the application duty-cycle reaches 0.0001. Optimizing and reducing the application duty-cycle lower than this does not bring much energy savings any

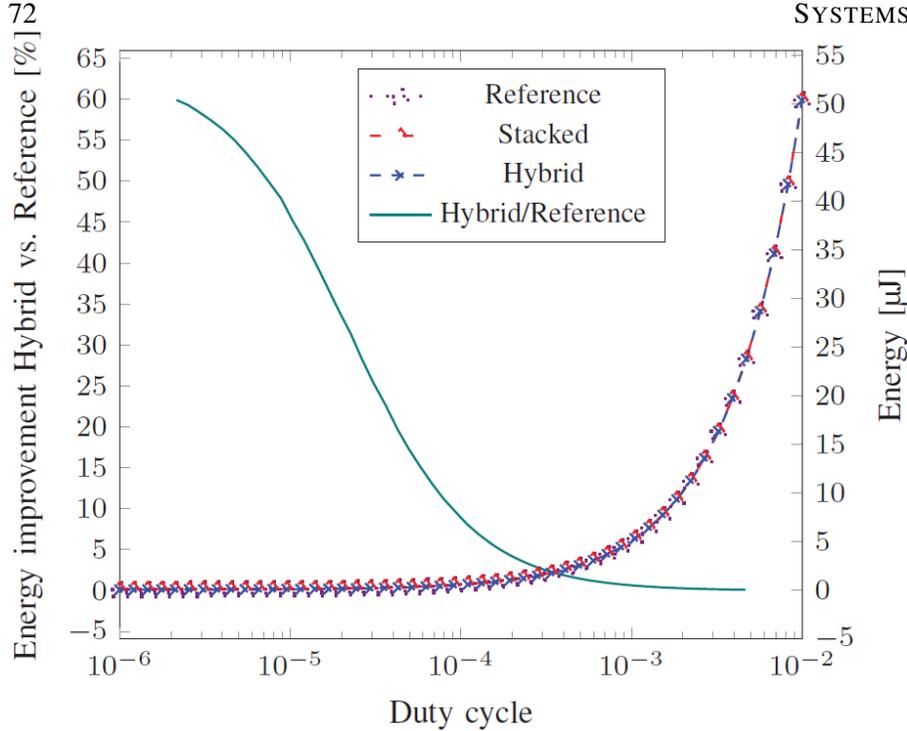


Figure 5.8: Energy Consumption versus Duty-cycle

longer. However, from this point further, for even lower activity applications the relative savings in energy consumption due to our power architecture when compared to High- V_T STs start to increase, reaching up to 90% lower energy consumption. Compared to the previous 3D-Stacked Hybrid PM architecture the overall energy improvement is about up to 50% higher.

5.5 Conclusions

In this paper, we diminished the leakage power of the NEMFET based 3D-stacked power management architecture by making use of NEMFETs also for the implementation of isolation cells, and power management controller design. We evaluated the practical implications of such an approach by implementing an embedded SoC platform running a bio-medical sensing application with NEMFET based power gating. We performed energy evaluation of the enhanced design and our results indicated a reduction of 7% over the

3D Hybrid architecture with the price of 4.7% increase in area and of about 15% energy reduction with respect to the "classic" 2D CMOS counterpart. Furthermore, for applications with lower activity, the potential energy improvement of the Leakage-enhanced architecture could deliver can reach up to 90%, with respect to the 2D CMOS reference design. Our results suggest that when utilized in conjunction with CMOS the NEMFETs can induce significant performance improvements over CMOS only approaches in terms of energy efficiency. This justifies the need for further efforts in evaluating other aspects of the novel power management architectures, such as reliability and technology integration issues.

Acknowledgment

This work has been jointly supported by the EU commission via the 7th Framework Programme project NEMSIC, reference 224525, and by Catrene and Agentschap NL via CT105 project 3DIM³.

Bibliography

- [1] D. Tsamados, Y. Singh Chauhan, C. Eggimann, K. Akarvardar, H. S. Philip Wong, and A. Mihai Ionescu, "Finite element analysis and analytical simulations of suspended gate-FET for ultra-low power inverters," *Solid-State Electronics*, vol. 52, no. 9, p. 13741381, 2008.
- [2] P. L. McEuen, M. S. Fuhrer, and H. Park, "Single-walled carbon nanotube electronics," *IEEE Transactions on Nanotechnology*, vol. 1, no. 1, p. 7885, 2002.
- [3] W. Lu and C. L. Lieber, "Nanoelectronics from the bottom up," *Nature Materials*, vol. 6, no. 11, p. 841850, 2007.
- [4] A. Dehon, "Nanowire-based programmable architectures," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 1, no. 2, pp. 109–162, Jul. 2005.
- [5] John Wei, "TSMC Technology Symposium - Leading Edge Technologies," TSMC, Amsterdam, NL, Tech. Rep., Jun. 2011.
- [6] *Low Power Methodology Manual*. Boston, MA: Springer US, 2007. [Online]. Available: <http://www.springerlink.com/index/10.1007/978-0-387-71819-4>
- [7] M. Enachescu, S. Cotofana, A. Genderen, D. Tsamados, and A. Ionescu, "Can SG-FET replace FET in sleep mode circuits?" *Nano-Net*, pp. 99–104, 2009.
- [8] M. Enachescu, G. Voicu, and S. Dan Cotofana, "Advanced NEMS-based power management for 3D stacked integrated circuits," in *2010 International Conference on Energy Aware Computing*, Cairo, Egypt, Dec. 2010.
- [9] G. R. Voicu, M. Enachescu, and S. D. Cotofana, "Towards "Zero-energy" using NEMFET-based power management for 3D hybrid stacked ICs," in *2011 IEEE/ACM International Symposium on Nanoscale Architectures*, San Diego, CA, USA, Jun. 2011, pp. 203–209.
- [10] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, "Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a

- metal-over-gate architecture,” in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [11] P. Garrou, *Handbook of 3D integration : technology and applications of 3D integrated circuits*. Weinheim: Wiley-VCH, 2008.
- [12] OpenCores, “openMSP430 softcore,” 2011. [Online]. Available: <http://opencores.org>
- [13] EP Limited, “ECG analysis software,” 2011. [Online]. Available: <http://www.eplimited.com/software.htm>
- [14] J. Pan and W. J. Tompkins, “A real-time QRS detection algorithm,” *IEEE Transactions on Bio-Medical Engineering*, vol. 32, no. 3, pp. 230–236, Mar. 1985, PMID: 3997178.
- [15] “Cadence design systems,” 2011. [Online]. Available: <http://www.cadence.com/us/pages/default.aspx>

6

Ultra Low Power NEMFET Based Logic¹

Abstract: In this paper, we introduce a Nano-Electro-Mechanical Field Effect Transistor (NEMFET) based logic family tailored to the implementation of low speed and ultra low energy functional units and processors. Basic Boolean gates implemented with NEMFETs only are analysed and compared against equivalent CMOS realisations. Our simulations suggest that the proposed short-circuit current free NEMFET gates exhibit up to 10x dynamic energy reduction and up to 2 orders of magnitude less leakage, at the expense of 10 to 20x slower operation, when compared with CMOS counterparts. We also analyse the fan-in influence on gate performance and observe that NEMFET the gate energy advantage increases with fan-in. Finally, we consider a 3D-Stacked hybrid NEMFET-CMOS computation platform running a heart-beat rate monitor application and demonstrate that NEMFET based logic is an enabling factor for the implementation of "zero-energy" operated systems.

Copyright © 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purpose must be obtained from IEEE by sending an email to *pubs – permissions@ieee.org*.

¹This chapter is based on the research article published as "Ultra Low Power NEMFET Based Logic" by M. Enachescu, M. Lefter, A. Bazigos, A. M. Ionescu, and S. Cotozana, *IEEE International Symposium on Circuits and Systems (ISCAS 2013)*, pp. 566 - 569, Beijing, China, May, 2013.

6.1 Introduction

While CMOS scaling substantially diminishes device delay the scalability frontier of the threshold voltage (V_T) limits the reduction of the supply voltage; thus, the energy consumption does not gracefully scale with feature size reduction. Moreover, feature size technology scaling implies high power density and leverages an abrupt leakage increase [1]. While for general purpose and high performance applications the throughput and/or latency is the most important figure of merit for many embedded applications, e.g., wireless sensor networks [2], the capability to operate without a battery change until they become obsolete is of premium importance. Such applications have a high idle rate and relaxed time constraints thus, to meet such a "zero-energy" operation requirement, they should be implemented with low speed and extremely low power technologies. Current CMOS technologies are not the best fit for their implementation as they even have large leakage and/or require high power supply thus tight "zero-energy" budgets cannot be satisfied.

With the advent of Nano-Electro-Mechanical Systems (NEMS) novel devices, e.g., NEM Field Effect Transistor (NEMFET) [3], have been proposed with ultra-low leakage currents, low active energy due to abrupt switching, and hysteresis. Such devices have been initially meant to be utilized in sensing [4]. However, their utilization as sleep transistors in power management schemes has been also investigated [5].

In this paper, we introduce a NEMFET based logic family tailored to the implementation of ultra-low energy functional units and processors. We introduce basic Boolean gates implemented with NEMFETs only, analyse their performance, and compare with CMOS counterparts. In our proposal, we make use of the NEMFET hysteresis and introduce short-circuit current free designs. Thus, the NEMFET gates do not experience short circuit path during switching, and that energy component is eliminated. Our simulations suggest that when compared with CMOS gates implemented in a technology with similar transistor size the NEMFET gates are between 10 to 20x slower but consume up to 10x less energy and have up to 2 orders of magnitude less leakage. We also perform an analysis of the fan-in influence on gate performance, which suggests that NEMFET gates are even more advantageous in terms of energy reduction as the number of inputs increases. We equipped the Trough Silicon Via (TSV) based 3D-Stacked hybrid NEMFET-CMOS computation platform in [5] with NEMFET implemented power management. This results in a further 5% energy reduction for the heart beat monitoring and 21% for

application with 10x lower activities rate [6]. Due to this reduction the size of the vibration energy generator needed to power the SoC in a "zero-energy" regime decreases from 2.15cm^2 to 2.03cm^2 and 1.69cm^2 , respectively. This clearly indicates that NEMFET-based logic is an enabling factor for the implementation of "zero-energy" operated systems.

The rest of the paper is organised as follows. In Section 7.3.2, we provide a brief NEMFET overview and introduce a novel, compact model that allows us to simulate basic NEMFET based logic gates. Section 6.3 introduces the proposed NEMFET based logic family, analyses its performance, and positions it versus CMOS implementations. In Section 7.5, we make use of the NEMFET logic for the implementation of the power management circuitry and we evaluate this implication by considering a 3D-Stacked hybrid NEMFET-CMOS computation platform running a heartbeat rate monitor application as a case study and evaluate its potential "zero-energy" operation. Finally, Section 7.6 presents some conclusions.

6.2 NEMFET Background and Compact Modeling

The Nano-Electro-Mechanical Field Effect Transistor (NEMFET) has a 3D geometry and cross-section as presented in Figure 6.1. It has ultra-low leakage currents, low active energy due to abrupt switching, and hysteresis.

Essentially speaking, the device behaves like an electromechanical switch which responds to gate bias changes as follows. When the gate voltage V_G is low, the gate-oxide capacitance is in series with the air-gap capacitance resulting in low electrostatic coupling of the gate to the channel, thus in a negligible drain current I_D . If V_G increases the situation remains unchanged until it reaches the pull-in voltage V_{PI} in which case the electrostatic force cannot be compensated anymore by the mechanical restoring force and the suspended gate (beam) snaps onto the gate oxide, thus turning on the device. After the pull-in, the I_D increase with V_G is comparable with the one of a standard MOSFET. On the other way around when V_G is decreased from some high value I_D starts decreasing until at a certain V_G value when the system becomes unstable due to combined electro-mechanical force and the beam is pulled-out. This causes an abrupt I_D decrease due to a large decrease in capacitance, i.e., the pull-out effect.

In our implementations, we make use of Fully Depleted Silicon on Insulator (FD SOI) NEMFETs and in order to analyse the performance and potential of

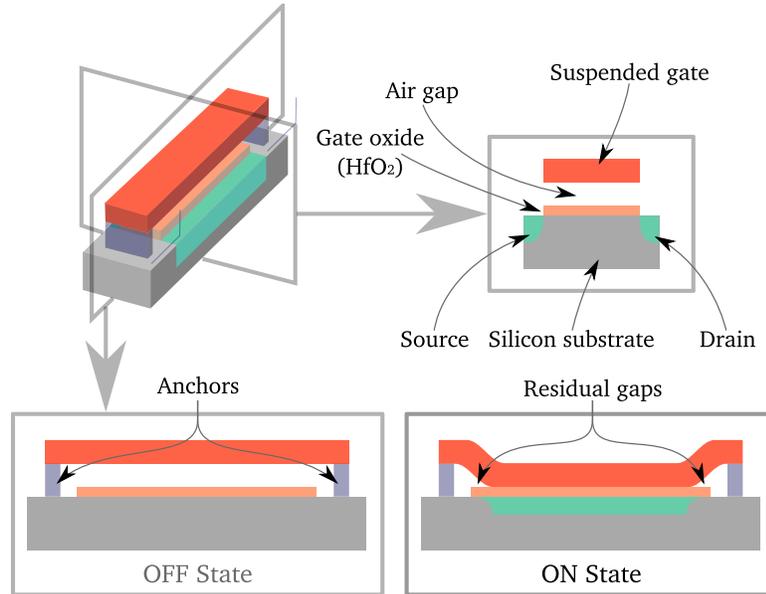


Figure 6.1: Illustrative cross-section of NEMFET. The two states, e.g., pull-out (OFF) and pull-in (ON), are depicted

our proposal we rely in circuit simulations. Thus, compact models are required to accurately capture the complex NEMFET behavior.

The model developed for the NEMFET consists of two main elements: (i) The BSIM-IMG FD SOI model [7, 8] is utilized for predicting the electrical behavior of the transistor; (ii) A Verilog-A-based compact model was developed to estimate the position of the suspended gate. Moreover, the variation of the capacitor in series with the gate was taken into consideration. As the nature of the device extends both in electrical and mechanical regimes, both such types of Verilog-A natures were addressed in the model. An in-depth description of the model can be found in [9].

The parameters of the model have been extracted after hybrid numerical simulations that combined multi-physics and semiconductor analyses [10, 11]. At the first step, the parameters of the BSIM-IMG FD SOI model were extracted, considering the NEMFET with a constantly pulled-in gate-beam. Subsequently, the parameters for the electromechanical addition of the suspended-gate model were as well extracted. In order to adapt the model to the TCAD results, further optimizations were required, i.e., adapting the initial nominal values for our parameters of interest. Since the model is written in Verilog-A it is portable in a wide variety of simulators. Therefore, two design-kits have

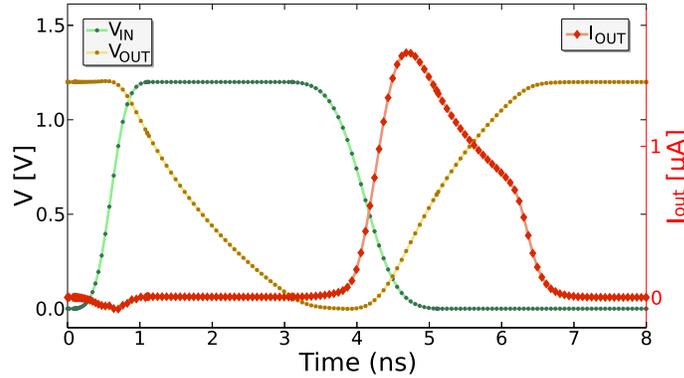


Figure 6.2: NEMFET inverter transfer characteristics

been implemented that easily introduce the NEMFET model to the Agilent ADS and the Cadence Virtuoso software platforms.

6.3 Short Circuit Free NEMFET-based Logic

This section provides an answer to the following question: "To which extent can CMOS alike NEMFET based logic supersede traditional logic?" Hence, we introduce NEMFET based Boolean gates tailored to the implementation of ultra low energy functional units and processors, analyze and compare their performance with CMOS counterparts. Finally, we address the following issues related to NEMFET-based logic: (i) energy efficiency against the "classic" CMOS logic, and (ii) scaling for improved performance.

By wiring two NEMFETs together in a traditional manner, we form an inverter circuit. The dynamic inverter operation at a clock frequency of 125 MHz is depicted in Figure 7.3, for a square-wave input signal supplied by a waveform generator. The NEMFET hysteresis allows for n/p NEMFET channel sizing that precludes the situation when the two NEMFETs are simultaneously open during the switching. In this way the most important power component, i.e., the power consumption given by the Short Circuit (SC) current, is diminished.

Figure 6.3 depicts the SC current analysis of an NEMFET inverter without and with hysteresis. First, if no proper attention is given to NEMFET sizing, the inverter behaves like a normal CMOS gate (left side of Figure 6.3). However, by proper dimensioning of the n/p-NEMFETs (Figure 6.3 right side) we can adjust the occurrence of the Pull-In (PI) and Pull-Out (PO) events as follows:

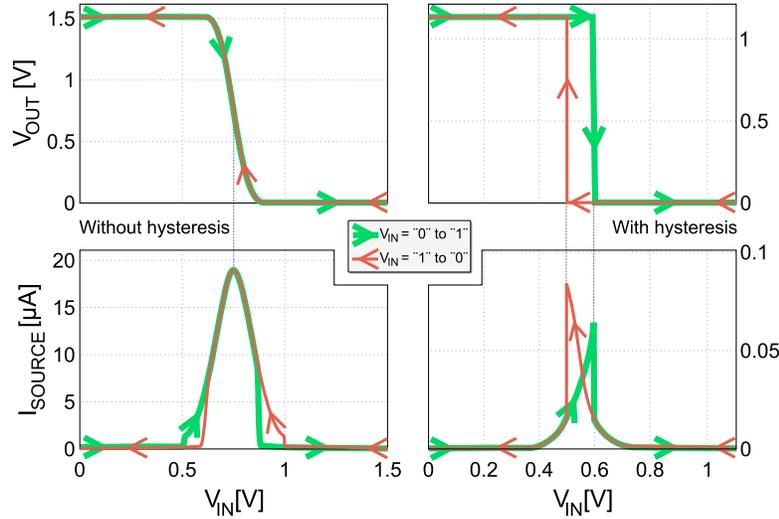


Figure 6.3: NEMFET inverter SC current analysis

(i) the nNEMFET PI takes place after the pNEMFET PO; (ii) the pNEMFET PI takes place before the nNEMFET PO. In this way, we arrange the switching moments such that the two transistors are never conducting at the same time, which is practically cancelling the short circuit current and results in 2 order of magnitude reduction of the energy consumption per switching event.

We can build upon the Inverter and construct basic Boolean gates operating according the same SC free paradigm. The voltage transfer functions for two-input (NAND, NOR) gates are depicted in Figure 6.4, for a 100 MHz clock period. Thus, the NEMFET gates do not experience the creation of a short circuit path during switching, and that energy component is eliminated.

Table 6.1 presents the delay and the static and dynamic energy parameters for various (INV, NAND, NOR) gates implemented with NEMFETs and in a commercial CMOS 65nm Low Power (LP) technology. The figures of merit of the CMOS High- V_T logic gates were obtained using Cadence Spectre while Agilent ADS was utilised for the NEMFET-based logic.

