



Design and Fabrication of 3D Hybrid CMOS/NEM relays technology for energy-efficient electronics

Giulia Usai

► To cite this version:

Giulia Usai. Design and Fabrication of 3D Hybrid CMOS/NEM relays technology for energy-efficient electronics. Micro and nanotechnologies/Microelectronics. Université Grenoble Alpes, 2019. English. NNT : 2019GREAT069 . tel-02529390

HAL Id: tel-02529390

<https://theses.hal.science/tel-02529390>

Submitted on 2 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Nanoélectronique et Nanotechnologies**

Arrêté ministériel : 25 mai 2016

Présentée par

Giulia USAI

Thèse dirigée par **Thomas Ernst, *PhD*, CEA Leti** et **Tsu-Jae King Liu, *PhD*, UC Berkeley**

Préparée au sein du **Laboratoire d'électronique et de technologie de l'information (CEA-Leti)** dans l'**École Doctorale électronique, électrotechnique, automatique et traitement du signal EEATS**

Conception et Fabrication hybride 3D monolithique de relais NEMS co- intégrés CMOS

Thèse soutenue publiquement le « **5 12 2019** »,
Devant le jury composé de :

Dr. Laurent, PICHON

Professeur, Université de Rennes 1 (Président)

Dr. Núria, BARNIOL BEUMALA

Professeur, Universitat Autònoma de Barcelona, Espagne (Rapporteur)

Dr. Damien, DELERUYELLE

Professeur, INSA Lyon, (Rapporteur)

Dr. Louis, HUTIN

Ingénieur-Docteur, CEA LETI, Grenoble (Encadrant)

Dr. Thomas, ERNST

Directeur de recherche au CEA-Leti, Grenoble (Directeur de thèse)

Dr. Tsu-Jae, KING LIU

Professeur, UC Berkeley, États-Unis (Co-Directrice de thèse)



*A mio Marito Giorgio Palma,
Fonte di infinito amore e sostegno.*

ABSTRACT

This manuscript focuses on Nano-Electro-Mechanical (NEM) relays with electrostatic actuation for advanced logic and memory applications. The use of Nano-Electro-Mechanical relays was recently proposed for digital logic circuits in order to overcome the fundamental energy-efficiency limitations that mainstream CMOS technology is currently facing. The cumulated benefits of essentially Zero Off-State current and ultimately abrupt DC switching characteristics enable alleviating the power-performance trade-off as the supply voltage V_{DD} is reduced. Additionally, for some particular switch designs (e.g. free of dielectric layers), an increased resistance to ionizing radiations is also anticipated, making such components valuable for defense or aerospace applications.

However, NEM relays have intrinsic limitations in terms of integration density, endurance and operation frequency. Therefore, rather than considering them as technology that could replace MOSFETs, we adopt an intermediate approach that consists in using NEM relays as a complement to CMOS circuits (e.g.: buffers, non-volatile elements for SRAM and CAM), which can be fabricated in a 3D co-integration scheme. This approach mitigates the area penalty issue.

The thesis explores the strength and the weakness of NEMS relays and identifies applications for which hybrid NEMS/CMOS circuits are potentially interesting.

This work includes the manufacturing of prototype devices designed to be proof of concept for the identified applications. At first, NV NEM relays design and dimensioning through modeling and simulations was performed. Then NV NEM/CMOS circuits were validated through simulations. This was followed by the tapeout and the process integration of monolithically co-integrated NEMS above CMOS. After wafer processing the devices were electrically characterized.

This all-inclusive works allows identifying some crucial challenges that NEMS relays still have to face.

RESUME

Ce travail de thèse porte sur les relais nano électromécaniques (NEMS en anglais). Afin de surmonter la limite d'efficacité énergétique inhérente à la technologie CMOS, l'utilisation de relais NEMS pour des circuits logiques a été récemment proposée. En effet, les bénéfices cumulés d'un courant nul à l'état bloqué et d'une commutation abrupte permettent d'améliorer le compromis puissance-performance lorsque la tension d'alimentation V_{DD} est réduite.

Cependant, les relais NEMS sont limités intrinsèquement par leur endurance, leur fréquence de fonctionnement ainsi que par leur faible densité d'intégration. De ce fait, une approche complémentaire basée sur la cointégration entre CMOS et relais (mémoire tampon, éléments non volatile) plutôt qu'un remplacement semble plus pertinent à court terme. Une intégration 3D monolithique permet l'ajout de tels éléments sur la même surface tout en bénéficiant d'une précision d'alignement lithographique. De plus, pour certains dispositifs (ex. sans couche diélectrique), une plus grande immunité aux radiations ionisantes est attendue, ouvrant des perspectives dans les domaines des applications militaires ou spatiales.

En particulier, ce travail de thèse explore les forces et les faiblesses de ces structures et identifie des applications à fort potentiel. Il comprend en outre la conception de prototypes correspondants, depuis leur modélisation et simulation permettant un dimensionnement fonctionnel, le perfectionnement de l'intégration 3D « NEMS sur CMOS », et enfin leurs tests électriques. A la lumière de ces analyses, des recommandations sont fournies pour le développement de circuits hybrides 3D CMOS/NEMS pour la logique et les mémoires avancés.

ACRONYMS

AFM	A tomic F orce M icroscope
ALD	A tomic L ayer D eposition
CBRAM	C onductive B ridge R andom A ccess M emory
CMOS	C omplementary M etal O xide S emiconductor
CMP	C hemical- M echanical P olishing
CPI	C atastrophic P ull- I n
CPU	C entral P rocessing U nit
BEOL	B ack E nd O f L ine
DIBL	D rain I nduced B arrier L owering
DL	D rain L eft
DR	D rain R ight
DRAM	D ynamic R andom A ccess M emory
DVS	D ynamic V oltage S caling
DUV	D eep U ltra V iolet
EBEAM	E lectron B EAM
EOT	E ffective O xide T hickness
FinFET	F in F ield E ffect T ransistor
FEA	F inite E lement A nalysis
FEM	F inite E lement M ethod
FEOL	F ront E nd O f L ine
FeRAM	F erro-electric R andom A ccess M emory
FDSOI	F ully D epleted S ilicon O n I nsulator
FPGA	F ield- P rogrammable- G ate A rrays
GAA	G ate A ll A round
GND	G rou N D
HP	H igh P erformance
IC	I ntegrated C ircuit
IoT	I nternet O f T hings
ITRS	I nternational T echnology R oadmap for S emiconductors

LP	Low Power
LSP	Low Standby Power
LUT	Look Up Table
MEMS	Micro Electro-Mechanical System
MEOL	Middle End Of Line
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
ML	Match Line
NEMS	Nano Electro-Mechanical System
NV	Non-Volatile
NVM	Non-Volatile Memory
OxRAM	Oxide Resistive Access Memory
PCM	Phase Change Memory
PECVD	Plasma-Enhanced Chemical Vapor Deposition
PEALD	Plasma Enhanced Atomic Layer Deposition
PI	Pull-In
RIE	Reactive Ion Etching
RRAM	Resistive Random Access Memory
RWL	Re-Write Line
RWLB	Re-Write Line Bar
SEM	Scanning Electron Microscope
SL	Search Line
SN	Storage Node
SRAM	Static Random Access Memory
SS	Subthreshold Swing
STT-RAM	Spin Transfer Torque Random Access Memory
SOI	Silicon On Insulator
SPDT	Single Pole Double Throw
SPM	Scribe line Primary Marks
TCAM	Ternary Content-Addressable Memory
TEOS	Tetra-Ethyl Ortho-Silicate

ACKNOWLEDGEMENTS

I would like to thank the jury members, Dr. Laurent Pichon, Dr. Núria Barniol Beumala, Dr. Damien Deleruyelle, for accepting to review my manuscript and for their relevant remarks.

Before starting the thesis, I was young and not at all aware of the huge amount of work, strength, discipline, perseverance that a thesis such as mine required. My internship supervisor, which then became my thesis supervisor (Dr. Louis Hutin) convinced me to embark on this journey.

For this, I would like to thank him. He helped me to push all my limits, allowing me to discover that I was able to grow more and more professionally but also personally. He challenged me every day in becoming always my better self.

I would love to thank Claude Tabone; his sentence: ‘Les voies du Seigneur sont impenetrable’ became my mantra. Claude with 40 years of experience taught me thousands of fabrications tricks that cannot be learnt from books.

I would love to thank Thomas Ernst, my thesis director. He was very busy, however during my thesis he always could free up some time the time to talk to me and help me. Thank you for listening and giving me useful advices. You taught me that a good thesis work is not done only through research but is also done through time management and smart use of communication skills.

I would love to thank my co-director Tsu Jae King Liu. You are the busiest person I ever met and yet you found time to be part of my thesis. Thank you to showing me that it is possible to be Dean of Engineering at UC Berkeley, Board of Directors of Intel, wonderful teacher and supervisor, woman and mother. Your achievements are amazing, it was a privilege to meet you. I would love to thank the rest of the MADN/MD2N team: Christophe Plantier, Sebastien Hentz, Jose Munoz Gamarra and Emmanuel Roland. Our team was dynamic and fun. Each one of you had experience in different topics, each one of you had different skills and knowledge. Together we rocked 😊. I loved our Friday meeting where I learnt how to find the root of a problem, be more precise and mostly to never give up (and I also learnt many useful French expression!). I would love to thank the Jedi council: Alex, Vincent, Julien, Mathilde, Jessy, Yoon Ji, Benoit, Lina, Sotiris, Carlos, Remy, Thomas, and the Padawans: Camila, Daphne, Rony. Thankfully I had you on my journey, I have so many things to say to each one of you that I could write another manuscript. Therefore, I will just say: “Padawans do not give up. The thesis will finish sooner than you expect! May the force be with you”. I would love to thank Urmita Sikder and the rest of the Tsu Jae’s groups, working with you has enhanced this thesis work and seeing how research is done on the other part of the world has enriched me professionally. I would love to thank Bernard P. *Master of Expida*. You take wonderful pictures, my switches never looked better 😊!

I would love to thank LICL, LMDK, LCM, DCOS and DACLE labs during my thesis I had the pleasure to interact with many of you.

I would love to thank all my Grenobloise friends with a special thanks to my dance buddy Sana Rachidi.

Thanks to all my long date friends because you are always with me, supporting me and giving me your love even if we live faraway.

Thanks to my family who is always there for me, no matter where I am or what I am doing. I love you so much. You are my heart.

Thanks to my family in law who support me and fills me with affections.

Finally, I am grateful of having the best husband of the world: Giorgio. Thanks for holding me tight when I was feeling down and celebrating with me when I was feeling up.

Thank you to each one of you.

CONTENTS

Abstract	3
Résumé.....	4
Acronyms	5
Acknowledgements.....	7
Contents	8
1 Introduction	12
1.1 MOSFET miniaturization: Purpose and limits.....	13
1.2 Moore’s Law	13
1.3 The Conventional MOSFET	14
1.3.1 MOSFET Scaling Limits.....	17
1.3.1.1 Performance-Leakage Trade OFF.....	18
1.4 Memory Hierarchy	21
1.5 General Review of Beyond CMOS devices	24
1.5.1 Beyond CMOS Logic Devices.....	25
1.5.1.1 Tunnel Field Effect Transistors.....	25
1.5.1.2 Impact Ionization MOSFET IMOS.....	26
1.5.1.3 Negative Capacitance Field-Effect Transistor (Neg-Cg FET).....	27
1.5.2 Beyond CMOS Emerging Non-Volatile Memories	29
1.5.2.1 Resistive RAM (RRAM)	30
1.5.2.2 NEMS.....	33
1.5.2.2.1 Electromagnetic	34
1.5.2.2.2 Electrothermal	35
1.5.2.2.3 Piezoelectric	36
1.5.2.2.4 Electrostatic.....	37
1.6 Thesis GOAL	39

2	NEMS Relay: State of Art.....	41
2.1	Introduction	42
2.2	Nanoelectromechanical Switches.....	44
2.2.1	Design	44
2.2.2	NEM Relay Actuation Mechanism	48
2.2.3	Materials.....	50
2.2.3.1	Sacrificial Layer.....	50
2.2.3.2	Structural Layer.....	51
2.2.3.3	Contact and Coating layer.....	54
2.2.3.3.1	Coating Layer.....	54
2.2.4	Metal Interconnection Air Gap for NEM relay: Are They Going to Simplify the Release Process?.....	56
2.2.4.1	Interconnection stack Air Gap Fabrication	56
2.2.4.2	Air gap reliability issues	58
2.2.4.3	Conclusion	59
2.2.5	Nano-contact	59
2.2.6	Endurance.....	61
2.2.7	Mechanical Delay.....	65
2.2.8	Scaling limits.....	65
2.2.9	Harsh Enviroment	68
2.3	Target Applications	70
2.3.1	Logic NEMS	70
2.3.1.1	Nano-Electro-Mechanical relays for FPGA routing	71
2.3.1.2	Logic NEMS Relay in SRAM	72
2.3.2	Non-Volatile NEM relay.....	73
2.3.2.1	Look-up Table.....	74
2.3.2.2	NV-SRAM	76

2.3.2.3	NV-TCAM	77
2.3.3	Power Gating	78
2.3.4	Remaining Challenges	80
2.3.4.1	Challenges for Reducing Relay Size	80
2.3.4.2	Challenges for Reducing Operating Voltages	81
2.3.4.3	Challenges for Monolithically NEMS CMOS Integration	81
3	Modeling of NV NEM Relays	82
3.1	Design and Modeling of NEM Relays	83
3.1.1	Parallel Plate Capacitor Approximation	83
3.1.2	Modeling of Mechanical Beam	87
3.1.2.1	Stiffness	88
3.1.2.2	Cantilever and Clamped-Clamped Beam Stifness and Resonant Frequency	89
3.1.3	Mechanical Delay	91
3.2	Working Margin Optimization	92
3.2.1	Pull In Voltage Modeling For “Beam” NEM Relay	93
3.2.2	Catastrophic Pull In Voltage	98
3.2.2.1	Working Margin Optimization Summary	100
3.3	Design Optimization for Non Volatile NEM Relay	100
3.4	Contact Instability Due to Scalability	106
3.5	Case Study: NV SRAM	112
3.5.1	6-T SRAM	112
3.5.1.1	Storing-Writing-Reading Operations	112
3.5.2	1-Relay (R)-8T NoN-Volatile (NV) SRAM	113
3.5.2.1	Re-Writing Operation	115
3.6	Conclusion	118
4	Focus On 3D monolithic NEMS/CMOS Cointegration Process Flow	120