One can observe in the table that NEMFET gates are between 10x to 20x slower but consume up to 10x less energy and have up to 2 orders of magnitude less leakage. Having these results in mind, we can conclude that the NEMFET logic application field is rather limited for the time being to ultra low energy, low frequency, low activity designs, e.g., wireless sensor node running a heart beat rate application [5]. However, since it has been already proven that the

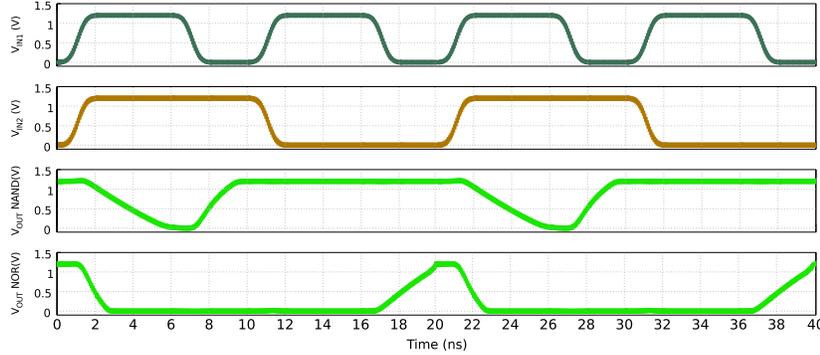


Figure 6.4: Transfer characteristics for NEMFET NOR/NAND

Table 6.1: CMOS vs. NEMFET gates

	NAND		NOR		INV	
	C	N	C	N	C	N
E dynamic [fJ]	3.27	0.65	2.72	0.57	1.33	0.41
P leakage [pW]	11.72	0.16	13.44	0.16	8.21	0.08
Delay [ns]	0.04	0.80	0.05	0.78	0.03	0.31

C, N stands for CMOS and NEMFET, respectively.

NEMFET beam could scale at least up $260nm \times 65nm$ [12] and the scaling is predicted to continue in the following years [13], we expect the NEMFET delay to decrease thus to make NEMFET logic more competitive.

We also perform an analysis of the fan-in influence on gate performance which results are presented in Table 6.2. The Table suggests that NEMFET gates are even more advantageous in terms of energy reduction as the number of inputs increases. For example while a NAND2 CMOS consumes 5x more dynamic energy than a NAND2 NEMFET this factor becomes 10x for the NAND4 case.

6.4 NEMFET-based Power Management Logic

State of the art ultra low power designs employ specific Power Management (PM) circuitry to turn off inactive units in order to achieve energy savings [14]. While PM is substantially reducing the idle parts leakage by separating them

Table 6.2: CMOS vs. NEMFET - variable fan-in NAND

	NAND2		NAND3		NAND4	
	C	N	C	N	C	N
E dynamic [fJ]	3.27	0.65	6.90	0.96	11.20	1.24
P leakage [pW]	11.72	0.16	13.28	0.22	14.42	0.28
Delay [ns]	0.04	0.80	0.06	1.20	0.09	1.55

C, N stands for CMOS and NEMFET, respectively.

from the supply voltage by means of High- V_T MOSFET or NEMFET [15] based Sleep Transistors (STs), it has been noticed in [15] that the always-active PM circuitry significantly contributes to the total energy consumption of low activity rate applications. Thus, in this section, we make use of the proposed SC-free NEMFET logic for the implementation of the PM circuitry, i.e., Isolation Cells (ISO), State Retention (SR) cells, and PM Controller (PMC). Additionally, to evaluate the impact of utilising SC-free NEMFET based logic for the PM circuitry implementation, we consider, as a case study, a 3D Through-Silicon Via (TSV) based hybrid NEMFET-CMOS computation platform running a heartbeat rate monitor application.

Next, we describe how the basic NEMFET logic gates can be utilised to implement the PM circuitry in a heterogeneous 3D stacked power management architecture composed out of a NEMFET and a CMOS tier.

Isolation Cells

Isolation cells (ISO) control the floating output of powered down blocks and clamp the outputs to a specific, legal value. Thus, they prevent short circuit current generation from the power gated block floating connections. NEMFET-based logic gates are employed for Retain 1/Retain 0 isolation cells as depicted in Figure 6.5(a), and Figure 6.5(b), respectively. An ISO group part of the same power domain can share the isolation control signal. Every isolation cell requires 2 TSVs, one for the input signal coming from the powered-down block, and the other for the output to be clamped. The isolation control signals are generated by the PMC and since it is implemented on the same NEMS die as the isolation cells, no TSVs are needed for the control signals. Hence, the total cost in terms of number of TSVs is given by $N_{TSVsISO} = 2N_{ISOCELLS}$, where $N_{ISOCELLS}$ is the total number of isolation cells.

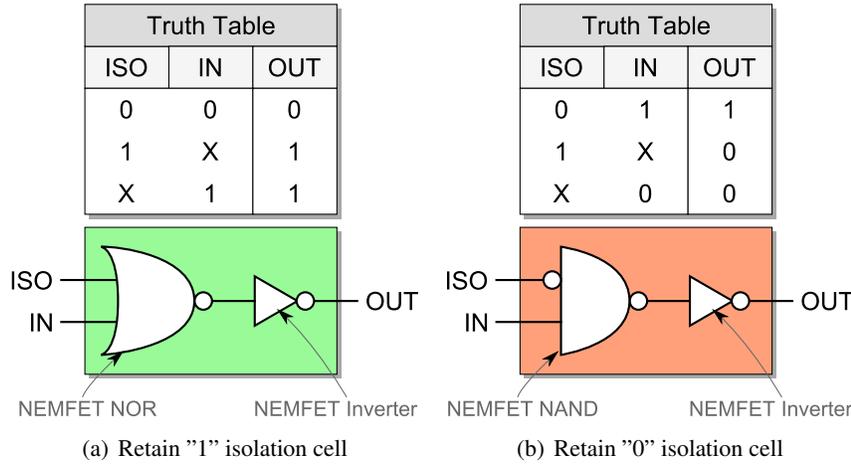


Figure 6.5: NEMFET-based isolation cells in commercial designs and associated truth table

State Retention

The storage capacity of the NEMFET inverter makes it suitable also for state retention schemes [16]. Thus, we propose a NEMFET-based state retention cell which preserves the state of a register before a module shuts down, with the hybrid NEMFET-CMOS structure depicted in Figure 6.6 and operating according with the signal waveforms from Figure 6.7. A specialised dual-function NEMFET inverter with the input connected to the to-be-saved register acts as a classic inverter while the gated block is powered on, and as a memory cell when the power is switched off. A multiplexer selects the value (normal operation or saved) to be written in the flip-flop register. The multiplexer and the flip-flop register, active during normal operation of the power gated block, are implemented on the CMOS tier. In contrast with the ISO case, the NEMFET part of the SR circuit together with the inter-die TSV links are inactive during normal powered operation. Hence, the scheme does not incur any delay penalty that might have detrimental influence on the clock period. Every state retention cell needs 2 TSVs, one for the saved signal value, and the other one for the restored value. The TSV required for the control of the specialised inverter can be shared across all the state retention cells that commute in the same time. The total number of TSVs needed for the SR cells is given by $N_{TSVsSR} = 2N_{SRCELLS} + N_{PD}$, where $N_{SRCELLS}$ is the total number of retention bits, and N_{PD} is the number of power domains.

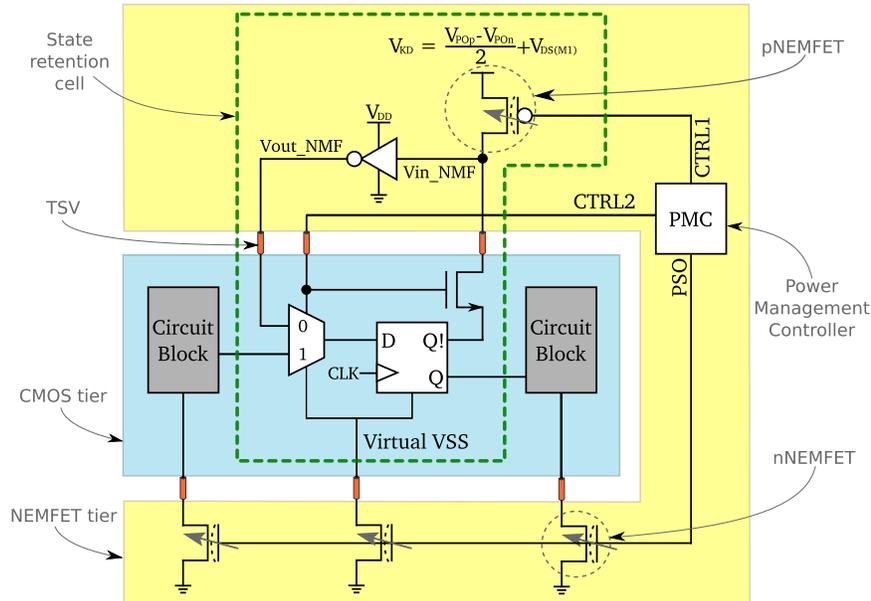


Figure 6.6: Heterogeneous state retention cell

Power Management Controller

The PMC is typically a finite state machine that controls the sequence of power-down and power-up events. Its functionality is highly dependent on the actual design implementation and the power modes specifications. The amount of communication between PMC and the design depends on the transition rules between power modes. At least a TSV is required to indicate the power state of each power domain. The delay penalty of the NEMFET logic is usually tolerable if the switching between different power modes is done not very frequently. Thus, the controller can run at a reduced clock rate without impeding the overall application performance. This is the case for the considered application presented in our example below.

6.4.1 Case Study

To get a more accurate insight on the "zero-energy" potential of the NEMFET-based power management circuitry, we consider a heart beat rate monitoring application running on a typical SoC for low power embedded devices [5]. The SoC consists of a 16-bit processor core, with program and data memory, peripherals, and PM circuitry implemented with either CMOS or NEMFET-

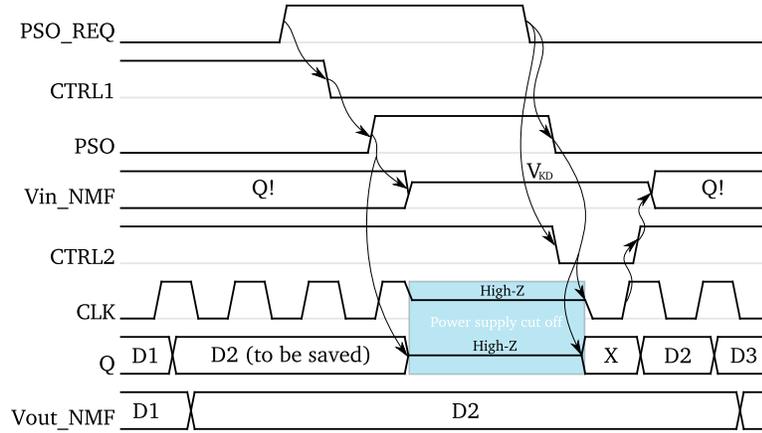


Figure 6.7: State retention signal wave forms

based logic.

For this platform, the PM circuitry is composed of isolation cells and PMC only as no SR cells are being present due to the fact that the system state is software-saved in the always-on data memory before the processor core is powered down. The PM circuitry consists of 74 isolation cells and 22 gates for the PMC.

The power consumption values obtained after simulation in best-case technology corner for CMOS, with respect to power consumption, and after simulation with the NEMFET model in Section 7.3.2, for both the ISO cells and PMC, are presented in Table 8.2. We can observe from Table 8.2 that while the active power consumption during operation and during idle is halved when we switch from CMOS to NEMFETs technology, for both ISO cells and PMC, the leakage power is diminished by about 2 orders of magnitude.

We note that for ISO cell implementation we relied on complex gates (see Figure 6.5(a)) for the CMOS case, while for the NEMFET case we cascaded simple gates, which resulted in a larger number of NEMFET gates even though the number of transistors is the same in both implementations. Although a proper area comparison between NEMFET and CMOS-based gates is difficult because of the differences in gate geometries, preliminary results indicate that a NEMFET transistor is typically four times larger than its CMOS counterpart for the same maximum I_{ON} current value. Thus our approach results in an area overhead on the NEMFET tier and an CMOS tier are reduction. Given that in 3D-Stacked technology the design footprint is the relevant area metric, our approach may even result in a footprint reduction. For the time being, NEM-

FET circuits operating frequency is limited to about 125MHz, which may be high enough given that for most application the PM circuitry operates at low frequency (the entire power management philosophy relies on the assumption that parts of the system stay idle for long periods of time). In case this is not the case, the utilisation of NEMFET based PM circuitry may result in some extra cycles, but with no fundamental implications on the application performance.

Table 6.3: PM Circuitry Power Consumption

	No. Gates	SoC ON	SoC OFF	
		$P_{dynamic}$ [uW]	$Leakage$ [nW]	$P_{dynamic}$ [nW]
ISO CMOS	74	15.63	195	136
ISO NEMFET	158	7.69	2.11	66
PMC CMOS	22	217.73	49.81	1.79
PMC NEMFET	22	107.12	0.58	0.88

Our simulations indicate that if we replace the CMOS PM circuitry of the System-on-Chip presented in [15] with NEMFET one, we can further reduce the energy budget by 5%. We note, however that for an application that has an activity factor 10x lower [6] than the proposed heart rate monitor the energy savings increases to 21%. Due to this reduction, the size of the vibration energy generator needed to power the SoC in a zero-energy regime decreases from 2.15cm^2 to 2.03cm^2 and 1.69cm^2 , respectively.

6.5 Conclusions

In this paper, we introduced a NEMFET based logic family tailored to the implementation of ultra low energy functional units and processors. Basic Boolean gates implemented with NEMFETs only were analyzed and compared with CMOS counterparts implemented in a technology with similar transistor size. Our simulations indicate that the proposed NEMFET based gates consume up to 10x less energy and up to 2 orders of magnitude less leakage, at the expense of up to 20x delay increase. Moreover, for larger fan-in gates, the energy consumption benefit is increasing. Finally, we considered a 3D-Stacked hybrid NEMFET-CMOS architecture running a heartbeat rate monitor application and demonstrated that NEMFET based logic is an enabling factor for the implementation of “zero-energy” operated systems.

Bibliography

- [1] S. Mukhopadhyay, H. Mahmoodi-Meimand, C. Neau, and K. Roy, "Leakage in nanometer scale cmos circuits," in *VLSI Technology, Systems, and Applications, 2003 International Symposium on*, 2003, pp. 307 – 312.
- [2] R. Min, M. Bhardwaj, S.-H. Cho, E. Shih, A. Sinha, A. Wang, and A. Chandrakasan, "Low-power wireless sensor networks," in *VLSI Design, 2001. Fourteenth International Conference on*, 2001, pp. 205–210.
- [3] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, "Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture," in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [4] X. M. H. Huang, M. Manolidis, S. C. Jun, and J. Hone, "Nanomechanical hydrogen sensing," *Applied Physics Letters*, vol. 86, no. 14, pp. 143 104 –143 104–3, apr 2005.
- [5] M. Enachescu, G. Voicu, and S. Cotofana, "Is the road towards "Zero-Energy" paved with NEMFET-based power management?" in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, may 2012, pp. 2561 –2564.
- [6] K. Romer and F. Mattern, "The design space of wireless sensor networks," *Wireless Communications, IEEE*, vol. 11, no. 6, pp. 54 – 61, dec. 2004.
- [7] D. Lu, "Phd dissertation: Compact models for future generation cmos," Ph.D. dissertation, EECS Department, University of California, Berkeley, May 2011. [Online]. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-69.html>
- [8] Prof. Chenming Hu and Prof. Ali Niknejad, "BSIM-IMG: surface potential based UTBSOI MOSFET model," UC Berkeley, 2011. [Online]. Available: http://ekv.epfl.ch/files/content/sites/ekv/files/workshop/2011/Karim_NanoTera_2011.pdf
- [9] M. Enachescu, A. van Genderen, S. D. Cotofana, D. Tsamados, and A. Ionescu, "Report on power management and power evaluations

- based on simulated NEM-FET characteristics,” Department of Software and Computer Technology, TU Delft, NEMSIC Deliverable 2.4.1, 2011. [Online]. Available: http://www.nemsic.org/members/pfn_v2/documents.html
- [10] D. Tsamados, Y. Chauhan, C. Eggimann, K. Akarvardar, H. Wong, and A. Ionescu, “Numerical and analytical simulations of suspended gate - fet for ultra-low power inverters,” in *Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European*, sept. 2007, pp. 167–170.
- [11] D. Tsamados, Y. S. Chauhan, C. Eggimann, K. Akarvardar, H.-S. P. Wong, and A. M. Ionescu, “Finite element analysis and analytical simulations of Suspended Gate-FET for ultra-low power inverters,” *Solid-State Electronics*, 2008.
- [12] S. Chong, K. Akarvardar, R. Parsa, J.-B. Yoon, R. T. Howe, S. Mitra, and H.-S. P. Wong, “Nanoelectromechanical (nem) relays integrated with cmos sram for improved stability and low leakage,” in *Proceedings of the 2009 International Conference on Computer-Aided Design*, ser. ICCAD '09. New York, NY, USA: ACM, 2009, pp. 478–484.
- [13] D. Tsamados, A. Ionescu, K. Akarvardar, H.-S. Philip Wong, E. Alon, and T.-J. King Liu, “ITRS - Nano-Electro-Mechanical Switches,” in *ITRS - Nano-Electro-Mechanical Switches*, September 2008.
- [14] N. H. E. Weste and D. M. Harris, *CMOS VLSI Design A Circuits and Systems Perspective*. Addison Wesley, 2011.
- [15] G. R. Voicu, M. Enachescu, and S. D. Cotofana, “Towards ”Zero-energy” using NEMFET-based power management for 3D hybrid stacked ICs,” in *2011 IEEE/ACM International Symposium on Nanoscale Architectures*, San Diego, CA, USA, Jun. 2011, pp. 203–209.
- [16] K. Akarvardar, C. Eggimann, D. Tsamados, Y. S. Chauhan, G. C. Wan, A. M. Ionescu, R. T. Howe, and H.-S. P. Wong, “Analytical modeling of the suspended-gate FET and design insights for low-power logic,” *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 48–59, Jan. 2008.

7

Low-Leakage 3D Stacked Hybrid NEMFET-CMOS Dual Port Memory¹

In this paper we evaluate a 3D stacked hybrid dual-port memory which combines the appealing Nano-Electro-Mechanical Field Effect Transistor (NEMFET) properties, i.e., ultra-low leakage currents and abrupt switching, with the CMOS technology versatility. The 3D stacked hybrid memory relies on a hysteretic NEMFET inverter to store data, and on adjacent CMOS based logic to allow for read/write operations, and data preservation. In the evaluation we performed a comparison in terms of footprint, access time, and energy, against state of the art CMOS dual-ports memories, considering small and large size memories implemented in various technology nodes. The 3D NEMFET-CMOS hybrid dual-port memory is on the average -25% , 8% , and 95% larger in terms of footprint, when compared to 90nm, 65nm, and 45nm CMOS implementations. The write access time is approximately $2\times$ higher, as it is dominated by the mechanical movement of the NEMFET's suspended gate, while the read access time is about 12% lower, when compared with 45nm CMOS counterparts. For small size memories our proposal results in 15% and 23% energy reductions for 100% and 50% data transition probability, respectively. For large size memories an energy reduction of about 40% was obtained, as in this case the static energy is predominant.