4.1	NEMS/CMOS Monolithic Co-Integration	121
4.2	c-Si NEMS ABOVE CMOS: Flow Overview	125
4.2.1	Preparation of The SOI Substrate	126
4.2.2	CMOS Substrate Preparation	129
4.2.3	Bonding, Grinding and Trimming.....	130
4.2.4	NEMS Patterning	133
4.2.5	Interconnections and Pads Definition	137
4.2.6	Nems Release	141
4.3	Fabricated Devices	143
4.4	I/V Characterization of Monocrystalline Stand-Alone NEMS relay	145
4.5	Conclusion.....	149
5	Conclusions and Perspectives	150
5.1	Conclusion.....	151
5.2	Perspectives	155
6	ANNEX 1	157
6.1	Poly-Si NEMS above CMOS	158
6.1.1	I/V Characterization of Polycrystalline Stand-Alone NEMS Relay	161
6.1.1.1	DC Characteristic	163
6.2	Cu BEOL NEMS.....	164
6.2.1	Switch Release	165
	Bibliography	170

1 INTRODUCTION

1.1 MOSFET MINIATURIZATION: PURPOSE AND LIMITS

1.2 MOORE'S LAW

Moore's law [MOORE 1965], which has been accepted by general consensus as a roadmap for the semiconductor industry for many decades states that: "The number of active transistors present in a single chip will double every 18 months". Indeed, the number of transistors integrated in a single chip has increased every year owing mainly to technology progress in device miniaturization. One of the messages of Moore's law is that the cost per transistor decreases with each new technology generation. This law also highlights the fact that, for each generation, there is an optimal number of components per integrated circuits that results into a minimum cost. Further increasing the number of components per circuit above that optimal number increases chip area and reduces the process yield (increasing the final cost). Since 1975, the semiconductor industry followed Moore's trends, and miniaturizations allowed to place billions of transistors on a single chip see Fig. 1.

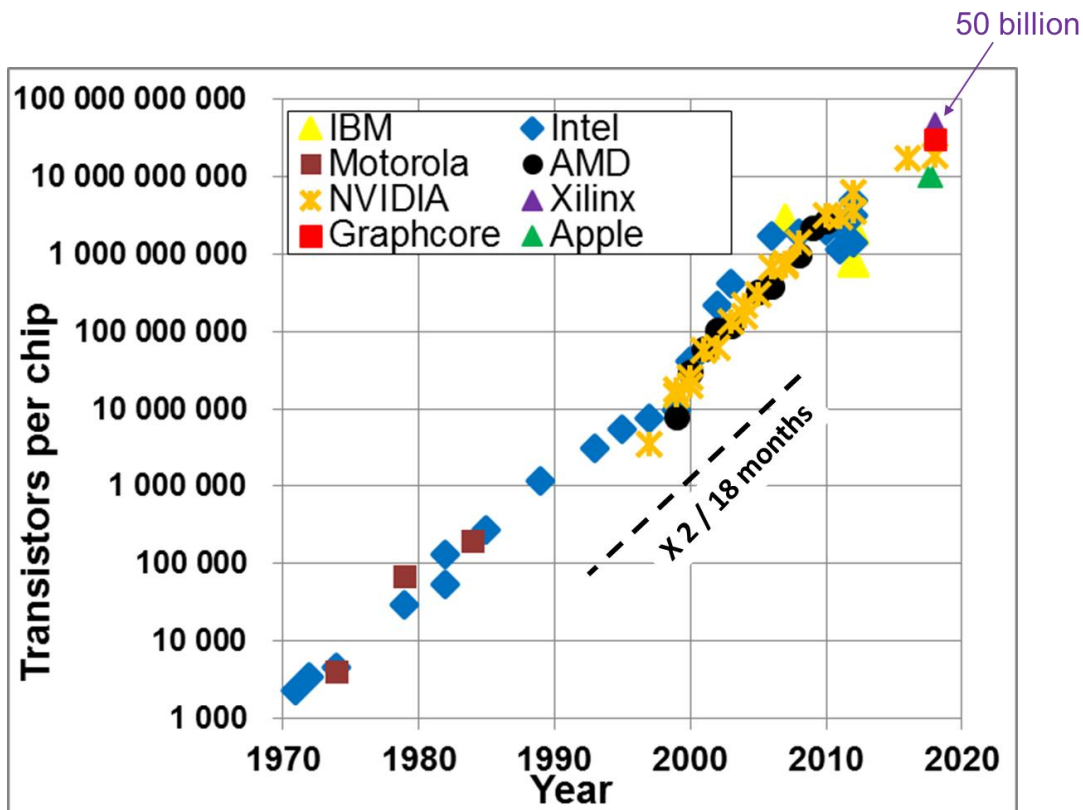


Fig. 1: Plot of transistor count vs year of introductions [Colinge 2016]

Moore's law drove the semiconductor market for at least 52 years. In 1965, there were no fundamental physical limits limiting transistors scaling. Once the transistor size reached the nano scale, however, parasitic effects started to appear, causing performance degradation and increasing device-to-device variability. Therefore, keeping extending Moore's Law requires working on innovation strategies. For example, to increase the channel mobility, strain and new materials such as SiGe have been introduced. The gate stack has been modified and the classical SiO₂/polysilicon gate stack has been substituted with high-k gate dielectric and metal gate to scale the effective oxide thickness (EOT). To improve the channel control new transistors structures have been proposed (*e.g.* FinFET, Gate All Around (GAA)). Additionally, new device solutions (*i.e.* beyond CMOS devices), new computational methods and new integration methods (*e.g.* 3D monolithic integration) have been suggested to help to keep Moore's law going.

1.3 THE CONVENTIONAL MOSFET

The key component of the integrated circuits, so far, has been the Metal–Oxide–Semiconductor Field-Effect Transistor (MOSFET) see schematic view in Fig. 2. The MOSFET has four electrodes, the gate, which controls the channel, the body, which connects the semiconductor to ground, the source and the drain. The transistor is called n-type or p-type depending on the dopant type implanted on source and drain, usually opposite to the channel dopant type. An oxide of thickness t_{ox} separates the gate from the channel region, which has a physical length L_G and a width w . When a voltage difference V_{GS} is applied between gate and source, an electric field is created to control and modulate the current density in a channel formed between drain and source. The carriers in the channel can be holes or electrons depending on the transistor type (respectively p-type or n-type).

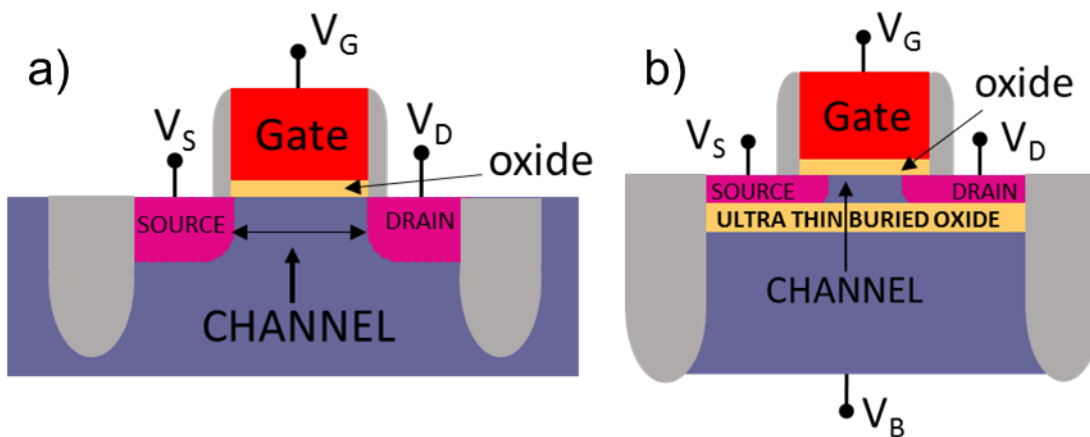


Fig. 2: Cross sections of MOSFET device a) typical Bulk MOSFET; b) Fully Depleted (FD)-Silicon On Insulator MOSFET with 4 terminals.