¹Parts of this chapter have been published in the research article "Energy Effective 3D Stacked Hybrid NEMFET-CMOS Caches" by M. Lefter, M. Enachescu, G.R. Voicu, and S.D. Cotofana, *ACM/IEEE International Symposium on Nanoscale Architectures (NANOARCH 2014)*, Paris, France, July 2014. The entire chapter is under submission as "Low-Leakage 3D Stacked Hybrid NEMFET-CMOS Dual Port Memory" by M. Enachescu, M. Lefter, G. R. Voicu, and S. Cotofana, to *IEEE Transaction on Emerging Topics in Computing Special Issue on Design and Technology of Integrated Systems in Deep Submicron Era*.

7.1 Introduction

With the number of transistors on a single silicon die crossing the one billion threshold, power dissipation in CMOS technology developed as the most important constraint in VLSI circuit design, overcoming traditional area and delay metrics [1]. Performance-driven technology scaling can only marginally reduce power consumption, since the MOSFET threshold voltage (V_T) scalability frontier limits the power supply voltage reduction. Moreover, with scaling, leakage increases abruptly and becomes a significant component of the overall power consumption [2].

As most area of up to date Systems on Chip (SoC) is utilised for storage, recent studies, e.g., [3], indicate that memory represents the major contributor to the SoC energy consumption, both static and dynamic. Thus, reducing the memory energy consumption directly implies an important overall energy benefit. As a result, memory power reduction techniques were proposed, e.g., forced transistor stacking [4], PMOS-based pull-up networks with domino logic [5], and supply voltage gating [6].

With the advent of emerging nano-technologies, alternative memory arrays have been proposed, which make use of Nano-Electro-Mechanical (NEM) devices [7], e.g., NEM Field Effect Transistor (NEMFET) [8], NEM Relay (NEMR) [9] [10], in conjunction with CMOS devices to substantially reduce their energy consumption. A preliminary study of NEMFET-based 1T DRAM memory cells was presented in [11]. Moreover, features such as electromechanical hysteresis and sticking induced by surface forces make NEM devices attractive for memory applications, e.g., [12, 13], while in [14] the authors replace SRAM's CMOS inverters with NEMFET equivalent counterparts reporting a clear advantage in terms of leakage reduction at the expense of a large area overhead. A less power effective approach, which replaces only the nMOS transistors with NEM relays resulting in a diminished area and delay overhead was presented in [15], while, the assessment of such memory cell for thermal management within a 3D Integrated many-core memory-processor system was presented in [16]. In [17] the authors identify the storage capability of a single NEMFET inverter and open the path towards a memory cell with reduced transistor count and a simple reading scheme.

In this paper, we follow the same line of investigation and propose a volatile low-power dual port (one port for write and one for read) multi-tier 3D stacked hybrid dual port NEMS-CMOS memory cell (3D-HdpMC) that combines the

appealing NEMFET properties, i.e., ultra-low leakage currents, abrupt switching, and hysteresis, with the versatility of the CMOS technology. The proposed memory cell relies: (i) on an energy effective NEMFET based inverter, designed in such a way that no short circuit current can occur during switching, to store the data, and (ii) on adjacent CMOS based logic to allow for read and write operations, and for data preservation. We make use of 3D integration in order to enable the potential co-integration of NEM and CMOS devices, which, for the time being, appears not to be feasible on the same die. We note that our proposal can be implemented with NEM-relays too, but we did not consider them as an option as they need an extra large body bias ($\approx 8V$) and exhibit a substantial ON state resistance increase after 10^4 actuations [9].

While the static energy of the 3D-HdpMC is drastically diminished due to the NEMFET's ultra-low leakage components, e.g., subthreshold and gate leakage, the dynamic energy is also reduced mainly due to the utilisation of only one inverter per memory cell instead of a cross coupled pair, hence, only one bitline is required for write.

To assess the practical implication of this memory cell, we analyzed at system level the potential implications of utilizing the hybrid memory cell in L1 and L2 cache implementation in [18]. However, a thorough investigation at circuit level is required. Hence, we compare memory arrays built with the proposed 3D-HdpMC and with the dual port SRAM memory cell (10T-DPMC) from [19] [20], the state of the art dual port memory cell for low switching activity, e.g., video processing, and low standby power applications. In the comparison, we consider relevant metrics for embedded memory cell design: (i) footprint, (ii) access latency, and (iii) leakage and dynamic energy consumption per access cycle. We implemented our designs as follows: (i) the 3D-HdpMC access logic, and the 10T-DPMC in 90nm, 65nm, and 45nm commercial CMOS technologies, and (ii) the 3D-HdpMC store logic in 90nm, 65nm, and 45nm predictive NEMFET technology.

In terms of footprint, when compared to a 90nm CMOS 10T-DPMC, the proposed hybrid memory utilises a memory cell 40% smaller, while for 65nm and 45nm, the 3D-HdpMC footprint is with 9% and 109% larger, respectively, for 2-die implementations.

The access time, when compared with a state of the art 45nm CMOS based dual port memory, is about 12% lower for read operations, while for write operations it is approximately $2\times$ higher, as it is dominated by the mechanical movement of the NEMFET's suspended gate.

In our energy consumption analysis, we have considered both small and large memories, as their energy consumption is dominated by dynamic and static components, respectively. For small memory sizes, e.g., 8-KB our proposal results in about 24% and 29% write energy reduction, and a read energy reduction of 4.5% and 16%, for 100%, 50% transition probability, respectively. For large memory sizes, e.g., 128-KB, we obtain an energy reduction of about 40%, regardless of the access type, as in this case the static energy is predominant. We have further considered different memory utilisation scenarios as applications may have read or write dominated memory access traces. When compared with 90nm, 65nm, and 45nm 10T-DPMC array implementations, 3D-HdpMC array's total energy reduces with 40%, 37%, 29% for the 99-1 write-read scenario, respectively. When more read operations occur, i.e., 1-99 write-read scenario, 3D-HdpMC array's dynamic energy reduces with 20%, 20%, and 16%, for 90nm, 65nm, and 45nm implementations, respectively.

The rest of the paper is organised as follows. First, in Section 7.2, we present the power dissipation issue, in particular for memories, followed by a brief presentation of NEMFET background and design challenges for the Short Circuit Current Free (SCCF) NEMFET-based inverter as storage element. Section 7.4 introduces the novel 3D stacked hybrid NEMS-CMOS dual port memory cell that builds upon the SCCF NEMFET-based inverter. In Section 7.5 we compare the proposed hybrid memory with state of the art low power dual port CMOS SRAMs. Finally, Section 7.6 presents some conclusions.

7.2 Background

In this section we first address the power dissipation issue and determine the major contributors and their impact on the overall memory energy consumption. Next, we present background information about NEMFET, i.e., its geometry and basic operation, followed by an effective way to eliminate the short circuit current of NEMFET based Boolean logic gates.

7.2.1 SRAM Energy Consumption

The per application normalised total energy consumption of an N -bit memory array addressed with a k -bit data width granularity can be expressed as:

$$E = af \cdot E_{td} + (1 - af) \cdot E_{ts}, \quad (7.1)$$

where E_{td} and E_{ts} are the memory array per access cycle dynamic and static energy consumption, respectively, and af is the activity factor, i.e., the percentage of cycles in which the memory is accessed (either for write or read) from the total number of clock cycles an application executes. The memory array dynamic and static energy consumption per access cycle can be further expressed as:

$$E_{td} = (N - k) \cdot E_s + k \cdot [tp \cdot E_d + (1 - tp) \cdot E_s], \quad (7.2)$$

$$E_{ts} = N \cdot E_s, \quad (7.3)$$

where E_s and E_d are the per access cycle static and the dynamic energy consumption of a memory cell, respectively, and tp the transition probability, i.e., the probability that a new operation writes/reads a different data than the previously accessed data. By substituting Equations (7.2) and (7.3) in Equation (7.1) it results that:

$$E = (N - k \cdot af \cdot tp) \cdot E_s + k \cdot af \cdot tp \cdot E_d. \quad (7.4)$$

It can be noticed from Equation (7.4) that the majority of the cells comprising a memory array are inactive and consume only static energy, i.e., mostly leakage, as N is much greater than k , and that af and tp do not have a major impact on the static energy value (they are both positive sub-unitary values). Nevertheless, af and tp greatly influence the term that comprises the dynamic energy from Equation (7.4). Note that, even though in practice E_d is few orders of magnitude higher than E_s , the memory array total dynamic energy can be substantially reduced due to low af and/or tp values.

7.2.2 NEMFET Background and Basic Operation

The Nano-Electro-Mechanical FET (NEMFET), firstly described in [8], is a rather complex device with a 3D geometry and cross-section as presented in Figure 7.1, where *gate-oxide* is the thickness of the gate oxide, H_{BEAM} is the thickness of the suspended gate, W_{BEAM} is the width of the beam, L_{BEAM} is the length of the beam, and *Air-gap (gap)* is the gap between the oxide and the suspended gate. The gate plate has a fixed position that leaves some air-gap between the gate and the oxide. If we apply a difference of potential between the gate and the oxide, the gate plate, at some point, pulls-in, and touches the oxide. This happens when the transistor is towards inversion. On the other hand, when the device is about to leave inversion towards depletion,

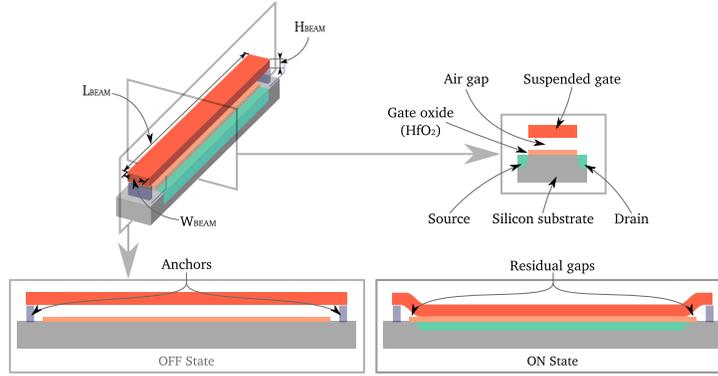


Figure 7.1: NEMFET Suspended-gate Illustrative Cross-section: the Two States, i.e., Pull-out (OFF) and Pull-in (ON), are Depicted

the electrical force gets smaller due to the reduction of the potential difference that generated the force in the first place and, the gate pulls-out to its original in-air position due to the spring force that pulls the gate plate towards its anchors. NEMFET has an extremely low OFF current and exhibits hysteresis as the pull-in and pull-out effects occur at different gate voltage values, denoted further as V_{PI} and V_{PO} , respectively [17].

When designing NEMFET-based logic, the Short Circuit Current (SCC) can be eliminated by carefully sizing the geometry of NEMFET's suspended gate, hence adjusting nNEMFET and pNEMFET hysteresis cycles [21] such that (as illustrated in Figure 7.2): (i) for an inverter with a positive ramping input voltage (V_{IN}) the pull-in voltage of the nNEMFET (V_{PI_n}) should follow the pull-out voltage of the pNEMFET (V_{PO_p}), and (ii) for a negative ramping V_{IN} , the pull-in voltage of the pNEMFET (V_{PI_p}) should follow the pull-out voltage of the nNEMFET (V_{PO_n}). Practically, from the NEMFET inverter transfer characteristic depicted in Figure 7.2 it can be noticed that when the nNEMFET beam is pulled-in, the pNEMFET beam is pulled-out, and vice versa, hence eliminating the short circuit current. On top of that, NEMFET's abrupt switching propriety mitigates also the Subthreshold Leakage (SL), while NEMFET's Gate Leakage (GL) is mitigated in OFF state due to the gap between the suspended gate and the oxide [22].

Given that for standard CMOS based Boolean logic gate implementations the SCC contributes to dynamic energy dissipation, while the SL and GL contribute to the static energy consumption, and that NEMFETs are able to diminish all these three components, we can conclude that they can be an energy

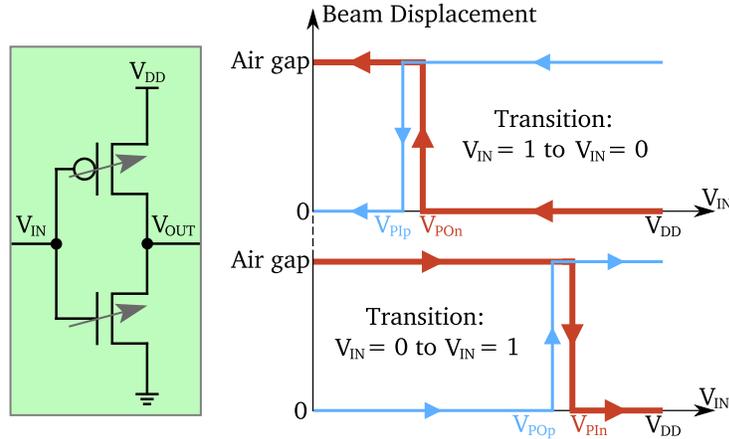


Figure 7.2: NEMFET Inverter Schematic and its Hysteretic Transient Behaviour

effective MOSFETs replacement in CMOS-like Boolean gate implementations.

Figure 7.3 presents the transfer characteristic of an SCC Free (SCCF) NEMFET inverter, obtained by Verilog-A compact model [21] based simulation. We can observe the hysteresis influence on the NEMFET inverter behaviour and current consumption. Practically, NEMFET's inverter current consumption mainly consists on the current required to charge and discharge load and internal capacitances, while SCC, GL, and part of SL have been eliminated [21]. As a consequence the SCCF NEMFET based inverter consumes $3\times$ less dynamic energy and has 2 orders of magnitude less leakage, at the expense of up to $10x$ delay increase, when compared with the CMOS counterpart [21]. Hence, one can conclude that, if the energy consumption is at a premium, the NEMFET inverter can be a very promising candidate for replacing the CMOS cross-coupled storage inverters within the SRAM cell.

To have a better insight on the NEMFET-based inverter energy-efficient design implications, we address in the following section relevant aspects to be considered at design time, particularly for NEMFET based memory cells and arrays.

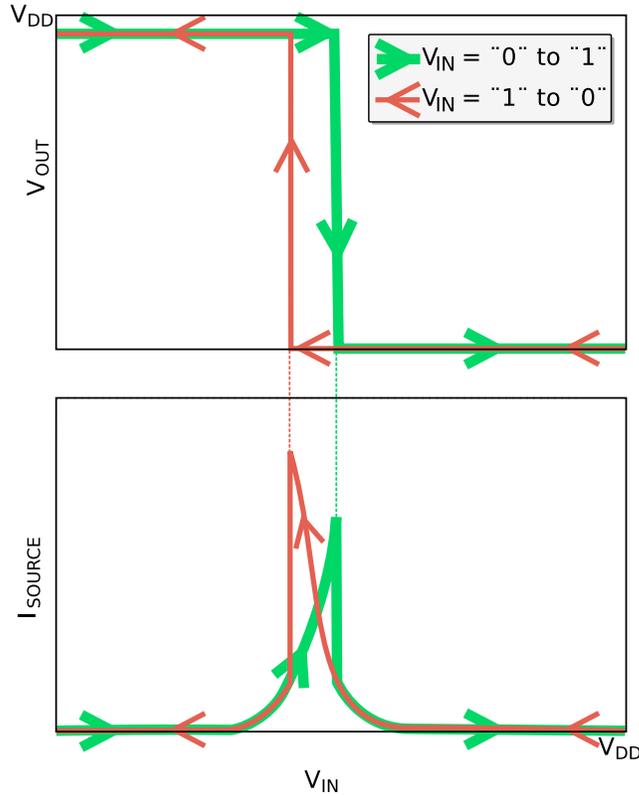


Figure 7.3: NEMFET-based Inverter Transfer Characteristics

7.3 NEMFET Inverter as Storage Structure

Essentially, as also briefly suggested in [17], the NEMFET inverter with its hysteretic behaviour depicted in Figure 7.2, can constitute the kernel of a memory cell. Based on the NEMFET inverter, we introduce the design of a novel memory cell in Section 7.4, but before doing that, we address the following NEMFET-based logic related concerns: (i) NEMFET based inverter noise margin, and (ii) NEMFET scaling and variability.

7.3.1 NEMFET Inverter Stability

The stability of an inverter is given by its *noise margin*, i.e., the maximum noise voltage on the inverter input which is not disturbing its output. The low and high noise margins, i.e., NM_L , and NM_H , respectively, are defined according

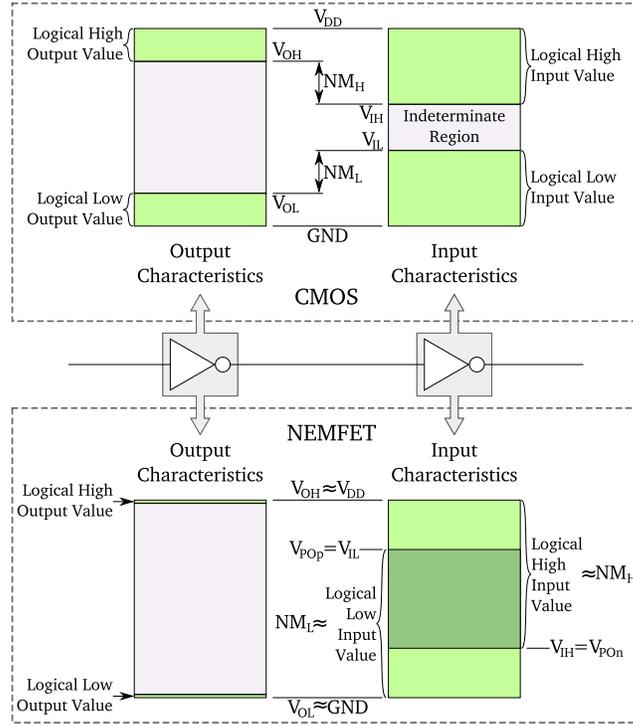


Figure 7.4: CMOS vs. NEMFET Inverter Noise Margin

to [23] as follows:

$$NM_L = V_{IL} - V_{OL} \quad (7.5a)$$

$$NM_H = V_{OH} - V_{IH} \quad (7.5b)$$

where V_{IH} is the minimum HIGH input voltage, V_{IL} is the maximum LOW input voltage, V_{OH} the minimum HIGH output voltage, and V_{OL} the maximum LOW output voltage. Based on the NEMFET-inverter hysteresis behaviour depicted in Figure 7.4, we can observe that the NM_L and NM_H are equal with V_{POp} , and $V_{DD} - V_{POn}$, respectively, since V_{OH} is approximately equal with logic "1", i.e., V_{DD} , and V_{OL} is approximately equal with logic "0", i.e., GND .

When designing the NEMFET-inverter, we can carefully select the geometry of the n/p NEMFET device, such that, the V_{POp} is closer to V_{DD} , and V_{POn} closer to GND , to maximise NM_L , and NM_H , respectively. Moreover, from Figure 7.4 it can also be observed that the NEMFET-inverter *noise margin* is larger than

the one of a typical MOSFET-based inverter. Even for the worst case scenario, when the $V_{POn}=V_{POp}$, this still holds true due to the NEMFET abrupt switching.

7.3.2 NEMFET Inverter Scalability and Variability

In state of the art SRAM architectures, the memory-cell array occupies about 60% of the total footprint [24], thus if NEMFET inverters are utilised instead of CMOS inverters, their area is of major importance. Hence, in this subsection, we present NEMFET’s inverter scalability and variability studies with respect to the design aspects presented above, i.e., the noise margin, and the relative position of the PI and PO events.