For logic applications, the ideal MOSFET electrical characteristics should be similar to those of the *ideal* switch shown in Fig. 3 a). If the voltage difference applied between source and gate V_{GS} is smaller than a certain threshold voltage V_{TH} , no channel is formed and there is no current flow between source and drain ($I_{OFF} \cong 0$ in a first approximation). Once $V_{GS} > V_{TH}$ a current path (channel) between source and drain is formed ($I_{ON} \neq 0$). In an ideal switch, $I_{OFF} = 0$ and the resistance of the channel is negligible ($R_{ON} = 0$). However, real MOSFETs always present finite I_{OFF} see Fig. 3 b). This OFF current is called the leakage current. The value of I_{OFF} depends on the temperature (diffusion current), on short channel effects (*e.g.* DIBL and V_{TH} shift), on ultra-short channel effects, on direct tunneling and on the junctions quality [CHEN 1987].

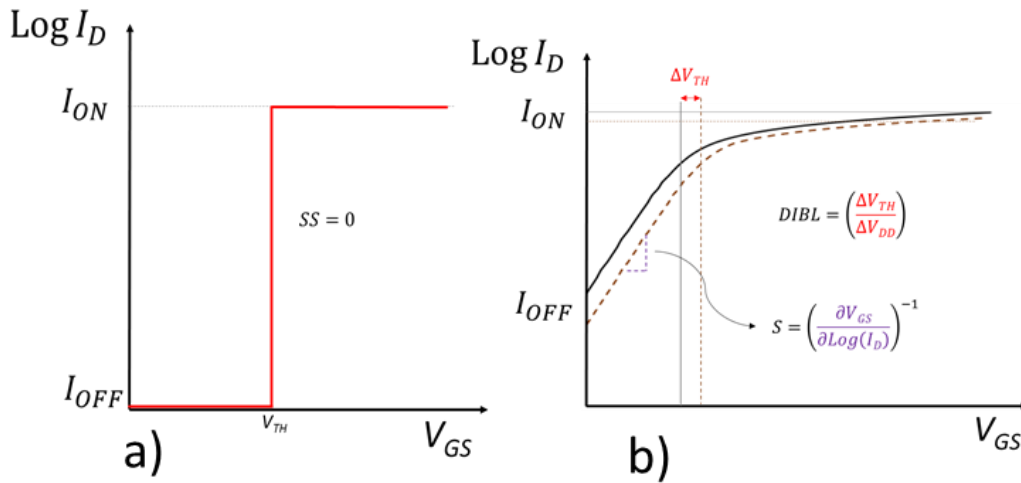


Fig. 3: **a)** Characteristic curve of an ideal switch: the switch presents a steep slope. **b)** n-MOSFET Characteristic curve, drain current vs. V_{GS} showing the subthreshold slope and the drain induced barrier lowering

The various MOSFET technologies are benchmarked with some merit factors that are listed below:

I_{OFF}	Leakage current
I_D	Drain current (corresponding to the saturation current I_{ON} for logic application)
$DIBL$	Drain-Induced Barrier Lowering
SS	Subthreshold Swing (S^{-1})

The subthreshold swing SS (inverse subthreshold slope) is defined as the change in gate voltage required changing the subthreshold drain current by one decade. In a bulk transistor, it is given by the following expression:

$$SS = \frac{\partial V_g}{\partial \psi_s} \times \frac{\partial \psi_s}{\partial \log I_{DS}} = m \times n = \frac{kT}{q} \ln 10 \times \left(1 + \frac{C_{dep}}{C_{ox}}\right) \quad Eq. 1$$

Where ψ_s is the potential at Si/insulator interface, m is called body factor, which depends from C_{dep} the depletion region capacitance, and C_{ox} the gate-to-channel (oxide) capacitance in MOSFET, $m = \frac{kT}{q} \ln 10$, k is the Boltzmann constant, T is the absolute temperature. In some devices such as SOI transistors or GAA MOSFETs, C_{dep} is replaced by the capacitance between the channel and ground, which can be much smaller than C_{ox} , in which case the body factor n can have a value close to 1.

A lower value of SS means that a lower gate voltage swing is required to turn the transistor from the OFF to the ON state (or: to reach V_{TH} for a given I_{OFF}). The subthreshold swing has a theoretical minimum value of 60mV/dec. for a conventional silicon device, at $T=300K$ [GODOY 2001].

1.3.1 MOSFET SCALING LIMITS

Moore's law reigned for more than 50 years, and each 18 months the semiconductor industry launched a new generation labeled by a technology node (specific manufacturing process and design rules). The technology node numerical value corresponds to the MOSFET gate length L_G until node 32nm. Nowadays, the node numerical name is chosen with respect marketing strategies and its value does not have to correspond to physical gate length see Fig. 5.

Since 1974, transistor scaling was driven by employing "Dennard scaling rules" that required to maintain a constant electric field in the transistor [DENNARD 1974] see Fig. 4. Those scaling rules allowed to increase transistor performance, and to reduce power consumption while decreasing linearly dimensions and supply voltage. In other words, they allowed maintaining power density constant.

TABLE I
SCALING RESULTS FOR CIRCUIT PERFORMANCE

Device or Circuit Parameter	Scaling Factor
Device dimension t_{ox} , L , W	$1/\kappa$
Doping concentration N_a	κ
Voltage V	$1/\kappa$
Current I	$1/\kappa$
Capacitance $\epsilon A/t$	$1/\kappa$
Delay time/circuit VC/I	$1/\kappa$
Power dissipation/circuit VI	$1/\kappa^2$
Power density VI/A	1

Fig. 4: Scaling relationship to maintain power density constant while improving performance. [DENNARD 1974]

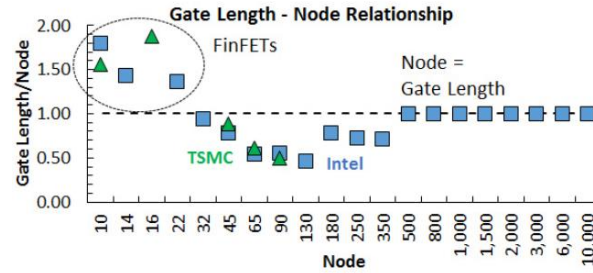


Fig. 5: Gate length vs technology node showing that scaling the equation gate length=node is no longer valid. [JONES 2017]

However, as the transistor size reached sub- μm dimensions undesirable effects, such as short channel effects started appearing Fig. 6. The short channel effects appear when the channel length is comparable with the depletion layer width of source and drain junctions, which results in lowering the threshold voltage. Additionally, if a high voltage is applied to the drain the barrier between drain/channel junctions is further reduced (Drain-Induced Barrier Lowering) reducing the electrostatic control of the transistor.

Further, in FDSOI and FinFET technologies, when the channel thickness reaches $\sim 7\text{nm}$ dimensions, quantum effects such as carrier confinement or tunneling start to influence the transistor behavior.