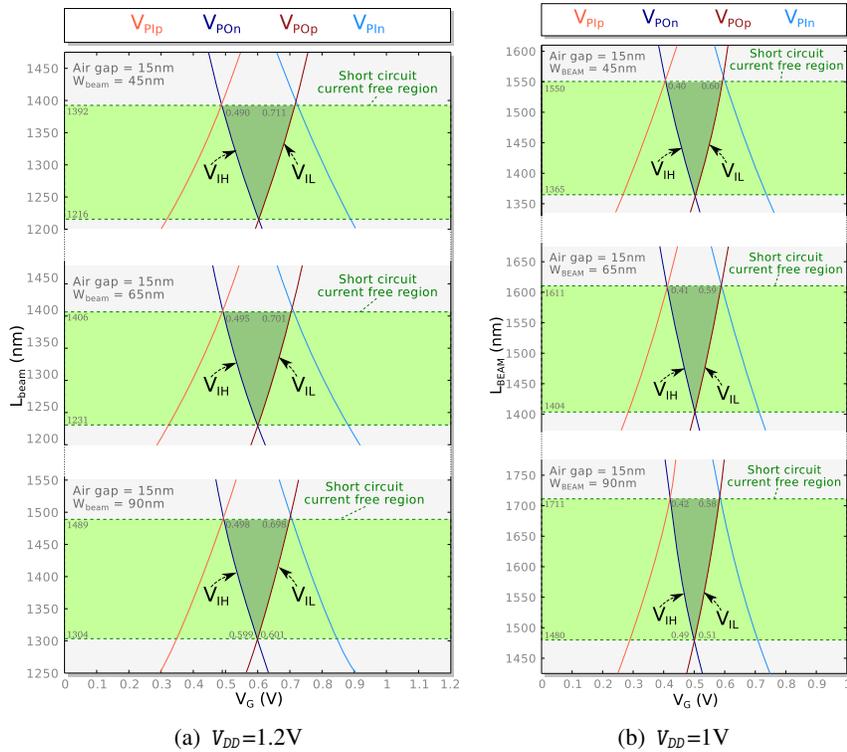


Figure 7.5: NEMFET Inverter Stability Analysis: $W_{BEAM}=45/65/90$ nm, $H_{BEAM}=10$ nm, and $gap=10/15/20$ nm

By varying NEMFET’s $BEAM$ geometry, i.e., L_{BEAM} , W_{BEAM} , the impact of NEMFET-scaling on its NM_H and NM_L is depicted in Figure 7.5(a). We note

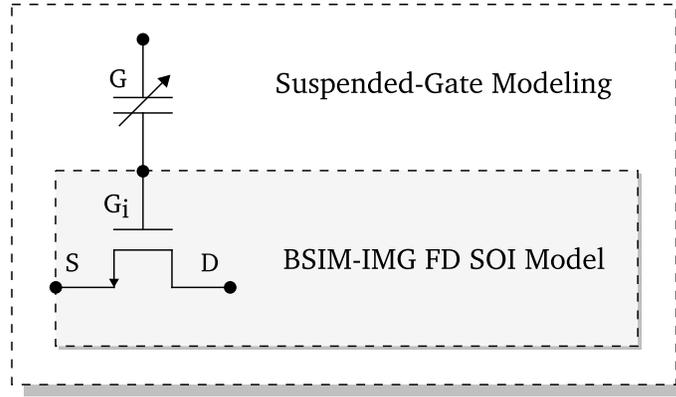


Figure 7.6: A Macro-Model representation of the compact modelling approach for the NEMFET

that the NEMFET Verilog-A compact model from [21] was utilised to simulate the NEMFET inverter in the context of Agilent Advanced Design System (ADS), in *dc* and *transient analysis*. For a given W_{BEAM} , i.e., 90nm, 65nm, and 45nm, a wide variety of L_{BEAM} values are considered to capture NEMFET's inverter Short-Circuit-Current Free (SCCF) operation interval, i.e., V_{PI} and V_{PO} values for which the n/p channel NEMFETs respect the following occurrence order: $V_{PIp} < V_{POn} < V_{POp} < V_{PIn}$.

NEMFET Compact Modeling and Evaluation Platform

The model used for the NEMFET covers both mechanical and electrical properties, as well as the discontinuous nature of the gate-beam pull-in and pull-out events. It consists of two main elements. At the core level, the BSIM-IMG FD SOI model [25,26] predicts the electrical behaviour of the transistor, while the position of the suspended gate has been addressed separately [21]. The model has been implemented in Verilog-A [27,28]. The hierarchical macro-model scheme is presented in Figure 7.6.

The gate-beam geometry and material properties determine the NEMFET mechanical behaviour [17,29]. Considering an orthogonal gate-beam geometry, with length, width, and thickness, L_{BEAM} , W_{BEAM} , and H_{BEAM} , respectively, and

Young's Modulus E , the NEMFET's mechanical properties are given by:

$$\begin{aligned} k &= \frac{32 \cdot E \cdot W_{BEAM} \cdot H_{BEAM}^3}{L_{BEAM}^3}, \\ k_s &= \frac{\pi^4 \cdot E \cdot W_{BEAM} \cdot H_{BEAM}}{L_{BEAM}}, \\ b &= \frac{3 \cdot \mu_v \cdot W_{BEAM}^2 \cdot L_{BEAM}^2}{2 \cdot \pi \cdot gap^3}, \end{aligned} \quad (7.6)$$

where μ_v is the viscosity of the air material between the gate-beam and the gate-oxide.

The electrical attractive force F_{el} and gate-beam spring resistance F_s , i.e., the main forces determining the gate-beam behavior, are statically calculated based on the following equations:

$$\begin{aligned} F_{el}(z, V_{g,gi}) &= \frac{\epsilon \cdot W_{BEAM}^2 \cdot V_{g,gi}^2}{2 \cdot (gap - \bar{z})^2}, \\ F_s(z) &= -(k \cdot \bar{z} + k_s \cdot \bar{z}^3), \end{aligned} \quad (7.7)$$

where z is the vertical displacement of the gate-beam, and \bar{z} is a smoothly clamped version of z within its physical limits. By utilizing \bar{z} , we increase the convergence of the model, and we constrain the solution interval within the physical limits. Hence, \bar{z} and z are identical when the device is in the no-contact state. To coherently model the pull-in and pull-out phenomena, the contact force, F_c was introduced, being proportional with the difference $z - \bar{z}$. For the transient domain, however, the solver has to address the following equation:

$$\sum_{i=el,s,c} F_i(z, V_{g,gi}) - b \frac{dz}{dt} = m \frac{d^2z}{dt^2}, \quad (7.8)$$

where m stands for gate-beam mass. Note that for a static analysis, the derivative terms of (7.8) are null, thus, the problem is simplified into balancing the forces.

It must also be mentioned that the macro-model in Figure 7.6 leaves the internal node of the inner gate (G_i) floating between the capacitances of the suspended gate (G) and the inner transistor. To assist the simulator converging towards a solution, G_i potential is set by NEMFET's representation as a capacitor divider, where the values for the inner and outer capacitances are provided by the simulator. As gate leakage current is negligible, the device DC behaviour is not affected, while the robustness of the compact mode increases.

The model parameters have been extracted after hybrid numerical simulations that combined multi-physics and semiconductor analyses [22, 30]. The BSIM-IMG FD SOI model parameters were extracted, taking into account NEMFET's constantly pulled-in gate-beam, and, consequently, NEMFET's suspended-gate parameters for the electromechanical addition were calculated. As a starting point the nominal parameters values were used, but in order to adapt the model to the TCAD results further optimizations were required. This compact model, written in Verilog-A, it is employed in the rest of this paper, and can be utilized by a wide variety of simulators, e.g., Agilent ADS and Cadence Virtuoso software platforms.

NEMFET Inverter's Evaluation

We can observe from Figure 7.5(a) that: (i) by scaling W_{BEAM} from $90nm$ to $65nm$ and $45nm$, the maximum NM_H and NM_L marginally increase with 0.4% and 1.8%, respectively. (ii) by increasing L_{BEAM} from the minimum to the maximum value for which the SCCF condition is met, i.e., from $1304nm$, $1231nm$, and $1216nm$ to $1489nm$, $1406nm$, and $1392nm$, for L_{BEAM} equal with $90nm$, $65nm$, and $45nm$, respectively, NM_H and NM_L are increasing with about 100mV from $0.5V_{DD}$ to about $0.6V_{DD}$.

Another NEMFET device parameter of interest is the *gap*. Hence, Figure 7.7 depicts NEMFET inverter stability analysis for *gap* values of $10nm$, $15nm$, and $20nm$, this time by focusing only on the minimum W_{BEAM} value, i.e., $45nm$. We can observe that by scaling *gap* from $20nm$ to $15nm$ and $10nm$, the maximum NM_H and NM_L decreases from $0.63V_{DD}$ to $0.59V_{DD}$, and $0.53V_{DD}$, respectively. However, L_{BEAM} decreases as well for the maximum NM_H and NM_L , from $1843nm$ to $1391nm$ and $1041nm$, respectively. Hence, at design time a trade-off between L_{BEAM} and $NM_{H/L}$ can be considered in order to fulfil footprint and/or stability requirements.

When the NEMFET and the "classic" CMOS technologies are integrated within a hybrid platform, as it is the case in our proposal, similar supply voltage for the two technologies is recommended to avoid voltage drops that could generate unwanted leakage currents. Scaling CMOS technology to even sub-threshold V_{DD} has already been proven [31], however, due to high leakage, for commercial $28nm$ low power CMOS technologies, the recommended typical supply voltage is still $1V$ [32]. Hence, the stability implications of scaling the supply voltage from $1.2V$ to $1V$ for $90nm$, $65nm$, and $45nm$ NEMFET technologies are depicted in Figure 7.5(b), which indicates that when the NEM-

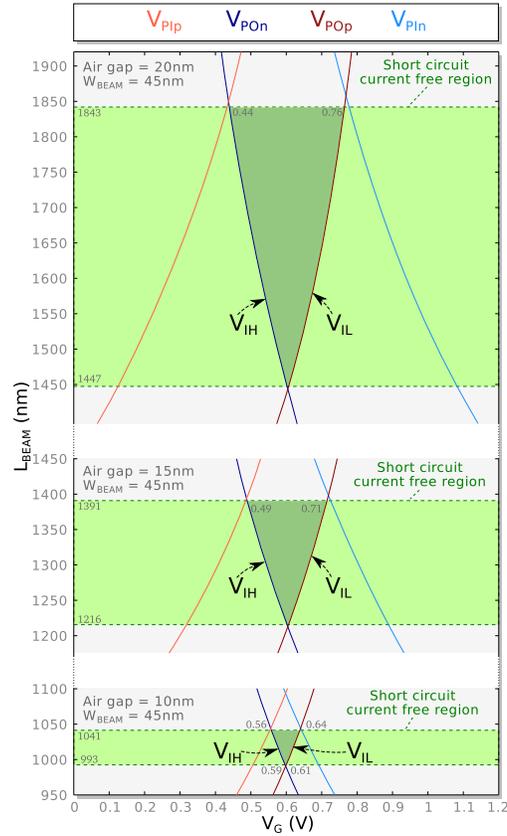


Figure 7.7: NEMFET Inverter Stability Analysis: $V_{DD}=1.2V$, $W_{BEAM}=45nm$, $H_{BEAM}=10nm$, and $gap=10/15/20$ nm

FET inverter’s supply voltage is reduced from 1.2V to 1V, its $NM_{H/L}$ is reduced with about 16% for all considered technology nodes.

For our study in the rest of the paper, to capture potential stability issues, we consider the NEMFET geometries in Table 7.1 corresponding to the average $NM_{H/L}$ within the NEMFET inverter’s short circuit current free operation region. We made use of 1.2V supply voltage for W_{BEAM} of 90nm and 65nm, and of 1V for 45nm. In Table 7.1 we also present the static and dynamic energy consumption of the considered NEMFET inverter geometries and of minimum size inverters implemented in commercial 90nm, 65nm, and 45nm CMOS high-Vt low-power technology nodes.

Table 7.1 indicates that, NEMFET inverters always outperforms CMOS coun-

terparts providing a dynamic energy reduction of 37%, 34%, and 32%, over Low Power (LP) high-Vt CMOS counterparts for 90nm, 65nm, and 45nm technology nodes, respectively. Additionally, the NEMFET implementations have extremely low static power consumption enabling a leakage reduction of $19\times$, $12\times$, and $25\times$, for 90nm, 65nm, and 45nm technology nodes, respectively.

We note that when scaling the NEMFET W_{BEAM} from 90nm to 65nm the power supply value is preserved at 1.2V and the NEMFET inverter dynamic energy consumption is reduced with 12.3%. However, when scaling W_{BEAM} from 65nm to 45nm the supply voltage is reduced to 1V, and in order to account for variability L_{BEAM} cannot be shrank but it has to be increased with about 10%. Hence, in this case, the dynamic energy reduction due to voltage scaling is less significant than expected and due to scaling the NEMFET inverter energy consumption reduction is only 9.6%.

Even if the energy advantages of the NEMFET logic are obvious, the dynamic energy consumption characterisation of both hybrid and classic SRAM cells is not straightforward, being influenced by the load of the NEMFET inverter, i.e., the number of memory cells connected to each bitline. Thus, a system level assessment is presented in Section 7.5.

Table 7.1: NEMFET and CMOS Inverter Static Power and Dynamic Energy Consumption

Design Type	W_{BEAM} / L (nm)	V_{DD} (V)	E_{SW} (aJ)	P_{LEAK} (pW)	L_{BEAM} (nm)	W (nm)
CMOS	90	1.2	1761	9.81	N/A	0.2
NEMFET	90	1.2	1110	0.5	1400	N/A
CMOS	65	1.2	1545	5.67	N/A	0.12
NEMFET	65	1.2	1014	0.47	1320	N/A
CMOS	45	1	1397	11.02	N/A	0.12
NEMFET	45	1	942	0.44	1460	N/A

7.4 3D-Stacked Hybrid NEMFET-CMOS Memory

The schematic diagram of the proposed hybrid NEMS-CMOS memory cell is presented in Figure 7.8, and has the following storage functionality: (i) the to-be-stored value (0/1) is transmitted at the input of the inverter, (ii) the inverter propagates the inverted value (1/0) at its output, and (iii) the inverter retains the output value unchanged as long as its input is kept at a voltage within the interval $[V_{PDP}; V_{PDn}]$ - see also Figure 7.9. We further denote this voltage as V_{KD} . There is a clear and natural separation between the read and the write paths, reflected also in Figure 7.8. The CMOS logic consists of five transistors: four are forming the transmission gates TG_{WA} and TG_{RA} and are utilised for write/read operations and one is utilised for state retention. In the following we detail the write and read operations and present a memory cell noise margin analysis.

Write Operation

In order to write to a memory cell the required value should be present on the write bitline (BL_W). When the write wordline (WL_W) is asserted the transmission gate TG_{WA} opens and the data item reaches the input of the NEMFET-based inverter, which further outputs its complementary value. When WL_W is de-asserted the pMOS transistor T_{KV} keeps the inverter input at V_{KD} such that the inverter output value is maintained unchanged. The write access time is mostly determined by the NEMFET switching time, already addressed in [33], which is rather large as due to mechanical considerations the NEMFET gate requires a certain time interval, i.e., 2.5ns for $W_{BEAM}=45\text{nm}$, $L_{BEAM}=1460\text{nm}$, $H_{BEAM}=10\text{nm}$, $Air-Gap=15\text{nm}$, to pull-in/pull-out when a state change occurs. We note that the reason for utilizing transmission gates instead of simple access transistors is to allow the passing of strong 1 and 0 values, in order to ensure a larger *noise margin*.

Read Operation

In order to read from a memory cell, the read wordline (WL_R) should be asserted. As a result, the transmission gate TG_{RA} opens and frees the stored data item on the read bitline (BL_R). Due to its large dimensions the NEMFET storage inverter behaves as a powerful driver during read, which has a positive impact on the read operation completion time as no extra drivers and sense

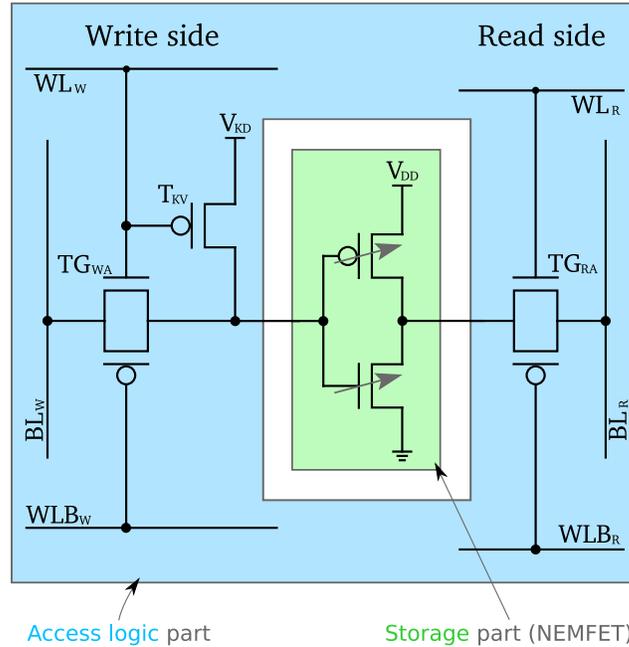


Figure 7.8: Hybrid NEMFET-CMOS Memory Cell Electric Scheme

amplifiers are required.

V_{KD} Noise Margin

The noise margin of the V_{KD} , i.e., hold MM_{VKD} , is determined by $(V_{PDp} - V_{PDn})/2$ as one can observe in Figure 7.10. Thus for an optimal hybrid NEMFET-CMOS memory cell design in terms of *hold noise margin*, i.e., maximum MM_{VKD} , one should choose the geometry of the n and p channel NEMFETs such that: (i) V_{PDp} values are as close to V_{DD} as possible, and (ii) V_{PDn} values are as close to GND as possible. Given the fact that, during data retention, T_{KV} operates in the linear region (V_{KD} is lower than the difference between V_{DD} and T_{KV} 's threshold voltage), the WL_w noise might influence the NEMFET' inverter input. If required, this noise can be reduced/mitigated by: (i) making use of an T_{KV} with a ultra high threshold voltage value, (ii) replacing T_{KV} with a cascade arrangement, or (iii) by utilizing FD SOI CMOS or FinFET technology and dynamically adjusting the T_{KV} threshold voltage value according to variation tolerant on-chip degradation sensors as described in [34].