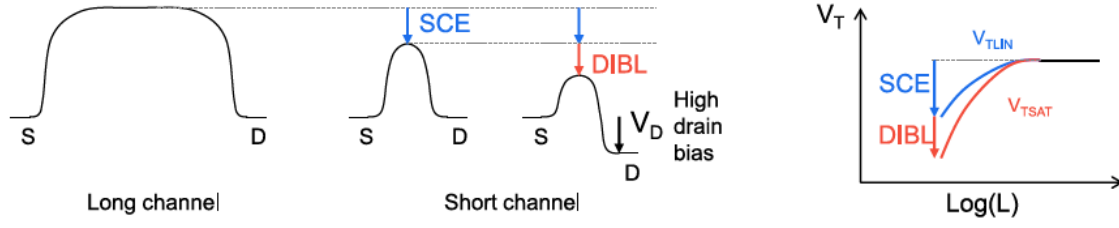


Fig. 6: Schematic showing the short channel effect and the DIBL due to the channel length downsizing figures from [BERTHELON 2018]

1.3.1.1 PERFORMANCE-LEAKAGE TRADE OFF

The CMOS power consumption per clock cycle can be divided in two components Eq. 2: Dynamic power consumption (Eq. 3), proportional to activity factor α , input signal frequency f , load capacitance C_{LOAD} ; and static power consumption, proportional to leakage current I_{leak} and supply voltage V_{DD} . The static power consumption represents the power consumed when the circuit is in standby.

$$P_{consumptions} = P_s + P_{dyn} \quad \text{Eq. 2}$$

$$P_{dyn} = C_{LOAD} * V_{DD}^2 * f_c * \alpha \quad \text{Eq. 3}$$

$$P_s = V_{DD} * I_{leak} \quad \text{Eq. 4}$$

For simplifications, we can consider that the *subthreshold leakage* is the only source of leakage (Eq. 5), since it is the main contributor to the MOSFET off-state current I_{OFF} . This current corresponds to the current measured on the drain when the transistor is OFF ($V_{gs}=0$) and $V_{DS}=V_{DD}$.

$$I_{leak} \approx I_0 e^{\frac{-V_{TH}}{SS}} \quad \text{Eq. 5}$$

I_0 is the current at threshold $V_G=V_{TH}$. Thus, the leakage current I_{OFF} is limited by the subthreshold swing at given V_{TH} and I_0 . For a chip with 100 million transistors, the average I_{leak} of turned-off devices should not exceed a few times 10nA, which corresponds to a total current of 1A and a standby power dissipation of 1 watt if $V_{DD}=1V$. In practice, because the subthreshold slope cannot be decreased below 60 mV/dec., the threshold voltage must be kept to values higher than approximately 0.3V at operating temperature [KISHORE, 2013].

Reducing dynamic power consumptions requires reducing V_{DD} , and while reducing V_{DD} , to maintain speed and, therefore, a high on current, it is necessary to reduce V_{TH} together with V_{DD} .

$$f_{perf} \propto \frac{I_{ON}}{C_{LOAD}} \propto \frac{(V_{DD} - V_{TH})^2}{C_{LOAD}} \quad Eq. 6$$

However, scaling V_{TH} results on an exponential increment of leakage, according to Eq. 6, augmenting static power consumption. Hence, decreasing dynamic power consumption increases static power consumption. Therefore, there is a limit in total power dissipated per operation for any CMOS based digital circuit.

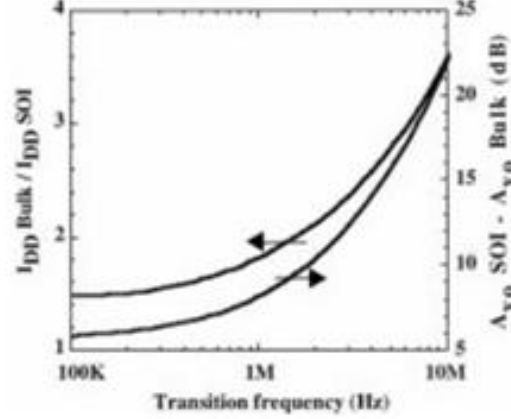


Fig. 7: Example of comparison of simulated total current dissipation and DC open-loop gain performance of the bulk and SOI cascoded CMOS OTAs as a function of the transition frequency. [FLANDRE 1999]

For the reason above, CMOS technology can be found in various flavors. Sorted by leaking/performance trade-offs, *i.e.* CMOS can be optimized for High-Performance (HP), Low Power (LP) and Low Standby Power (LSP) applications (see Fig. 8).

- **Increasing Performance:** *e.g.* for high-performance computing, CPU speed comes at the expense of power consumption.
- **Decreasing Power Consumptions:** application that requires Low Standby Power LSP, are for example mobile, IoT sensors, or Low Operating Power LOP devices, *e.g.* smart phone processors.

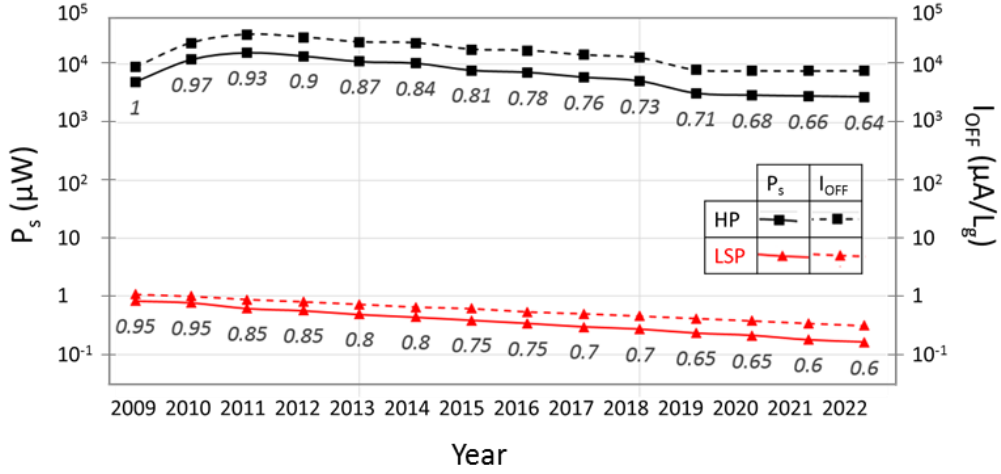


Fig. 8: Single device static power consumption (P_s) and leakage current VS year for High Performance and Low Standby Power application, data adapted from IRDS 2017 Reports.

Another problem of scaling is an overall increase of delay at the circuit level (to not be confused by the delay at the device level). Scaling increases the transistor density by reducing physical dimensions; metal line width and thickness is also reduced, but the length of global interconnects is not. As a result, the metal interconnection delay (RC) keeps increasing with scaling and interconnect delay becomes dominant over the gate delay (Fig. 9). Therefore, the global circuit delay is degraded due to the dominance of the capacitance of the metal lines. To address this challenge, copper has replaced aluminum [INFOWORD, 1997] and, to decrease the capacitance coupling between metal lines air gaps are introduced [SHIEH 1998]. However, interconnects delay nowadays, still limits the circuit performance. Monolithic 3D integration processing, such as CoolCube™, can help in reducing the interconnect delay by shortening the overall metal interconnection length. The transistor performance can be maintain using a less aggressive transistor scaling, but the transistor density can be preserved or improved by stacking devices on the top of each other.