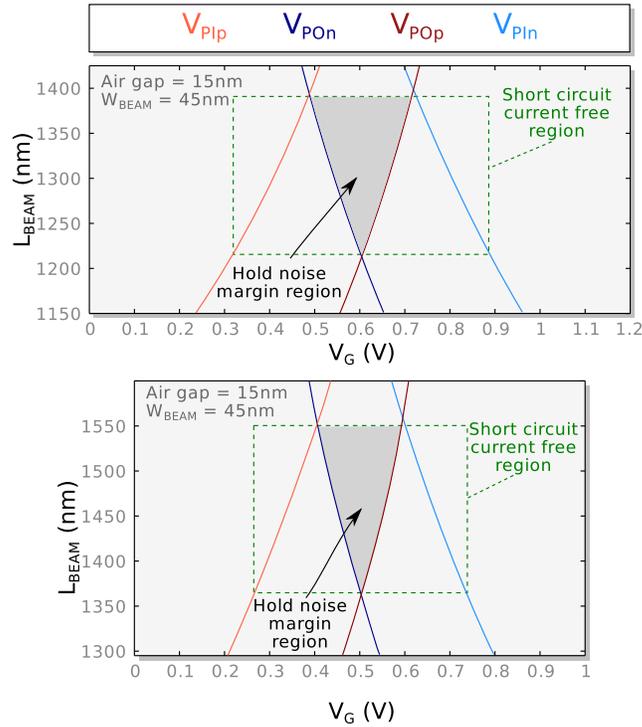


Figure 7.10: NEMS-CMOS Memory Cell Stability Analysis: $V_{DD}=1/1.2V$, Air-gap=15nm, and $W_{BEAM}=45$ nm

This allows for the reduction of the 3D memory chip footprint as row/column circuitry and their associated wires, traditionally placed around the memory array can be accommodated on the area available within the CMOS tier.

7.5 3D Hybrid NEMFET-CMOS Dual Port Memory vs. 2D Dual Port SRAM

In this section we compare the proposed 3D-HdpMC against the most effective, to the best of our knowledge, state of the art low-power SRAM cell [19], with respect to relevant metrics for battery operated SoCs, i.e., energy, delay, and footprint. Moreover, at the memory array level, we investigate the implications of different memory utilisation scenarios, as applications may have read or write dominated memory access traces, on the overall energy consumption.

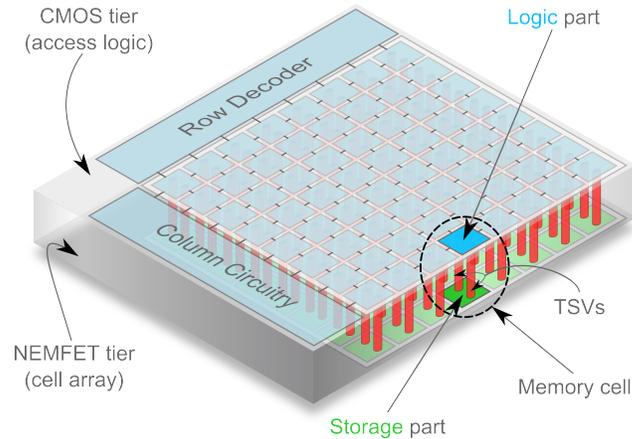


Figure 7.11: Proposed 3D Hybrid NEMFET-CMOS Memory

Taking into consideration that for the proposed 3D-HdpMC only one write bit-line is necessary, we expect to get benefits due to the reduction of: (i) the number of major active energy contributors, i.e., write bitline drivers with their associated bitlines capacitive loads, and (ii) the leakage component associated with the inactive memory array cells. Even though the above theoretical advantages of the 3D-HdpMC are obvious, the evaluation of the actual impact of such a 3D implementation on the overall memory cell array energy consumption, area, and latency, it is of interest as it gives a better insight on the practical implications of our proposal. Thus, further-on, we investigate the efficiency of our proposed 3D hybrid memory, at cell and cell array level, when compared with the 10T non-precharge SRAM cell from [19], depicted in Figure 7.13.

7.5.1 Evaluation Methodology

We use as discussion vehicle the implementation of a dual port (1 for write and 1 for read) memory array, for a wide memory array design space analysis, i.e., from 8-KB up to 128-KB. The CACTI 6.5 memory simulator [35] was utilised to derive the optimal memory partitioning and collect area and write latency information for each of the 45nm, 65nm, and 90nm CMOS implementations. For a more in depth read latency, active energy, and leakage analysis the CMOS-based memory cell was designed using Cadence Virtuoso [36], in a commercial 45nm, 65nm, and 90nm Low Power (LP) Multi-Threshold CMOS (MTCMOS) technologies. *DC* and *transient* simulations were performed us-

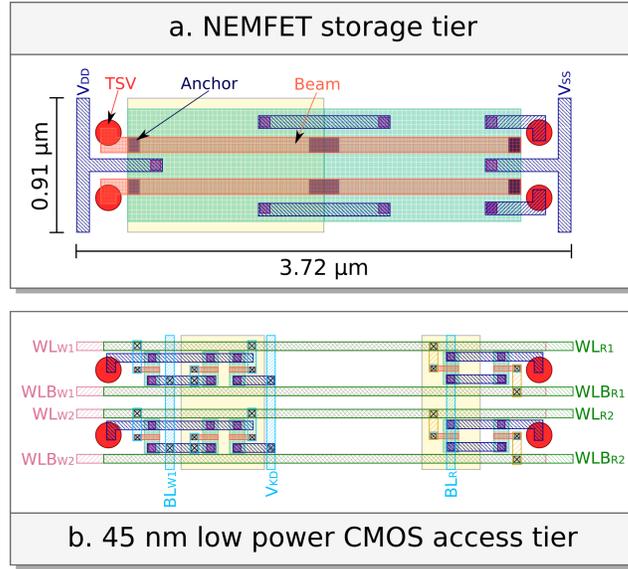


Figure 7.12: Two 3D-HtmpMCs Layout: NEMFET Inverters Tier (top) and 45nm CMOS Access Logic Tier (bottom)

ing the Cadence Spectre simulator [36], and Agilent ADS [37] in conjunction with Cadence Spectre, for the CMOS and for the hybrid NEMFET-CMOS memory cell, respectively. For the hybrid cell, we considered the write bitline driver output signal (having as load the bitline and the NEMFET inverter equivalent capacitances) generated by the Spectre simulator, as input for the NEMFET inverter simulated with Agilent ADS, by means of the NEMFET's Verilog-A compact model from [21]. The TSV contribution was also considered, by means of an RLC model from [38], tailored to a TSV diameter of $0.2\mu\text{m}$ [39]. Moreover, the bitline and wordline RC parasitics, including the bitline coupling capacitance, for each memory array aspect ratio, were extracted from layout, and included in the SPICE simulation. Our simulations were performed in typical case conditions, i.e., typical device models, 1V supply voltage for 45nm, 1.2V supply voltage for 65nm, and 90nm CMOS implementation as well as for all hybrid CMOS-NEMFET implementation, and 27°C .

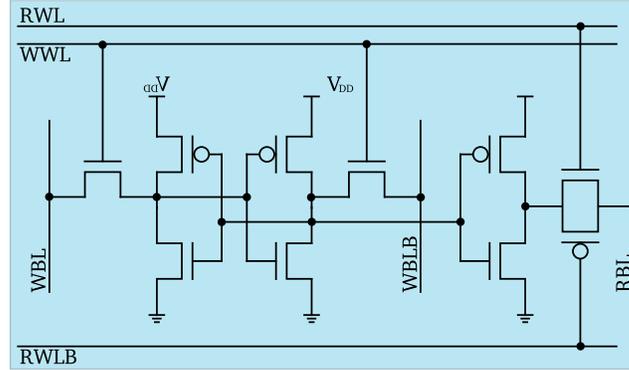


Figure 7.13: Schematic of 10T-DPMC from [19]

7.5.2 Memory Cell

In theory, the footprint of the 3D-HdpMC is determined by $2 \cdot A_{TSV} + \max(2 \cdot A_{T_{NEMFET}}, 5 \cdot A_{T_{MOSFET}})$, where A_{TSV} is the TSV area, and $A_{T_{NEMFET}}$ and $A_{T_{MOSFET}}$ the area of one NEMFET/MOSFET transistor, respectively. However, due to NEMFET's relatively large size, which even if scaled has a footprint larger than that of a MOSFET counterpart (owing this to its mechanical nature - see the layout from Figure 7.12), we can consider $2 \cdot A_{TSV} + 2 \cdot A_{T_{NEMFET}}$ as the 3DHdpMC footprint. The smallest TSV pitch reported in the literature equals $0.4 \mu\text{m}$ [39]. Table 7.2 compares the footprint of hybrid NEMFET with CMOS for dual-port memory cell implementations. When compared to a CMOS 90nm dual port SRAM cell [20], the footprint of a 3D-HdpMC is with 40% smaller, while for 65nm and 45nm [19], the footprint of 3D-HdpMC is with 9% and 109% larger, respectively. The reason behind such a large footprint increase for 45nm is that for the chosen NEMFET geometry to secure an average *noise margin*, L_{BEAM} is not scaling as expected and increases when V_{dd} is scaled from 1.2V to 1V (see also discussion in Section 7.3 as well as Figure 7.5(a) and Figure 7.5(b)).

Table 7.2: Dual-port Memory Cell Footprint

	90nm @ V_{dd} 1.2V Hybrid CMOS	65nm @ V_{dd} 1.2V Hybrid CMOS	45nm @ V_{dd} 1V Hybrid CMOS
Height (μm)	0.5	0.475	0.455
Width (μm)	3.6	3.44	3.72
Footprint (μm^2)	1.8	1.634	1.692
	3.0096	1.5	0.8077

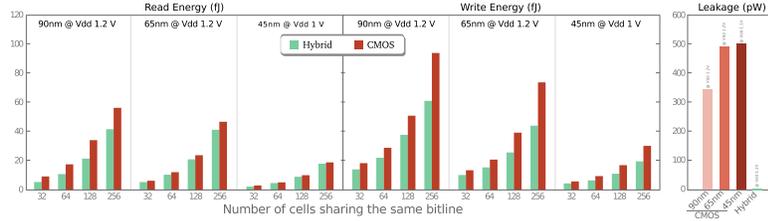


Figure 7.14: 3D-HdpMC vs 10T-DPMC Dynamic Energy and Leakage for Different Loads

Figure 7.14 presents the write/read energy comparison between the 10T-DPMC and the 3D-HdpMC, with respect to four different loads, i.e., 32, 64, 128, and 256 memory cells connected to each bitline (equivalent with the number of rows of a memory array). We considered a worst case scenario for the dynamic energy, with read/write operation data transition probability of 100%, i.e., each write operation changes previously stored data, and each read operation requires a change of the bitline value.

It can be noticed from Figure 7.14 that regardless of technology node or load, 3D-HdpMC outperforms 10T-DPMC counterpart providing a write energy reduction of 23%, 23%, 26%, 33%, of 24%, 30%, 34%, 40%, and of 22%, 30%, 34%, 35%, and read energy reduction of 41%, 38%, 37%, of 26%, 13%, 13.5%, 12%, 12%, and of 11%, 11%, 9%, 5% for 32, 64, 128, 256 memory cells connected to each bitline, and for 90nm, 65nm, and 45nm technology nodes, respectively. Additionally, the 3D-HdpMC implementations have extremely low static power consumption enabling a leakage reduction of 2 orders of magnitude when compared with the 10T-DPMC counterpart, for all considered technology nodes.

We note that when scaling the NEMFET W_{BEAM} from 90nm to 65nm, the bitline capacitance for both 10T-DPMC and 3D-HdpMC decreases accordingly while the power supply value is preserved at 1.2V, hence the NEMFET inverter write and read energy consumption are reduced with about 30% and 25%, respectively. When scaling W_{BEAM} from 65nm to 45nm the supply voltage is reduced to 1V. Hence, the write and read energy consumption energy reduction due to voltage scaling is more significant than expected and the due to scaling 3D-HdpMC inverter energy consumption reduction is about 60%.

7.5.3 Memory Array

Table 7.3 presents the footprint and the area efficiency, i.e., the percentage of the memory total area exclusively populated with memory cells, of an 8-KB, 32-KB, and 128-KB memory arrays implemented using: (i) the proposed NEMS-CMOS approach and (ii) 90nm, 65nm, and 45nm CMOS technologies. The area of the 2-tier 3D-HdpMC array is determined by the area of the NEMFET tier, hence it is always larger than the one of the CMOS implementations. However, at array level, the cell footprint difference from Section 7.5.2, i.e., 9% and 109% larger for 3D-HdpMC when compared to its 65nm and 45nm CMOS counterparts, is reduced to about 8% and about 50%, respectively, independent of the memory array size. This is due to the fact that since the footprint of auxiliary circuitry is equivalent for both 3D-HdpMC and 10T-DPMC implemented in 45nm technology node, the area efficiency for the 3D-HdpMC array is becoming higher when compared with 10T-DPMC (see Table 7.3), i.e., 71.67%, 81.56%, and 89.58% when compared with 65.86%, 74.84%, and 85.39%, for 8-KB, 32-KB, and 128-KB memory arrays, respectively. To completely eliminate the area penalty, one could benefit from the free real estate available within the memory cells and fit the row/column circuitry (see the discussion in Section 7.4 and Figure 7.12a), or it could distribute the NEMFETS on two tiers (one for *p*NEMFETs and one for *n*NEMFETs) a 3-tier embodiment. Moreover, we expect NEM devices scaling (addressed and sustained in [40]) will further alleviate this issue, once the technology becomes more mature. In addition, in [41] the scaling of anchor-free NEMS devices is addressed.

Table 7.3: Memory Footprint and Area Efficiency

	90nm			65nm			45nm					
	Hybrid		CMOS	Hybrid		CMOS	Hybrid		CMOS			
	Footprint (mm ²)	Efficiency (%)	Footprint (mm ²)									
8-KB	0.243	48.41	0.303	64.95	0.178	59.88	0.165	59.54	0.154	71.67	0.080	65.86
32-KB	0.820	57.41	1.091	72.13	0.598	71.45	0.552	71.08	0.542	81.56	0.282	74.84
128-KB	2.565	73.58	3.770	83.70	2.056	83.30	1.894	83.03	1.981	89.58	0.991	85.39

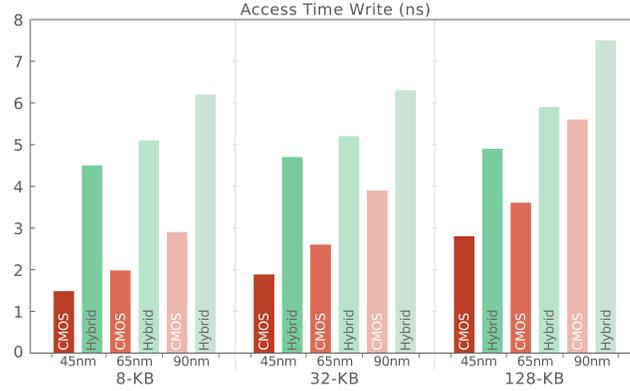


Figure 7.15: 64-bit Word Width Memories Write Access Time

The 3D-HdpMC write operation access time is determined by: (i) the access logic delay (similar to CMOS SRAM as it is implemented in the same technology), (ii) the TSV delay, and (iii) the NEMFET inverter delay. We note that the NEMFET switching time is mostly determined by the mechanical movement of the gate, which is dominant when compared with the RC delay of the "classical" CMOS devices [42], thus limiting its operation to hundred of MHz regime.

To get better insight into the access time comparison 8-KB, 32-KB, and 128-KB memory arrays with a 64-bit data I/O were simulated. Figure 7.15 presents the write access time values for different 10T-DPMC and 3D-HdpMC arrays. Summing up all the above latency components, an 8-KB 3D-HdpMC array requires about 3.3ns, 3.1ns, and 3ns longer write time when compared to its CMOS-based counterpart implemented in 90nm, 65nm, and 45nm technology nodes, respectively.

At a first glance the fact that is not possible to read a value at an address that was just written until the gate stabilisation is not finished may result in performance penalty. However, given that the occurrence of an immediately read after write access to the same memory location is quite improbable in state of the art processors the proposed hybrid memory would actually perform as fast as the SRAM counterpart as there is no need to wait for the write completion before launching the read operation. We note that if such a situation is expected to happen the memory can be extended with a small SRAM buffer to store the most recent written values, and in case an immediately read after write to the same location occurs it can directly provide the requested value.

The size of the buffer is in the order of few words (can be determined broadly by dividing the gate stabilisation time with the desired access time) and the associated logic is simple and has a negligible impact on the energy consumption. Another method to eschew the read after write problem is to rely on software techniques, e.g., compiler optimisations, that eliminate such situations. However, both the buffer and the software methods are outside of the paper's scope. We have studied the implications of a longer write time at the system level in [18].

In order to achieve the similar read and write access time for both the 3D-HdpMC and the 10T-DPMC one could add an additional write port to the 3D-HdpMC. This can be done without any footprint increase as, as previously discussed, enough CMOS tier real estates is unutilised.

The latency to charge and discharge the read bitline capacitance is proportional with its driver's p/n channel transistor L/W ratio. For the 3D-HdpMC, the read bitline driver is the NEMFET inverter, while for the 10T-DPMC it is a CMOS inverter designed with minimum technology rules. The access time values for the read operation within an 8-KB, 32-KB, and 128-KB memory arrays with a 64-bit data I/O are depicted in Figure 7.16. Given a worst case scenario for our proposal, i.e., 8-KB memory array architecture, the read latency for the 3D-HdpMC array is from 12% up to 20% lower when compared with the 10T-DPMC array counterpart implemented in 45nm up to 90nm technology node, respectively. Hence, for read dominated operations, e.g., Conjugate Gradient (CG), and discrete 3D Fast Fourier Transform (3DFFT) [43], the write latency penalty could be compensated by the read latency gain. Hence, for these type of applications, 3D-HdpMC arrays will outperform 10T-DPMC arrays in terms of total computational time.

When analysing the energy consumption of a memory array the following aspects should be considered: (i) the static vs dynamic energy ratio, (ii) the data transition probability, and (iii) the memory activity factor. Figure 7.17 presents the relation between the static and the total energy consumption for various 64-bit word memories. It is obvious that 3D-HdpMC is by far the best choice in all cases, i.e., for all memory array sizes, when leakage is the metric of interest. For 8-KB 10T-DPMC arrays, the leakage represents 3.5%, 4.8%, and 6.8% for write operation, and 7.1%, 11.4%, and 14% for read operation in 90nm, 65nm, and 45nm technology nodes, respectively. For 32-KB 10T-DPMC arrays, the leakage is starting to represent an important component out of the read/write dynamic power, and from 128-KB 10T-DPMC arrays, the



Figure 7.16: 64-bit Word Width Memories Read Access Time

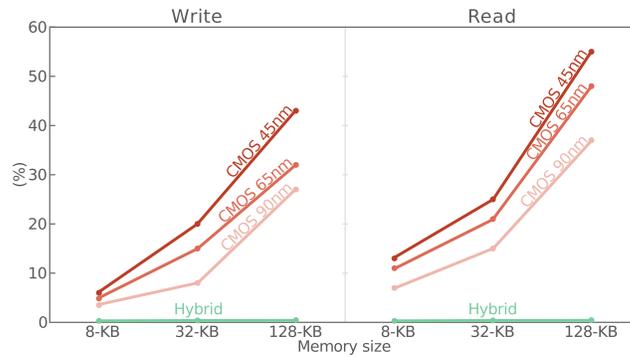


Figure 7.17: Static Energy (Leakage) Contribution to Total Energy Consumption (64-bit Memories)

leakage and the dynamic power start becoming equally important, representing 27%, 32%, and 43% for write operation, and 37%, 48%, and 55% for read operation in 90nm, 65nm, and 45nm technology nodes, respectively. For 3D-HdpMC arrays, however, the leakage, mainly given by the NEMFET's inverter off power, is below 1% of the overall memory energy. Thus, as expected, the larger the memory size, the more beneficial 3D-HdpMC-based implementations are. Moreover, while for CMOS the amount of leakage is increasing with technology scaling, for the 3D-HdpMC remains negligible.