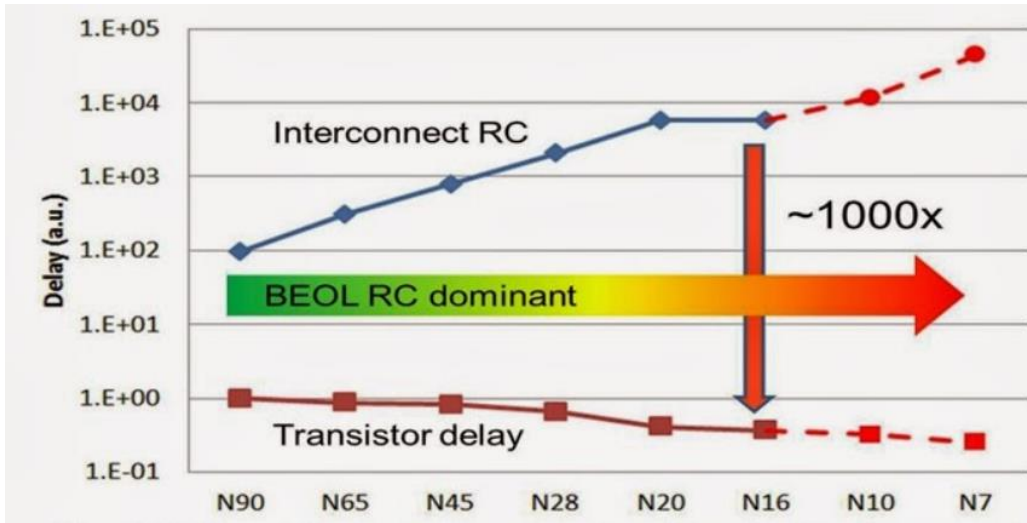


Fig. 9: Projection of on-chip electrical interconnect delays with technology scaling [Yeap 2013]

Despite the power challenges presented above, there still are many opportunities to optimize and to extend the life of CMOS for quite some time. However, the minimum energy point in CMOS between active and static energy consumption is obstacle in the way of achieving ultra-low power and energy-efficiency circuits. One way to improve this minimum energy point is to decrease the subthreshold slope of the switching device. Several devices proposed as alternatives have demonstrated a subthreshold swing below 60mV/dec.

Among these devices, we will focus on nanoelectromechanical relay that shows a competitive energy-efficient technology. NEMS relays exhibit abrupt turn-on/turn-off behavior and can operate at very low voltages (*e.g.* ~50mV [OSOBA 2016]) providing for low dynamic power consumption. NEMS relay also exhibits zero I_{OFF} (leakage current), which eliminates static power consumption.

1.4 MEMORY HIERARCHY

The scaling race has created a large gap between the central process unit CPU access time and speed (<1ns and 10GB/s) and out of CPU memory (*e.g.* Dynamic Random Access Memory DRAM ~10-30ns, Flash ~1ms). In general, memory improvements have mostly been in terms of data density (Fig. 10), rather than in terms of read/write data bandwidth (transfer rates), as shown in Fig. 11 and accordingly a “memory hierarchy” has been established.

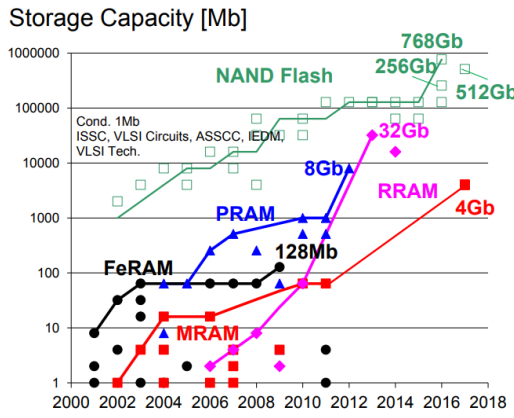


Fig. 10: Memory capacity trend of emerging non-volatile memories from [ISSCC 2017]

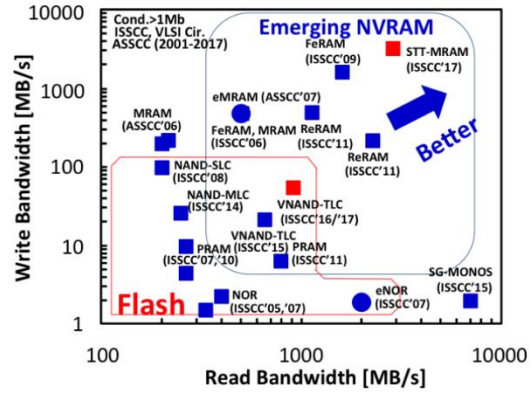


Fig. 11: Read/write bandwidth comparison of non-volatile memories [ISSCC 2017]

Fig. 12 shows the memory hierarchy, which is established considering size (bytes), access time and memory type. Designing computers for high performance requires considering the memory restrictions showed by the memory hierarchy, *i.e.* the size and capabilities of each component. CPU are very fast ($>2.5\text{GHz}$) but need to spend much of their time idling, waiting for memory I/O transfers to complete. This bandwidth gap between processor and memories it is known as the Von Neumann bottleneck [BACKUS 1977]. For example, there is a large gap between Flash memories and DRAM. Flash main issues are latency and low write bandwidth, as shown in Fig. 11. In addition, typical endurance parameter for FLASH is about 10^6 cycles whereas DRAM requires 10^{15} cycles endurance [PALMA 2013]. Another limitation is the high operating voltage of FLASH memories that is a problem for advanced CMOS node. As result, novel memory candidates are under investigations. With lower latency, higher bandwidth and process compatibility with CMOS technologies.

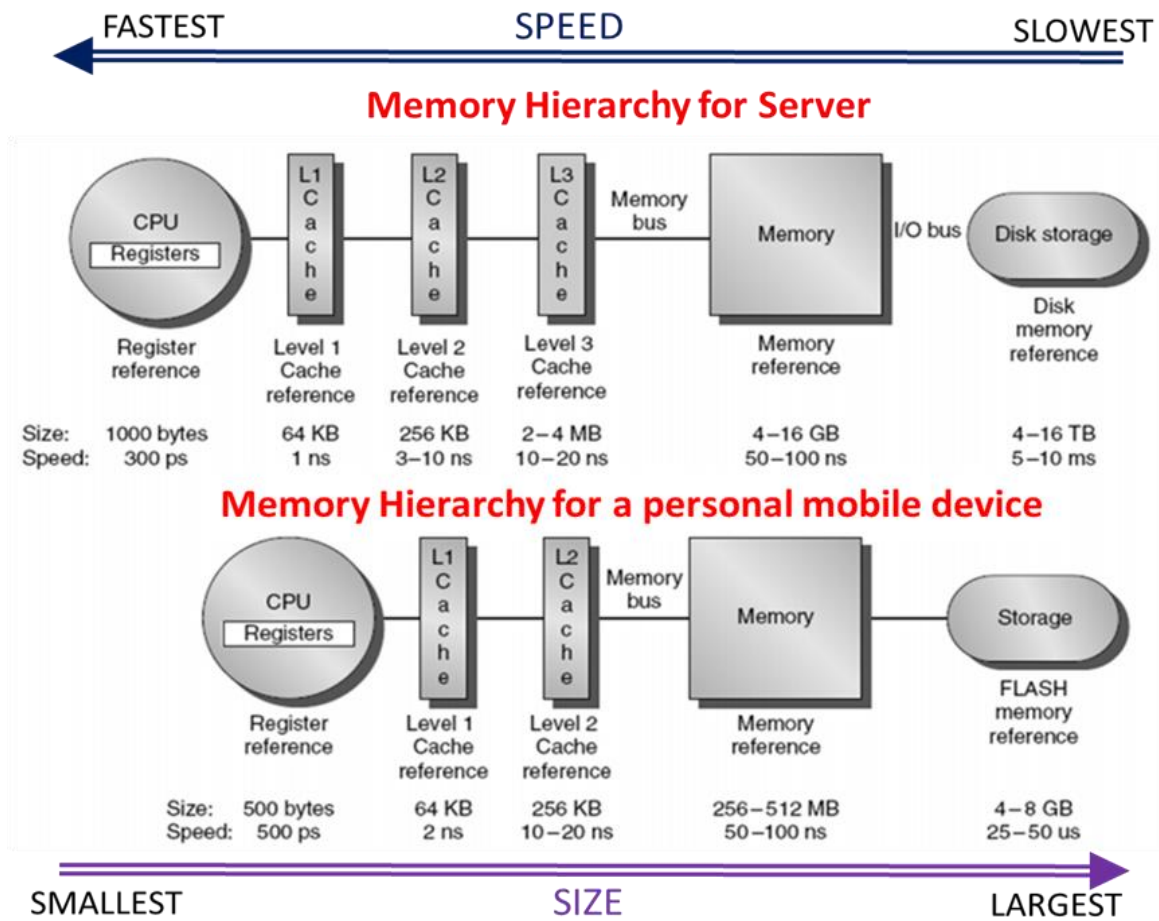


Fig. 12: Schematics of Memory hierarchy system for server and personal mobile device, adapted from [Silvano 2012]

1.5 GENERAL REVIEW OF BEYOND CMOS DEVICES

The International Technology Roadmap for Semiconductors 2.0 (ITRS) has identified devices with the potential capability to achieve SS below 60 mV/dec., those devices are in the categories of “Beyond-CMOS” devices [2013 ITRS]. ITRS has also identified nonvolatile memory devices that have high performance and robustness, that are nonvolatile and have a low cost per bit. Some of the beyond CMOS devices and memories are prototypical and some are still under study for both existing and new functions to extend the IC scaling. Beyond CMOS technologies have to increase efficiency, functionality and performance of the integrated circuit. Those devices, to enter on the productions have to fulfill some conditions: firstly, they have to be competitive with the MOSFET or have to introduce new functionality to the MOSFETs circuitry. Secondly, Process fabrication of these steep slope devices must be compatible with the co-integration of MOSFETs and monolithic 3D power scaling.

In Fig. 13 are listed two classes of beyond CMOS devices, the logic and the nonvolatile memory. For logic, one of the major challenges is to reduce the subthreshold swing, in order to reduce the power consumption, targeting ultra-low power applications. For memories, the challenges are that they should be scalable above the current limits of SRAM and Flash, CMOS process flow compatible, and should combines the best features of current memories: non-volatility with high speed and high density.

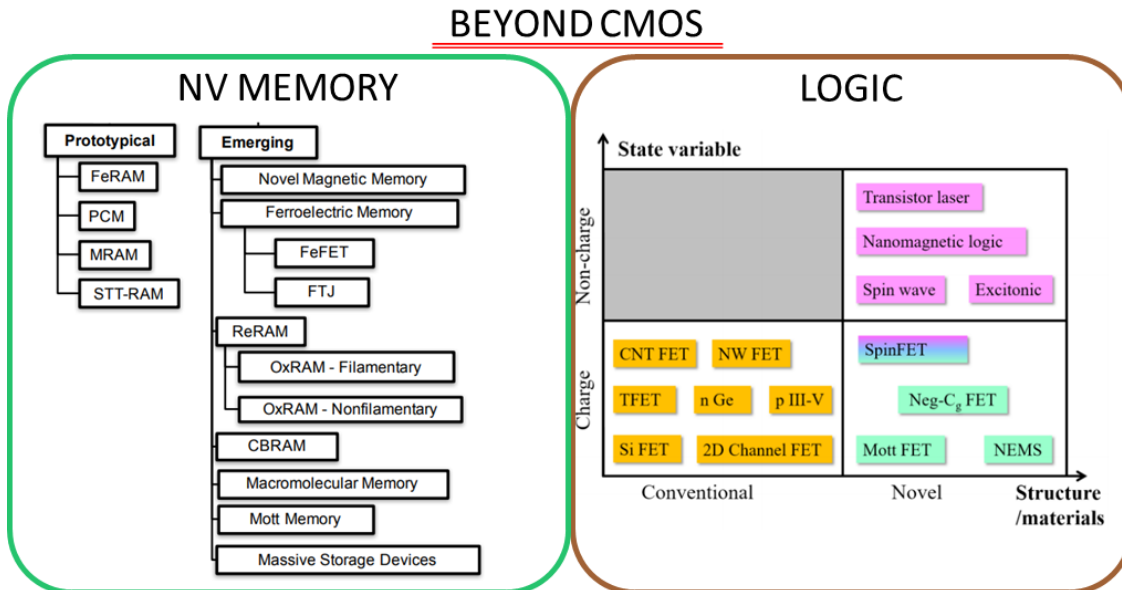


Fig. 13: Overview of the leading Beyond CMOS memories and logic devices adapt from [IRDS 2017]

1.5.1 BEYOND CMOS LOGIC DEVICES

Ultra-low power applications need devices capable to abruptly switch ON and OFF without spending a great amount of supply voltage. The figure of merit SS gives an immediate information about this capability because it indicates how much voltage is needed to increase the I_{ON} current one decade (see Eq. 1 for MOSFET). For beyond CMOS devices Eq. 1 is no longer valid, either because of their different structure or because different actuation mechanism and/or current path/emission with respect the MOSFET.

Device such as Tunnel FET and Impact Ionization MOS have a smaller n factor than the MOSFET because they are not driven by thermionic emission and they are not limited by Maxwell-Boltzmann distribution. Devices such as negative capacitance Neg-Cg and nanoelectromechanical NEM relays have a very steep slope, with body factor below unity.

However, lowering SS is not enough to be completely competitive with MOSFET, it is also necessary to be competitive in terms of cost and in terms of speed and scaling for a given applications. In the next paragraphs a brief description of beyond CMOS steep slope devices, their operation methods and physics is given.

1.5.1.1 TUNNEL FIELD EFFECT TRANSISTORS

The tunnel Field-Effect Transistor (FET) switch mechanism relies on the modulation of band-to-band tunneling to inject carriers and obtaining an I_{ON} current. Therefore, TFETs can potentially have a subthreshold swing smaller than 60mV/dec. [REDDICK 1995]. Their physical structure is similar to that of a MOSFET but the source and the drain are differently doped (*e.g.* in homojunction TFETs the source is N doped for P-TFETs and it is P doped for an N-TFET, see Fig. 14 (left). The TFET lowest measured sub-threshold swings approach 20 mV/decade, however this result is obtained for a narrow range of subthreshold currents [TOMIOKA 2012]. Additionally, in TFETs, low subthreshold slope it is not always accompanied by high ON current comparable to the CMOS I_{ON} . The highest TFET I_{ON} current is observed at 1nA/ μ m, but for a $SS=60$ mV/dec. [DEWEY 2011]. Otherwise, for $SS<60$ mV the highest I_{ON} current is ~ 0.1 nA/ μ m [LU 2014] see Fig. 14 (right).