The dynamic energy consumption per write/read access cycle depends on: (i) the changes operated on the read/write bitlines, and (ii) the status of the cross-coupled CMOS inverters for 10T-DPMC, and of the NEMFET inverter for 3D-HdpMC. Figure 7.18 presents the write/read energy consumption per

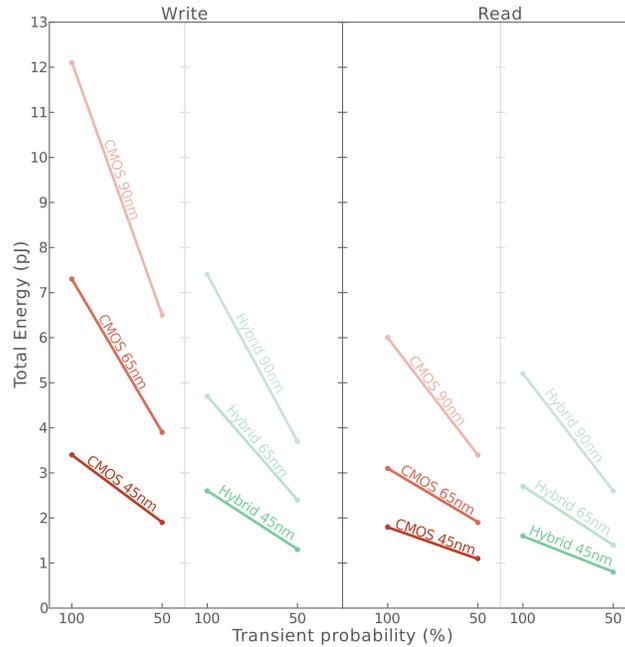


Figure 7.18: Transient Probability Influence on the Total Energy (8-KB Memory Array)

cycle for an 8-KB memory array with a 64-bit data I/O for access traces when 100% (worst case scenario) and 50% (typical case scenario) of the bit values are changing between two consecutive cycles. When compared with the 10T-DPMC array, the 3D-HdpMC array write (read) energy consumption is reduced with 39%, 34%, and 24% (13.5%, 10.7%, and 4.5%) for 90nm, 65nm, and 45nm technology nodes, respectively. For 50% bit values transition probability, the leakage during read or write operation becomes twice more significant, hence the 3D-HdpMC array write (read) energy consumption reduces even more to 40%, 37%, and 29% (20%, 20%, and 16%) for 90nm, 65nm, and 45nm implementations, respectively.

When the total energy consumption is computed according to Equation (7.1) a comparison between the write energy consumption for 3D-HdpMC and the 10T-DPMC based arrays, for various αf values (typically, αf is about 0.2 in current SoCs), is presented in Figure 7.19. The worst case scenario with respect to 3D-HdpMC is when the memory arrays are implemented in 45nm technology node, and when the memory array dissipates low leakage, i.e., 8-KB memory. Figure 7.19 indicates that for small αf values, the total energy

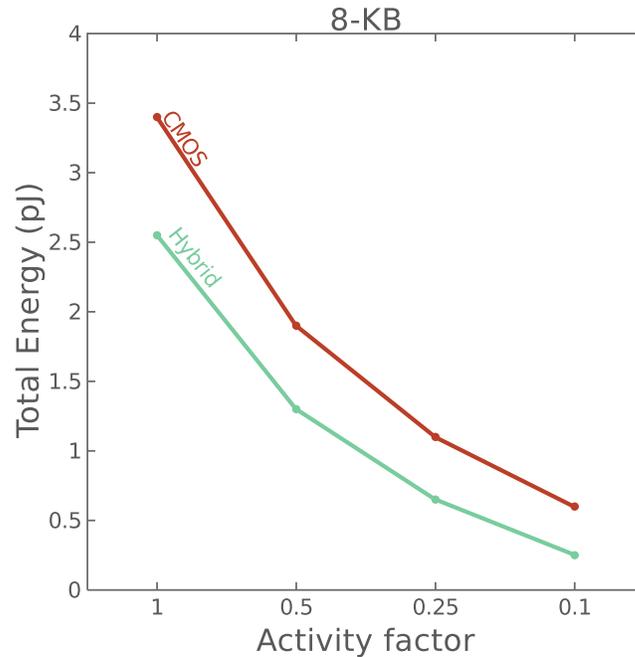


Figure 7.19: Activity Factor Impact on the Total Energy 8-KB Memory Sizes

becomes leakage dominated, hence the 8-KB 3D-HdpMC array energy reduction is increasing from 25% for $af=1$ to 33%, 40%, and 54% for af values of 0.5, 0.25, and 0.1, respectively.

Given that practical applications may have memory access traces with different ratios of read and write operations, being mostly read dominated operations, i.e., CG and 3DFFT, or write dominated, e.g., decompression algorithms, it is of interest to study the impact of the read-write access ratio on the memory overall energy consumption. To cover a large spectrum of practical situations we considered an extensive variety of scenarios, i.e., 99%, 90%, 75%, 50%, 25%, 10%, and 1% write and read operations, respectively. Figure 7.20 presents the energy consumption of 3D-HdpMC and 10T-DPMC 64-bit word 8-KB memory arrays being accessed as previously mentioned.

Overall, it can be noticed from Figure 7.20 that, for both 8-KB memory array implementations, the more reads there are, the lower the total energy is. On the other hand, the more writes there are, the more energy effective the 3D-HdpMC array is. When compared with 90nm, 65nm, and 45nm 10T-DPMC array implementations, 3D-HdpMC array's total energy reduces with

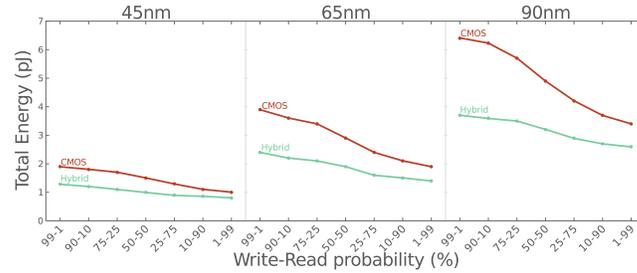


Figure 7.20: Total Energy for Various Write-Read Ratio and 50% Constant Transition Probability (8-KB Memory Array)

40%, 37%, 29% for the 99-1 write-read scenario, respectively. When more read operations occur, i.e., 1-99 write-read scenario, 3D-HdpMC array's dynamic energy reduces with 20%, 20%, and 16%, for 90nm, 65nm, and 45nm implementations, respectively.

7.6 Conclusions

In this paper we proposed a dual port 3D stacked hybrid memory that combines the appealing Nano-Electro-Mechanical Field Effect Transistor (NEMFET) properties, i.e., ultra-low leakage currents, abrupt switching, and hysteresis, with the CMOS technology versatility. The proposed memory relies on NEMFET based Short Circuit Current Free (SCCF) inverters to store data, and on adjacent CMOS based logic to allow for read and write operations, and data preservation. By utilising only one inverter per memory cell, instead of a cross coupled pair, a low write energy is achieved, as only one bitline is required. Furthermore, the static energy is drastically reduced due to NEMFET's extremely low *OFF* current. The proposed 3D NEMFET-CMOS hybrid memory relies on a memory cell 40% smaller in terms of footprint, when compared to a 90nm CMOS dual port memory cell, while for 65nm and 45nm, its footprint is with 9% and 109% larger, respectively, for a 2-die implementations. The access time, when compared with a state of the art 45nm CMOS based dual port memory, is about 12% lower for read operations, while for write operations it is approximately 2× higher, as it is dominated by the mechanical movement of the NEMFET's suspended gate. In our energy consumption analysis, we have considered both small and large size memories and compared CMOS and 3D NEMFET-CMOS implementations. For small memory sizes 8-KB our proposal resulted in about 24% and 29% write energy reduc-

tion, and a read energy reduction of 4.5% and 16%, for 100%, 50% transition probability, respectively. For large memory sizes, e.g., 128-KB, we obtained an energy reduction of about 40%, regardless of the access type, as in this case the static energy is predominant. We have further considered different memory utilisation scenarios for an 8-KB memory, case in which our proposal resulted in about 20% and up to 40% energy reduction for read and write dominated memory access traces, respectively.

Bibliography

- [1] T. Kuroda, “Low-power, high-speed cmos vlsi design,” in *Computer Design: VLSI in Computers and Processors, 2002. Proceedings. 2002 IEEE International Conference on*, 2002, pp. 310 – 315.
- [2] S. Borkar, “Exponential challenges, exponential rewards - the future of moores law,” in *Int. Conf. on Very Large Scale Integration of System-on-Chip*, Dec. 2003, p. 2.
- [3] V. Sharma, S. Cosemans, M. Ashouei, J. Huisken, F. Catthoor, and W. Dehaene, “Ultra low energy SRAM design for smart ubiquitous sensors,” *IEEE Micro*, pp. 1–1, 2012.
- [4] Y. Ye, S. Borkar, and V. De, “A new technique for standby leakage reduction in high-performance circuits,” in *VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on*, jun 1998, pp. 40 –41.
- [5] F. Hamzaoglu and M. Stan, “Circuit-level techniques to control gate leakage for sub-100 nm cmos,” in *Low Power Electronics and Design, 2002. ISLPED '02. Proceedings of the 2002 International Symposium on*, 2002, pp. 60 – 63.
- [6] J. Tschanz, S. Narendra, Y. Ye, B. Bloechel, S. Borkar, and V. De, “Dynamic sleep transistor and body bias for active leakage power control of microprocessors,” *Solid-State Circuits, IEEE Journal of*, vol. 38, no. 11, pp. 1838 – 1845, nov. 2003.
- [7] O. Y. Loh and H. D. Espinosa, “Nanoelectromechanical contact switches,” *Nature Nanotechnology*, vol. 7, no. 5, p. 283295, Apr 2012.
- [8] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, “Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture,” in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [9] R. Nathanael, V. Pott, H. Kam, J. Jeon, and T. Liu, “4-terminal relay technology for complementary logic,” in *IEEE International Electron Devices Meeting*, 2009, pp. 1–4.
- [10] S. Chong, B. Lee, K. Parizi, J. Provine, S. Mitra, R. Howe, and H.-S. Wong, “Integration of nanoelectromechanical (nem) relays with silicon

- cmos with functional cmos-nem circuit,” in *Electron Devices Meeting (IEDM), 2011 IEEE International*, Dec 2011, pp. 30.5.1–30.5.4.
- [11] N. Abele, R. Fritschi, K. Boucart, F. Casset, P. Ancey, and A. M. Ionescu, “Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor,” in *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 2006, pp. 479–481.
- [12] D. Tsamados, A. M. Ionescu, Y. Ye, K. Akarvardar, H. S. Philip Wong, E. Alon, and T. J. King Liu, “{ITRS} - nano-electro-mechanical switches,” Tech. Rep., 2008.
- [13] W. Young Choi, H. Kam, D. Lee, J. Lai, and T.-J. K. Liu, “Compact nano-electro-mechanical non-volatile memory (memory) for 3d integration,” in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, Dec 2007, pp. 603–606.
- [14] H. F. Dadgour and K. Banerjee, “Design and analysis of hybrid NEMS-CMOS circuits for ultra low-power applications,” in *Proceedings of the 44th annual Design Automation Conference*, 2007, pp. 306–311.
- [15] S. Chong, K. Akarvardar, R. Parsa, J.-B. Yoon, R. T. Howe, S. Mitra, and H.-S. P. Wong, “Nanoelectromechanical (nem) relays integrated with cmos sram for improved stability and low leakage,” in *Proceedings of the 2009 International Conference on Computer-Aided Design*, ser. ICCAD '09. New York, NY, USA: ACM, 2009, pp. 478–484.
- [16] X. Huang, C. Zhang, H. Yu, and W. Zhang, “A nanoelectromechanical-switch-based thermal management for 3-d integrated many-core memory-processor system,” *IEEE Transactions on Nanotechnology*, vol. 11, no. 3, pp. 588–600, May 2012.
- [17] K. Akarvardar, C. Eggimann, D. Tsamados, Y. S. Chauhan, G. C. Wan, A. M. Ionescu, R. T. Howe, and H.-S. P. Wong, “Analytical modeling of the suspended-gate FET and design insights for low-power logic,” *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 48–59, Jan. 2008.
- [18] M. Lefter, M. Enachescu, G. Voicu, and S. Cotofana, “Energy effective 3d stacked hybrid nemfet-cmos caches,” in *Nanoscale Architectures (NANOARCH), 2014 IEEE/ACM International Symposium on*, July 2014.

- [19] H. Noguchi, S. Okumura, Y. Iguchi, H. Fujiwara, Y. Morita, K. Nii, H. Kawaguchi, and M. Yoshimoto, "Which is the best dual-port SRAM in 45-nm process Technology? 8T, 10T single end, and 10T differential," in *IEEE International Conference on Integrated Circuit Design and Technology and Tutorial*, 2008, pp. 55–58.
- [20] H. Noguchi, Y. Iguchi, H. Fujiwara, Y. Morita, K. Nii, H. Kawaguchi, and M. Yoshimoto, "A 10T non-precharge two-port SRAM for 74% power reduction in video processing," in *VLSI, 2007. ISVLSI'07. IEEE Computer Society Annual Symposium on*, 2007, p. 107112.
- [21] M. Enachescu, G. R. Voicu, and S. D. Cotofana, "Ultra low power ncmfet based logic," *IEEE International Symposium on Circuits and Systems (ISCAS 2013)*, 2013, to appear.
- [22] D. Tsamados, Y. Singh Chauhan, C. Eggimann, K. Akarvardar, H. S. Philip Wong, and A. Mihai Ionescu, "Finite element analysis and analytical simulations of suspended gate-FET for ultra-low power inverters," *Solid-State Electronics*, vol. 52, no. 9, p. 13741381, 2008.
- [23] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. USA: Addison-Wesley Publishing Company, 2010.
- [24] K. Kanda, H. Sadaaki, and T. Sakurai, "90% write power-saving SRAM using sense-amplifying memory cell," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 6, p. 927933, 2004.
- [25] D. Lu, "Phd dissertation: Compact models for future generation cmos," Ph.D. dissertation, EECS Department, University of California, Berkeley, May 2011. [Online]. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-69.html>
- [26] Prof. Chenming Hu and Prof. Ali Niknejad, "BSIM-IMG: surface potential based UTBSOI MOSFET model," UC Berkeley, 2011. [Online]. Available: http://ekv.epfl.ch/files/content/sites/ekv/files/workshop/2011/Karim_NanoTera_2011.pdf
- [27] "Verilog-A Language Reference Manual," Available: <http://www.verilog.org/verilog-ams/htmlpages/public-docs/lrm/VerilogA/verilog-a-lrm-1-0.pdf>.

- [28] G. J. Coram and M. Dingsur, "Recent Achievements in Verilog-A Compact Modeling," *Presentation at MOS-AK/GSA Workshop on 9 Dec., 2009, Baltimore, 2009*.
- [29] E. Chan, E. Kan, R. Dutton, and P. Pinsky, "Nonlinear dynamic modeling of micromachined microwave switches," in *1997 IEEE MTT-S International Microwave Symposium Digest*, vol. 3. IEEE, Jun. 1997, pp. 1511–1514.
- [30] D. Tsamados, Y. Chauhan, C. Eggimann, K. Akarvardar, H. Wong, and A. Ionescu, "Numerical and analytical simulations of suspended gate - fet for ultra-low power inverters," in *Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European*, sept. 2007, pp. 167–170.
- [31] E. Laulainen, M. Turnquist, J. Makipaa, and L. Koskinen, "Adaptive sub-threshold timing-error detection 8 bit microcontroller in 65 nm cmos," in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, May 2012, pp. 2953–2956.
- [32] T.-C. Huang, T.-W. Chung, C.-H. Chern, M.-C. Huang, C.-C. Lin, and F.-L. Hsueh, "8.4 a 28gb/s 1pj/b shared-inductor optical receiver with 56reduction in 28nm cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, Feb 2014, pp. 144–145.
- [33] B. Pruvost, K. Uchida, H. Mizuta, and S. Oda, "Design optimization of nems switches for suspended-gate single-electron transistor applications," *Nanotechnology, IEEE Transactions on*, vol. 8, no. 2, pp. 174–184, March 2009.
- [34] Y. Wang, M. Enachescu, S. Cotofana, and L. Fang, "Variation tolerant on-chip degradation sensors for dynamic reliability management systems," *Microelectronics Reliability*, vol. 52, pp. 1787–1791, September 2012.
- [35] N. Muralimanohar, R. Balasubramonian, and N. P. Jouppi, "CACTI 6.0: A tool to model large caches," HP Laboratories, Tech. Rep., 2009.
- [36] "Cadence design systems," 2011. [Online]. Available: <http://www.cadence.com/us/pages/default.aspx>

- [37] “Advanced design system (ADS) | agilent,” 2012. [Online]. Available: <http://www.agilent.com/>
- [38] H. Chaabouni, M. Rousseau, P. Leduc, A. Farcy, R. El Farhane, A. Thuair, G. Haury, A. Valentian, G. Billiot, M. Assous *et al.*, “Investigation on TSV impact on 65nm CMOS devices and circuits,” in *Proc. 2010 IEEE International Electron Devices Meeting*, 2010, pp. 6–8.
- [39] A. Topol, D. La Tulipe, L. Shi, S. Alam, D. Frank, S. Steen, J. Vichiconti, D. Posillico, M. Cobb, S. Medd, J. Patel, S. Goma, D. DiMilia, M. Robson, E. Duch, M. Farinelli, C. Wang, R. Conti, D. Canaperi, L. Deligianni, A. Kumar, K. Kwietniak, C. D’Emic, J. Ott, A. Young, K. Guarini, and M. Jeong, “Enabling soi-based assembly technology for three-dimensional (3d) integrated circuits (ics),” in *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, dec. 2005, pp. 352–355.
- [40] H. Kam, T.-J. K. Liu, V. Stojanovic, D. Markovic, and E. Alon, “Design, optimization, and scaling of mem relays for ultra-low-power digital logic,” *Electron Devices, IEEE Transactions on*, vol. 58, no. 1, pp. 236–250, Jan 2011.
- [41] R. Vaddi, V. Pott, G. L. Chua, J. Lin, and T. Kim, “Design and scalability of a memory array utilizing anchor-free nanoelectromechanical non-volatile memory device,” *Electron Device Letters, IEEE*, vol. 33, no. 9, pp. 1315–1317, Sept 2012.
- [42] V. Pott, H. Kam, R. Nathanael, J. Jeon, E. Alon, and T.-J. King Liu, “Mechanical computing redux: Relays for integrated circuit applications,” *Proceedings of the IEEE*, vol. 98, pp. 2076–2094, 2010.
- [43] D. H. Bailey, E. Barszcz, J. T. Barton, D. S. Browning, R. L. Carter, R. A. Fatoohi, P. O. Frederickson, T. A. Lasinski, H. D. Simon, V. Venkatarishnan, and S. K. Weeratunga, “The NAS parallel benchmarks,” *The International Journal of Supercomputer Applications*, Tech. Rep., 1991.

8

Is the Road Towards “Zero-Energy” Paved with NEMFET-based Power Management?¹

Abstract:

In this paper we explore the potential that the use of state-of-the-art Nano-Electro-Mechanical (NEM) devices, i.e., NEMFETs and NEM Relays, in the implementation of power management circuitry, in combination with efficient energy harvesters through 3D stacking integration, have in meeting the tight energy budgets of “Zero-Energy” autonomous sensor systems. We propose various 3D hybrid embodiments of an openMPS430 embedded processor augmented with NEMFET and/or NEM Relay based power management mechanisms and investigate their energy consumption when executing a heart beat detection application. Our investigations indicate that the hybrid NEMFET-oriented approach, which relies on sleep transistors and associated management logic implemented on a dedicated NEMFET die, is the most promising in terms of energy consumption and reliability. Moreover, when combined with a thermal energy harvester, potentially implementable on the same die, it can enable the road towards autonomous computing.

Copyright © 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purpose must be obtained from IEEE by sending an email to pubs - permissions@ieee.org.

¹This chapter is based on the research article published as “Is the Road Towards “Zero-Energy” Paved with NEMFET-based Power Management?” by M. Enachescu, G. R. Voicu, and S. D. Cotofana, *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2561 - 2564, Seoul, Korea, May, 2012. it Finalist for the best student paper award.

8.1 Introduction

Nowadays, CMOS technology is approaching its physical limits, i.e., the scalability frontier of the threshold voltage (V_T) forces a saturated supply voltage owing to the fact that the thermal voltage $k_B \cdot T/q$ does not scale, therefore the power density of integrated circuits increases abruptly [1].

As an alternative, emerging technologies have been proposed and investigated to supersede the basic CMOS device, e.g., Nano-Electro-Mechanical Systems (NEMS), carbon nano tubes, quantum-dot cellular automata, spintronics, ferromagnetic logic, and single-electron devices [1].

The goal of our work is to create an ultra low-power programmable processing platform for embedded smart nodes for pervasive computing (see Figure 8.1). In achieving long life, low energy embedded systems, we cannot realize an architecture based exclusively on eco-friendly (relatively speaking) emerging technologies due to the fact that at this time, only low complexity circuits were successfully designed with the above technologies. In fact, we claim that the road to achieving “green computing” is paved with hybrid solutions, in which CMOS and emerging devices cohabitate.

A first attempt towards our goal was presented in [2], however a more efficient solution is required. In this paper we continue our investigation on energy effective 3D stacked hybrid computation platforms and concentrate on the feasibility of autonomous smart sensor nodes implementations relying on such a hardware infrastructure.

In particular, we propose and evaluate the “zero-energy” potential of an improved version of the 3D-Stacked NEMS based power management architecture in [2], achieved via the following three embodiments: (i) an ultra low power processing core appropriate for wireless sensor nodes, (ii) NEMS based devices for the implementation of sleep transistors (STs) and the additional power management low frequency circuitry necessary for power gating, and (iii) energy harvesters to provide enough energy for the processing core in a low duty cycle application.

In this line of reasoning we propose a number of 3D embodiments of an openMSP430 [3] embedded processor equipped with different power gating mechanisms and evaluate the possibility to power it, when executing a heart beat monitoring application, only by energy harvesting means. We consider the following approaches, all powered at 1 V: (i) a NEMFET-oriented embodiment with both STs and always-on power gating cells implemented with

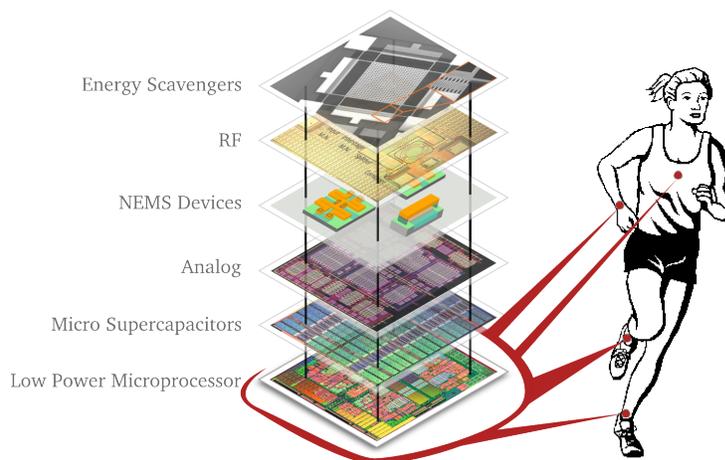


Figure 8.1: Emerging autonomous hybrid 3D stacked bio-sensor embodiment

NEMFETs, (ii) a NEM Relay power gating approach with STs implemented with NEM Relays and always-on power gating cells in CMOS technology, and (iii) a mixed NEMFET/NEM Relay solution with NEM Relay STs and NEMFETs always-on power gating cells. We make use of 3D stacking technology with Through-Silicon-Vias, which greatly reduces the drop-out voltage on the power supply lines between the CMOS logic circuit and the STs on the NEMS die. Moreover, it facilitates further integration for the entire sensor node, through stacking of dies containing analog sensing circuits, energy harvesters, capacitors, etc.

We implement the three proposed designs and compare them with the original 3D Stacked hybrid NEMFET/CMOS power management architecture from [2]. Our investigation suggests that the reduction in energy consumption for the three approaches are: 21.4%, 17.9%, and 23%, respectively. We note that the NEM Relays need an additional on-chip control voltage of at least 6 V for proper operation, and they degrade their ON resistance with switching cycles, drawbacks that could cancel out their energy savings advantages. Thus, the hybrid NEMFET-oriented approach is, for the time being, the best suited solution, and coupled with a thermal energy harvester of only 0.23 cm^2 can enable the road towards autonomous computing.

The rest of the paper is organized as follows. First, in Section 8.2, we give a overview of the state-of-the-art NEMS devices used in digital ICs. Section 8.3 describes the energy budget provided by up-to-date harvesters techniques. In Section 8.4 we present and discuss experimental results and finally, Section 8.5 presents the conclusions.

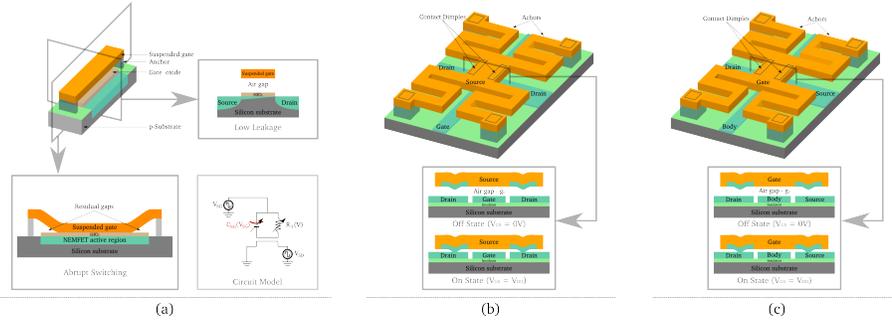


Figure 8.2: Schematic diagram of (a) NEMFET, (b) 3T NEM-Relay, and (c) 4T NEM-Relay

8.2 Nano-Electro-Mechanical Devices as Replacement for MOSFET

The switching energy efficiency of a device is given by the effective sub-threshold swing (S_{eff}). With this consideration in mind, alternative transistor designs which offer perfectly abrupt off-to-on transition are attractive for energy-efficient electronics, since they provide high on/off current ratio with a smaller supply voltage, i.e., a small S_{eff} value. Two such devices are: (i) the Nano-Electro-Mechanical FET (NEMFET) [4], which utilizes the pull-in and pull-out behavior of a mechanical beam (Suspended Gate) to achieve a perfectly abrupt switching transition, and (ii) the micrometer-scale mechanical switches that have been developed for radio frequency electronics (NEM Relay) [5], which function on the same principle as NEMFET with the difference that a mechanical beam is used as the Source of the device to enhance the on/off current ratio.

8.2.1 Nano-Electro-Mechanical Field Effect Transistor

The Nano-Electro-Mechanical FET (NEMFET) is a rather complex device with a 3D geometry and cross-section as presented in Figure 8.2a.

Essentially speaking the device behaves like an electromechanical switch which responds to gate bias changes as follows. When the gate voltage V_G is low the gate-oxide capacitance is in series with the air-gap capacitance resulting in low electrostatic coupling of the gate to the channel, thus in a negligible drain current I_D . If V_G increases the situation remains unchanged until it reaches the pull-in voltage V_{PI} in which case the electrostatic force cannot

be compensated anymore by the mechanical restoring force and the suspended gate (beam) snaps onto the gate oxide, thus turning on the device. After the pull-in, the I_D increase with V_G is comparable with the one of a standard MOSFET. On the other way around when V_G is decreased from some high value I_D starts decreasing until at a certain V_G value when the system becomes unstable due to combined electro-mechanical force and the beam is pulled-out. This causes an abrupt I_D decrease due to a large decrease in capacitance, i.e., pull-out effect.

As indicated in [6] NEMFET devices can potentially replace High- V_T sleep transistors (ST) due to their ultra low leakage characteristics. Note that for power gated designs two major features are desirable for the sleep transistors: (i) low sub-threshold leakage current to minimize the static power consumption, and (ii) low "ON" state resistance to minimize the voltage difference between the virtual and the real power supply nodes.

A comparison between the "ON" state resistance R_{ON} and "OFF" state leakage current I_{OFF} of an optimal NEMFET in terms of area and on-to-off ratio having $w_{Beam}=250$ nm, $L_{Beam}=7.5$ μm , $air_{gap}=20$ nm, $V_{PT}=1.1$ V and 65 nm High- V_T CMOS based STs was presented in [2]. It suggests that the NEMFET ST R_{ON} tends to become equal with the CMOS ST R_{ON} , while NEMFET ST I_{OFF} (leakage) is about 2 orders of magnitude smaller than the one of the CMOS counterpart.

8.2.2 NEM Relays

The scalability of the NEMFET due to the presence of air-gap and short channel effects is limited. In face of this limitation, micro-electro-mechanical relays, also termed "micro-relays" [7], appear to be an attractive alternative for zero-standby power logic applications. The attractiveness of micro-relays stems from the fact that a mechanical switch offers nearly ideal switching characteristics: zero off state drain-to-source and gate leakage currents, and perfectly abrupt off-to-on switching transition. Since there is no trade-off between off-state leakage current and on-state drive current, the relay threshold voltage and therefore V_{dd} can be in principle reduced much more aggressively than for a MOSFET in order to improve the energy efficiency.

The 3T NEM Relay described in [8] is a device with a rather similar 3D geometry when compared with the NEMFET. It essentially is an electrostatic switch having three terminals, as illustrated in Figure 8.2b. In the off state, when the

source and the drain electrodes are separated by an air gap, the behavior is similar with an open relay, thus no current can flow between the electrodes. When the amplitude of the gate-to-source voltage V_{GS} is sufficiently large, above V_{PT} , the source electrode is actuated downward into contact with the drain electrode so that current can flow under the influence of the drain-to-source voltage V_{DS} . Various 3T switch designs have been reported in the literature [8,9]. However, their performance is limited by the high operating voltage larger than 5 V), and a slow actuation of about 100 ns [5].

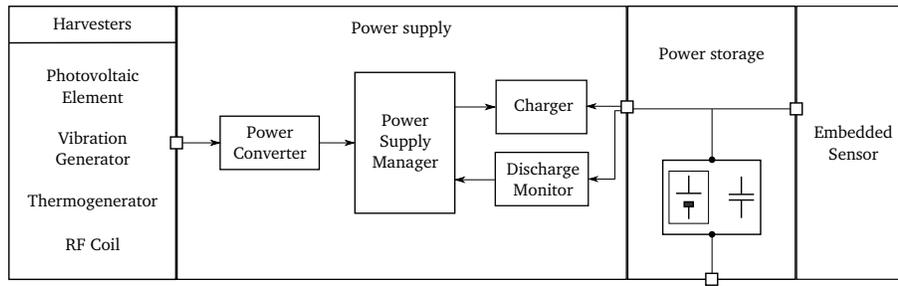
A 4T relay, depicted in Figure 8.2c, has been also proposed. To cope with high operation voltage and actuation time, the 4T relay design, a normally off device, adds a body terminal. With the beam acting now as the gate, the voltage between the gate and the body V_{GB} determines the state of the relay. When the amplitude of the V_{GB} is higher than V_{PT} , the relay will be turned on. Thus, a fixed voltage, and independent from the source and drain voltages is required for the body terminal to limit the variation of the gate switching voltage. Low switching voltages are desirable for low active power consumption. V_{PT} can be reduced by applying more advanced manufacturing techniques, i.e., increasing the actuation area ($W \times L$), decreasing the thicknesses of the movable structure, or decreasing the height of the fabricated actuation gap g_0 . However, all of these negatively affect the device density. Another post-process solution is to tune the gate switching voltages via body biasing, but extra circuitry and power consumption are required in order to generate on-chip large V_B values, approximately 8 V as presented in [10], for a device suited for 10 MHz circuitry. Moreover, the ON state resistance of the 4T Relays is increasing from $k\Omega$ to tens of $k\Omega$ after only 10^4 actuations, practically limiting the digital implementation range only to low activity circuits.

8.3 Power budgeting of energy harvesters

To address the problem of finite node lifetime of battery-powered smart systems energy harvesting is employed. Different forms of energy from the ambient environment or from other sources are converted to electrical energy which powers the node. If the harvested energy is large enough and continuously available, a smart node can even be powered forever (or at least until a failure in the power supply circuitry appears). Furthermore, by monitoring the amplitude and periodicity of the harvested energy, the smart sensor can self-allocate its available energy and alter at run-time the functioning duty-cycle to

Table 8.1: Various energy sources and harvested power densities [11]

Source type	Source power	Harvested power
Indoor ambient light	0.1 mW/cm ²	10 μW/cm ²
Vibration/motion	0.5 m @ 1 Hz 1 m/s ² @ 50 Hz	4 μW/cm ²
Thermal	20 mW/cm ²	30 μW/cm ²
Radiofrequency	0.3 μW/cm ²	0.1 μW/cm ²

**Figure 8.3:** System-level power supply architecture

match the supply energy.

The architecture of the power supply of the sensor node is depicted in Figure 8.3. The harvester transforms the ambient energy (e.g., photonic, kinetic, thermal, RF energy) in a variable voltage. A power converter performs voltage rectification and smoothens/regulates the voltage to the desired value. The power supply manager controls the operation of the converter and the optional charging circuits, when a storage element (battery, super-/ultra-capacitor) is needed.

Since our end goal is to achieve a fully autonomous smart sensor node we need to know the power budgets that harvesters can produce. Note that the form in which the energy is harvested, the energy source characteristics, and the conversion efficiency dictate the power budget. Table 8.1 summarizes the typical produced and expected harvested power densities for the most used energy sources in a human ambient environment.

Harvesting photovoltaic energy offers zero or close to zero output during nighttime, and a battery is required for continuous operation. A temperature gradient can be converted in electrical energy via micro-machined thermopiles integrated in the same package with the sensor node. The human skin is a fa-

favorite energy source since it provides a significant temperature difference from the surrounding environment. However, it requires a more complex package, with one surface being a plate with the exterior side in thermal contact with the human skin, and the interior side in direct electrical contact with one end of the thermopiles. Due to their ease of integration within the same package, micro-machined piezoelectric resonant vibration devices are the most promising in harvesting motion energy for smart sensor nodes. Finally, radiofrequency (RF) energy, though covering large surfaces of the inhabited land nowadays offers the less power compared with all other sources.

8.4 Results Evaluation and Discussion

The base 3D stacked power management architecture described in [2] consists of two dies connected with Through-Silicon Vias. The bottom CMOS tier comprises the active digital logic circuit, and can be power gated through the NEMFET STs incorporated on the top tier.

We first consider an Enhanced architecture which improves the energy efficiency of the platform by two means: (i) relocation of the openMSP power management controller (PMC) and power gating overhead circuitry, i.e., isolation and state-retention cells, on the NEMS die, and (ii) reduction the supply voltage from 1.2 V to 1.0 V. The first approach entails redesigning the referred blocks in the ultra low leakage NEMS technology, hence alleviating part of the leakage overhead associated with the use of CMOS devices. Thus, the bottom CMOS tier comprises the active logic circuit, while the top tier incorporates the always-on cells and the PMC implemented in NEMS technology. Moreover we consider two additional platforms both having STs implemented with NEM Relays and always-on power gating cells implemented in CMOS and NEMFET technology, respectively.

For evaluation we use the same test vehicle used to validate the base architecture, i.e., the *Hybrid* design, a typical SoC for low power embedded devices running a heart rate monitor application [2]. To assess the impact of using NEM Relays as sleep transistors, we use the same methodology previously applied for NEMFET-based power gating.

The energy consumption values obtained after simulation in worst-case technology corner, with respect to power consumption, for both the Enhanced NEMFET- and NEM Relays-based power management architectures, as well as the original reference *Hybrid* are presented in Table 8.2. One can observe

Table 8.2: Energy budgeting

Implementation	Energy [μ J]	Required Harvester Area [cm^2]			
		Light	Vibration	Thermal	RF
<i>NEMFET-based STs</i>					
Hybrid [2], $V_{DD} = 1.2$ V	8.611	0.86	2.15	0.29	86.11
Enhanced Hybrid	6.761	0.67	1.69	0.23	67.61
<i>NEM Relay-based STs</i>					
Always-on cells w/ CMOS	7.073	0.71	1.77	0.24	70.73
Always-on cells w/ NEMFET	6.579	0.66	1.64	0.22	65.79

that due to the supply voltage reduction and the placement of the low-activity power management logic on the NEMS die, the *Enhanced Hybrid* NEMFET-based implementation provides an energy consumption reduction of 21.4%.

For the NEM relay based implementations we have to make use of MOSFETs or NEMFETs for the implementation of the PMC block. This is motivated by the fact that, even for the considered biomedical application, which has a rather low switching activity of 200 cycles per second, the NEM Relays' R_{ON} is increasing with one order of magnitude within tens of seconds of circuit life time. This makes NEM Relays not suited to implement the logic gates from the always-on cells and the PMC block.

Even though at circuit level the NEM relay based approach does not provide a significant energy improvement over the NEMFET counterpart, this can be mostly explained by the fact that the considered application has a rather high duty cycle, and the ON power term is dominant in the total energy equation. Seen from device level, the advantages of the NEM Relays are obvious when compared with NEMFETs, as they offer almost ideal switching characteristics, “zero” off state leakage current, “infinite” on current, and abrupt off-to-on and on-to-off switching transitions. However for the time being their integration in a hybrid CMOS/NEMS architecture for low power embedded smart sensor nodes requires the on-chip generation of high voltages of approximately 6 V. Hence such an approach adds extra circuitry overhead which will most likely cancel out its advantages over the NEMFET solution. Moreover NEM Relay are less reliable and substantially increase their R_{ON} after tens of thousands switching cycles. This result in an IR drop increase that may make the entire circuit to malfunction.

Based on the total energy requirements we also compute in Table 8.2 the area

that an energy harvester needs to have to be able to autonomously power our platform. One can observe in Table 8.2 that the NEMS based schemes we proposed can be potentially powered from harvesting sources thus opening the road towards the implementation of zero-energy autonomous computing nodes. Light and thermal energy sources seem to be the most promising candidates to supply miniature autonomous smart sensors. While vibration generators do not offer the same power density, they do not need to be in visible or thermal contact with the environment so they are more easily to be integrated in future smart micro-sensors with a stacked layout.

8.5 Conclusions

In this paper we explored the possibilities to achieve autonomous, self-powered embedded smart sensors, by an in-depth investigation of the use of NEMS devices as power gating circuitry, followed by an energy budgeting of autonomous smart sensor nodes powered by various types of energy harvesters. Our investigations suggests that due to obvious advantages of the NEMS switches at device level, efficient energy harvesters design and 3D stacked integration technology we are entitled to answer affirmatively to our research question.

Bibliography

- [1] ITRS, “Emerging Research Devices,” 2009. [Online]. Available: <http://www.itrs.net/>
- [2] G. R. Voicu, M. Enachescu, and S. D. Cotofana, “Towards ”Zero-energy” using NEMFET-based power management for 3D hybrid stacked ICs,” in *2011 IEEE/ACM International Symposium on Nanoscale Architectures*, San Diego, CA, USA, Jun. 2011, pp. 203–209.
- [3] OpenCores, “openMSP430 softcore,” 2011. [Online]. Available: <http://opencores.org>
- [4] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, “Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture,” in *International Symposium on Quality Electronic Design*, 2002, pp. 496–501.
- [5] V. Pott, H. Kam, R. Nathanael, J. Jeon, E. Alon, and T.-J. King Liu, “Mechanical computing redux: Relays for integrated circuit applications,” *Proceedings of the IEEE*, vol. 98, pp. 2076–2094, 2010.
- [6] M. Enachescu, S. Cotofana, A. Genderen, D. Tsamados, and A. Ionescu, “Can SG-FET replace FET in sleep mode circuits?” *Nano-Net*, pp. 99–104, 2009.
- [7] K. Akarvardar, D. Elata, R. Parsa, G. Wan, K. Yoo, J. Provine, P. Peumans, R. Howe, and H. Wong, “Design considerations for complementary nanoelectromechanical logic gates,” in *IEEE International Electron Devices Meeting*, 2007, pp. 299–302.
- [8] H. Kam, V. Pott, R. Nathanael, J. Jeon, E. Alon, and T. Liu, “Design and reliability of a micro-relay technology for zero-standby-power digital logic applications,” in *IEEE International Electron Devices Meeting*, 2009, pp. 1–4.
- [9] H. Kam, “MOSFET replacement devices for energy-efficient digital integrated circuits,” DTIC Document, Tech. Rep., 2009.
- [10] R. Nathanael, V. Pott, H. Kam, J. Jeon, and T. Liu, “4-terminal relay technology for complementary logic,” in *IEEE International Electron Devices Meeting*, 2009, pp. 1–4.

- [11] R. Vullers, R. van Schaijk, I. Doms, C. Van Hoof, and R. Mertens, "Micropower energy harvesting," *Solid-State Electronics*, vol. 53, no. 7, p. 684693, 2009.

9

Conclusions and Future Work

In this thesis we focused on improving the energy efficiency of electronic products, especially portable, battery-powered, and autonomous ones, by making use of the well-established versatile CMOS devices in conjunction with an emerging leakage proof technology. More specifically, we proposed and assessed the potential practical impact and feasibility of hybrid CMOS-NEMFET circuits and systems. We also demonstrated that the CMOS-NEMFET device synergy enabled by 3D stacked integration can deliver substantial energy savings for power managed ICs as well as foster alternative implementations of Boolean functions and Static-RAM (SRAM) memory arrays. We pursued 3 main lines of research: NEMFET based Boolean logic, NEMFET-CMOS dual port memory, and hybrid NEMFET-CMOS power management architectures. First, we performed a design space exploration in order to identify the most promising NEMFET geometries and to evaluate their potential performance in terms of switching delay, current capability, and leakage. Moreover we compared those parameters of interest with the ones offered by traditional transistors utilized in up to date CMOS technologies. Second, we assessed the NEMFET potential when utilized as sleep transistor in circuits featuring 2D cell based power gating, and find out if NEMFETs constitute a viable alternative to *High* - V_{TH} FETs in sleep mode circuits. Furthermore, we proposed a novel 3D power management approach that attempts to alleviate some issues associated with the NEMS utilization as sleep transistor in CMOS power gated integrated circuits. Given the two designs, we evaluated the 2D and 3D NEMFET based power management implementations energy efficiency when embedded into a computation platform executing a bio-medical sensing application for heart rate detection. Third, we introduced a NEMFET based logic family tailored to the implementation of ultra-low energy functional units and processors. Fourth, we proposed a

memory cell that relies on a NEMFET based inverter designed in such a way that no short circuit current can occur during switching. Finally, we proposed and evaluated the "zero-energy" operation scenario potential of an improved version of the 3D-Stacked NEMS based power management architecture augmented with energy harvesters.

This chapter is organized as follows. Section 9.1 presents an overview of the thesis and summarizes the major contributions of this thesis and Section 9.2 proposes future research directions.

9.1 Summary

In this dissertation we investigated low-energy circuits and architectures implemented with Nano-Electro-Mechanical Field Effect Transistors (NEMFETs) only, or with both MOSFETs and NEMFETs. The work presented in this thesis can be summarized as follows:

In Chapter 1 we discuss the necessity to make use of emerging leakage proof technologies in conjunction with the versatile well-established CMOS counterpart for the implementation of portable, battery-powered, and autonomous electronic products. We introduced the essential knowledge of the power dissipation in such electronic products and pointed out that the efficiency of the current leakage reduction techniques is determined by the MOSFET leakage and the designer ability to control it. We analyzed the MOSFET capability to effectively reduce the overall energy consumption as well as the ability of state of the art *More than Moore* devices in an attempt to identify potential candidates that are better fitted for portable, battery powered, autonomous systems implementation. We identified NEMFET, which is potentially co-integrable with MOSFET, as being the best candidate to reduce the overall energy consumption of ultra low power electronic products and formulated our research questions.

In Chapter 2 we introduced NEMFET background and we presented a preliminary assessment of the NEMFET potential if utilized as Sleep Transistor (ST) in real life circuits, e.g., microprocessors. We first evaluated various NEMFET instances in terms of switching delay, current capability, and leakage. Subsequently, we compared these figures with the ones provided by traditional switch transistors utilized in CMOS technologies. According to our simulation results, NEMFET based sleep transistors enable substantial leakage reductions due to their extremely low *OFF* currents (4 orders of magnitude

lower than FET) at the expense of a $4\times$ larger active area for the same I_{ON} capability. Finally, in Chapter 3 we evaluated the potential implications of the utilization of NEMFETs as sleep transistors in a 90 nm CMOS technology 32-bit Adder. Our simulations indicated that the leakage is mitigated, at the expense of a sleep transistor active area increase of 130%.

In Chapter 4 we introduced a novel power management 3D architecture, which relies on the synergistic utilization of: (i) Nano-Electro-Mechanical (NEM) devices, i.e., the NEM Field Effect Transistor (NEMFET), as sleep transistors, and (ii) 3D stacking, which allows for placing the sleep transistors (the entire power management infrastructure) on a dedicated tier of the 3D-stacked hybrid platform. Subsequently we demonstrated that 3D integration is an efficient way to make use of the NEMFET characteristics, i.e., abrupt switching and extreme low leakage, in effective NEMS-based power/energy/thermal management. As a test case, in Chapter 5 we considered the 3D embodiment of an embedded *openMSP430* processor based SoC platform running a bio-medical sensing application for heart rate detection and measured the effects of the 3D hybrid architecture on sensitive metrics used in power gating designs, e.g., delay degradation, power-up and power-down behavior, and overall energy consumption. Our experiments indicated that when compared with a 2D embodiment, the system idle energy is decreased by $2.74\times$ for the same footprint due to the fact that the sleep transistors were relocated on the NEMS tier. The energy-delay product of the *openMSP430* processor based SoC executing the heart rate detection application was reduced by 9%, but we demonstrated that a potential reduction of up to 60% could be achieved for applications with lower activity, e.g., wireless sensor networks. Furthermore, we demonstrated that the 3D stacked architecture prevents clock period degradation issues, since the IR Drop is reduced with a factor of 4 when compared with the 2D embodiment.

In Chapter 6, we introduced a Short-Circuit-Current Free (SCCF) NEMFET based logic family tailored to the implementation of low speed and ultra low energy functional units and processors. We analysed and compared basic Boolean gates implemented with NEMFETs against equivalent CMOS realisations. Our simulations suggested that the proposed SCCF NEMFET gates are between 10 to $20\times$ slower, but provide up to $10\times$ dynamic energy reduction and up to 2 orders of magnitude less leakage, when compared with CMOS counterparts. We also analysed the fan-in influence on gate performance and observed that the NEMFET gates energy advantage is increasing with fan-in.

In Chapter 7, we introduced a dual port 3D stacked hybrid memory that combines the ultra-low power NEMFETs capabilities with the MOSFET technology versatility. The proposed memory relies on NEMFET based SCCF inverters to store data, and on adjacent CMOS based logic to allow for read and write operations, and data preservation. By utilising only one inverter per memory cell, instead of a cross coupled pair, a low write energy is achieved, as only one bitline is required. Furthermore, the static energy is drastically reduced due to NEMFET's extremely low I_{OFF} . The proposed dual port 3D NEMFET-CMOS hybrid memory relies on a memory cell with 140% and 30% footprint increase, for 2-die and 3-die implementations, respectively. However, by placing the memory array column and row circuitry within the memory cells, the total footprint of an 8-KB memory increases with only 60% for a 2-die implementation and decreases with 20% for a 3-die implementation. The access time, when compared with a state of the art CMOS based dual port memory, is equivalent for read operations, while for write operations it is approximately $4\times$ higher, as it is dominated by the mechanical movement of the NEMFET's suspended gate. However we propose a solution to achieve a similar write access time by adding extra write ports, which, due to the available CMOS tier real estate, has no negative impact on footprint and read access time. We also compared the energy consumption of standard and hybrid memory arrays as follows: (i) for small memory sizes of 2-KB and 8-KB our proposal results in about 10% and 30% write energy reduction, and a read energy reduction of 10% and 13%, respectively, and (ii) for large memory sizes, e.g., 128-KB, we obtain an energy reduction of 58%, regardless of the access type, as in this case the static energy is predominant. We have further considered different memory utilisation scenarios for an 8-KB memory, case in which our proposal results in up to 22% and up to 35% energy reduction for read and write dominated memory access traces, respectively.

In Chapter 8, we proposed and evaluated the "zero-power" operation paradigm potential of an improved version of the 3D-Stacked NEMS, i.e., NEMFETS and NEM-Relay, based power management architecture when executing a heart beat detection application. The platform builds upon the following three embodiments: (i) an *openMSP430* ultra low power processing core appropriate for wireless sensor nodes, (ii) NEMS devices for the implementation of Sleep Transistors and of the additional power management low frequency circuitry necessary for power gating, and (iii) energy harvesters to provide enough energy for the processing core when executing a low duty cycle application. Our investigations indicated that the hybrid NEMFET-oriented approach, which

relies on sleep transistors and associated management logic implemented on a dedicated NEMFET die, is the most promising in terms of energy consumption and reliability. Moreover, when combined with a thermal energy harvester, of 0.23 cm^2 potentially implementable on the NEMFET die, it can enable the road towards autonomous operation.

9.2 Future Research Directions

As a continuation of the research presented in this thesis we suggest the following future research directions:

1. Given that NEMFET's switching speed is dominated by mechanical phenomena the distance between the suspended gate beam and the gate oxide should be as small as possible in order to minimise the beam traveling distance associated with device switching. However, to avoid the metal beam to dielectric stiction the elastic beam force should be larger than the Casimir force, which for air gaps below 150nm [1] can be two orders of magnitude higher than the gate potential induced electrostatic force. Given that the NEMFET device geometry has to fulfil certain conditions in order to avoid stiction it would be interesting to develop alternative methods to increase the switching speed by other means than shrinking. One potential solution that could be considered is to make use of an input gate voltage control strategy that we propose to call progressive switching, which takes advantages of the NEMFET abrupt switching and hysteresis. While for a normal FET the input gate value can change from 0V to V_{DD} or the other way around for a NEMFET this set can be augmented with the V_{PI}^- and V_{PO}^+ values, where V_{PI}^- is slightly smaller than the gate pull-in voltage and V_{PO}^+ is slightly larger than the gate pull-out voltage. In this way the gate of an already open (non conducting) NEMFET can be kept at V_{PI}^- and the one of an already closed (conducting) NEMFET can be kept at V_{PO}^+ . In this way when the NEMFET should switch, the gate is already partially charged/discharged, hence the latency to fully charge/discharge can be significantly decrease.
2. Current NEMFET compact models are geared towards the simulation of low complexity schemes. In order to better simulate NEMFET abrupt switching and hybrid CMOS-NEMFET circuits a NEMFET compact

model compatible with commercial simulators. e.g., Spectre, is required.

3. Given the NEMFET large size it is expected that it would also go through a device shrinking process. In view of this it would be interesting to investigate the potential performance of NEMFET scaled devices and the NEMFET scaling process implications of the scheme we proposed in this thesis.
4. While our main focus was on the digital part of energy confined embedded platforms, e.g., wireless sensor nodes, they also contain analog circuitry, which may take a big share of the energy budget. In view of this it would be of interest to investigate the porting of such analog schemes into the NEMFET technologies.
5. Although we proved the efficiency of NEMFET-CMOS dual port memory, no detailed TSV size implication evaluation was performed. This is certainly of interest given that TSV shrinking follows a different agenda than the device one. Moreover, given that silicon real-estate is still available on the CMOS tier it would also be of interest to extend the number of read/write ports and investigate the implications of this augmentation in terms of memory access time, energy consumption, and availability.
6. Our experimental results were obtained by means of simulations only. To get a better grasp in the practical implications of our proposals it would be of interest to implement a simple 3D stacked power management architecture composed out of a CMOS tier for high frequency operations and NEMFET tier for low activity, low frequency operations.

Bibliography

- [1] H. De Los Santos, “Impact of the casimir force on movable-dielectric rf mems varactors,” in *Nanotechnology, 2003. IEEE-NANO 2003. 2003 Third IEEE Conference on*, vol. 2, Aug 2003, pp. 900–903 vol. 2.

List of Publications

Book Chapter

1. M. Enachescu, A. van Genderen, D. Tsamados, A. Ionescu, S. D. Cotofana, **Can SG-FET Replace FET in Sleep Mode Circuits?**, *Nano-Net*, Vol. 20, pp. 99-104, 2009.

International Conferences

1. M. Enachescu, A.J. van Genderen, S.D. Cotofana, **Suspended Gate Field Effect Transistor Based Power Management - A 32-Bit Adder Case Study**, *IEEE International Semiconductor Conference (CAS 2009)*, Sinaia, Romania, October 2009. **Honorific Mention.**
2. M. Enachescu, G.R. Voicu, S.D. Cotofana, **Advanced NEMS-based Power Management for 3D Stacked Integrated Circuits**, *IEEE International Conference on Energy Aware Computing (ICEAC 2010)*, Cairo, Egypt, December 2010.
3. G.R. Voicu, M. Enachescu, S.D. Cotofana, **Towards "Zero-energy" using NEMFET-based Power Management for 3D Hybrid Stacked ICs**, *IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH 2011)*, San Diego, USA, June 2011.
4. M. Enachescu, G.R. Voicu, S.D. Cotofana, **Leakage-enhanced 3D-Stacked NEMFET-based Power Management Architecture for Autonomous Sensors Systems**, *IEEE 15th International Conference on System Theory, Control and Computing (ICSTCC 2011)*, Sinaia, Romania, October 2011. **Best Paper Award for PhD Students.**
5. M. Enachescu, G.R. Voicu, S.D. Cotofana, **Is the Road Towards "Zero-Energy" Paved with NEMFET-based Power Management?**, *IEEE International Symposium on Circuits and Systems (ISCAS 2012)*, Seoul, Korea, May 2012. **Finalist for Best Paper Award for PhD Students.**
6. M. Enachescu, M. Lefter, A. Bazigos, A. Ionescu, S.D. Cotofana, **Ultra Low Power NEMFET Based Logic**, *IEEE International Symposium on Circuits and Systems (ISCAS 2013)*, Beijing, China, May 2013.

7. M. Lefter, M. Enachescu, G.R. Voicu, S.D. Cotofana, **Energy Effective 3D Stacked Hybrid NEMFET-CMOS Caches**, *ACM/IEEE International Symposium on Nanoscale Architectures (NANOARCH 2014)*, Paris, France, July 2014.

International Journals

1. M. Enachescu, M. Lefter, G. R. Voicu, and S. D. Cotofana, **Low-Leakage 3D Stacked Hybrid NEMFET-CMOS Dual Port Memory**, to be submitted at *IEEE Transaction on Emerging Topics in Computing Special Issue on Design and Technology of Integrated Systems in Deep Submicron Era*.

Other International Journals and Conferences non related with this thesis

1. Y. Wang, M. Enachescu, S.D. Cotofana, L. Fang, **Variation tolerant on-chip degradation sensors for dynamic reliability management systems**, *Microelectronics Reliability*, volume 52, issue 9-10, September 2012 [Journal Paper].
2. Al. Rusu, D. Dobrescu, M. Enachescu, C. Burileanu, A. Rusu, **The Small Signal Amplification of the Gated Diode Operated in Breakdown Regime**, *IEEE 34th International Semiconductor Conference (CAS 2011)*, Sinaia, Romania, October 2011.
3. G.R. Voicu, M. Enachescu, S.D. Cotofana, **A 3D Stacked High Performance Scalable Architecture for 3D Fourier Transform**, *30th IEEE International Conference on Computer Design (ICCD 2012)*, Montreal, Canada, October 2012.
4. M. Lefter, G.R. Voicu, M. Taouil, M. Enachescu, S. Hamdioui, S.D. Cotofana, **Is TSV-based 3D Integration Suitable for Inter-die Memory Repair?**, *Design, Automation and Test in Europe Conference and Exhibition (DATE 2013)*, Grenoble, France, March 2013.
5. G.R. Voicu, M. Lefter, M. Enachescu, S.D. Cotofana, **3D Stacked Wide-Operand Adders: A Case Study**, *24th IEEE International Conference on Application-Specific Systems, Architectures and Processors (ASAP 2013)*, Washington D.C., USA, June 2013.

-
6. C. Chen, M. Enachescu, S.D. Cotofana, **Enabling Vertical Wormhole Switching in 3D NoC-Bus Hybrid Systems**, *18th Design, Automation and Test in Europe conference (DATE 2015)*, Grenoble, France, March 2015.

Curriculum Vitae

Marius Enachescu was born in Bucharest, Romania, on 22nd of October 1982. After his high school graduation ("Costin C. Kiritescu" College, Bucharest, Romania) in 2002, he started studying electrical engineering at the University Politehnica of Bucharest (UPB), Romania. He received the Engineer Degree (equivalent with Bologna Master) in Microelectronics from UPB in 2007. While a student, between the year 2004 and 2007, Marius worked as an intern in analog design and layout at Atmel in Bucharest. Subsequently, after graduation, between 2007-2008, he was an analog design engineer at Atmel in Bucharest. In August 2008, he joined the Department of Software and Computer Technology at the Faculty of Electrical Engineering, Mathematics, Computer Science at Delft University of Technology (TU Delft), the Netherlands, to pursue a Ph.D. under the supervision of Dr. Sorin Cotofana. In 2014, between January and May, Marius was a PhD intern in signal integrity at ARM in Cambridge, United Kingdom. Starting September 2014, Marius works as a senior analog designer at Microchip in Bucharest.