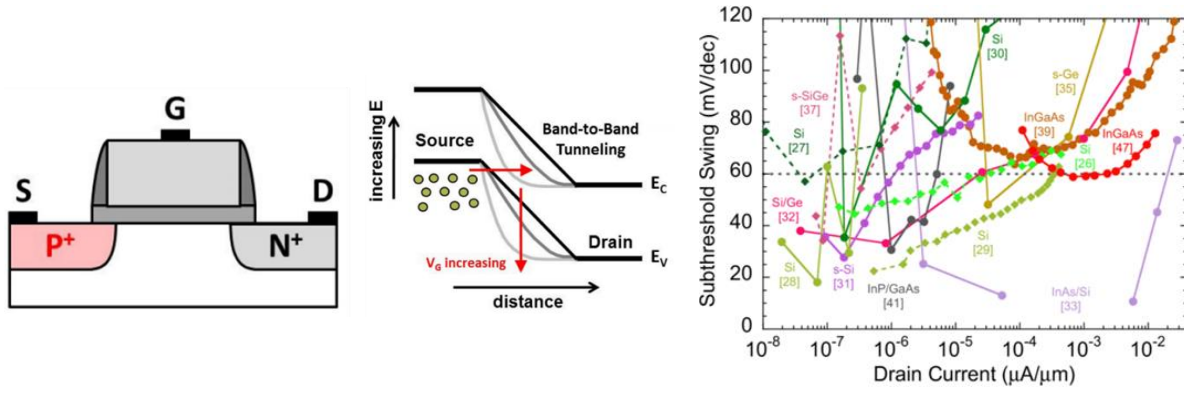


Fig. 14: (Left) Schematic cross section and operation mechanism of TFET, the turn on operation is based on gate-voltage modulation of the channel potential with carrier injection through the potential barrier via band to band tunneling [Kim 2012]. (Right) Comprehensive plot showing TFET sub-threshold swing Vs I_D (drain current) per unit width for TFETs having SS below 60 mV/decade [Lu 2014].

The discrepancy between the measured characteristics and the simulated curves is due to the fact that in the simulation many secondary interface effects are neglected. For example, a typical neglected secondary effects is the trap-assisted tunneling (TAT) which is an interface and temperature dependent effect causing band to band tunneling even when the device is supposed to be off, resulting on generation of currents degrading the subthreshold slope [DIAZ LLORENTE 2018].

1.5.1.2 IMPACT IONIZATION MOSFET IMOS

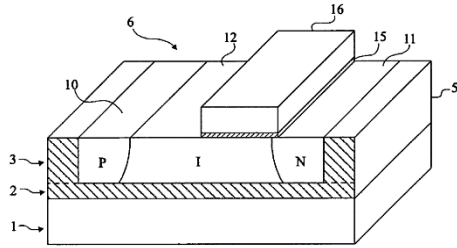


Fig. 15: Schematic cross view of a IMOS-type transistor [Charbillet 2006]

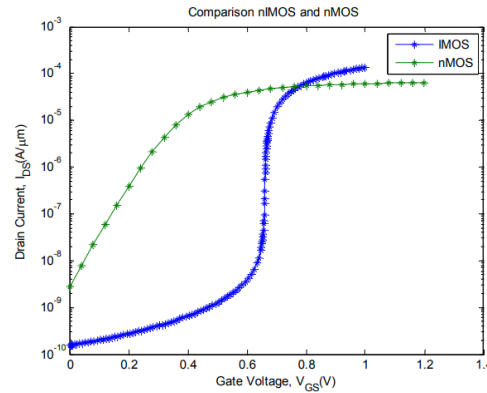


Fig. 16: I_D versus V_{GS} characteristic for an n-channel IMOS and n-MOS [SENG 2012A]

The impact-ionization field effect transistor (IMOS) modulates voltage breakdown of a gated p-i-n structure in order to turn ON and OFF the transistor, using impact ionization and avalanche breakdown [GOPALAKRISHNAN 2002]. The avalanche breakdown is a form of electric current multiplication enabling a steep increase of carrier multiplications (high I_{ON} current). The avalanche process occurs when the electric field, due to the gate voltage, accelerates source carriers until their energies is sufficient to create mobile or free electron-hole pairs via collisions with bound

electrons of the intrinsic region. Very steep slopes have been demonstrated: on silicon-on-insulator wafers (SOI), using a standard CMOS process, a subthreshold swing of 6mV/dec. and 2 mV/dec. respectively for n and p I-MOS [MAYER 2006]. The challenges are that the avalanche breakdown requires very high electric field leading to high drain bias ($\sim 20\text{V}$) [GOPALAKRISHNAN 2002]. Additionally, IMOS devices are intrinsically slow due to the statistical avalanche charge multiplication process, and speed limitations due to carrier multiplication delay and statistical retardation delay. There is also the problem that the hot carrier injected in the gate dielectric causes a voltage threshold and SS degradation.

1.5.1.3 **NEGATIVE CAPACITANCE FIELD-EFFECT TRANSISTOR (NEG-CG FET)**

The majority of dielectrics have a linear relation between the electric displacement field \vec{D} and electric field \vec{E} (Eq. 7), where ϵ_0 is the vacuum permittivity and ϵ_r is the relative permittivity. The permittivity is also related to the susceptibility $\epsilon_r = (1 + \chi)$ that is the proportional factor between the electric field and polarization density \vec{P} of individual particles on the material Eq. 8 and Eq. 9. In paraelectric materials, the permittivity is not linear, and it depends on the strength of the electric field as Eq 9.

$$\vec{D} = \epsilon_0 \epsilon_r \vec{E} \quad \text{Eq. 7}$$

$$\vec{D} = \epsilon_0 (1 + \chi) \vec{E} = \epsilon_0 \vec{E} + \vec{P} \quad \text{Eq. 8}$$

$$\epsilon_r = \frac{\vec{P}}{\epsilon_0 \vec{E}} \quad \text{Eq. 9}$$

Ferroelectric materials, besides being paraelectric, have an intrinsic or spontaneous nonzero polarization even when no electric field is applied. Applying a strong electric field on the opposite direction can reverse this spontaneous nonzero polarization. The final polarization depends on the value of the previous polarization and on the applied electric field, giving a hysteresis loop. Fig. 17 shows the polarization as a function of the applied electric field, for the case of a linear polarization, paraelectric polarization and ferroelectric polarization, where the permittivity is the slope of the $P(E)$.

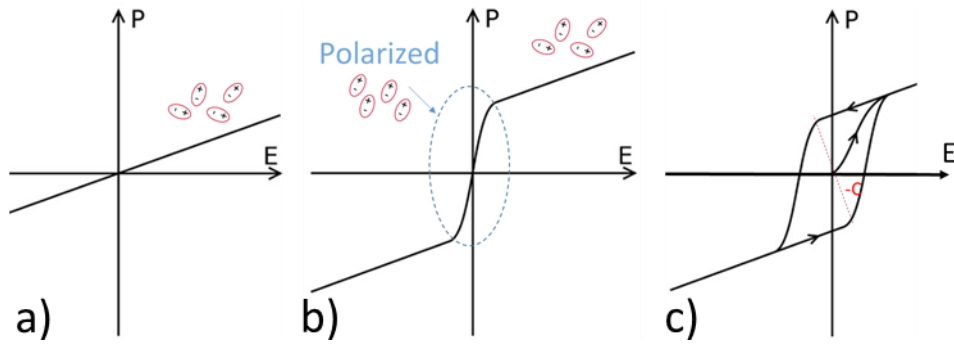


Fig. 17: a) Linear Polarization b) Paraelectric Polarization c) Ferroelectric Polarization

Thanks to their behavior, ferroelectric materials can be inserted in the gate stack of MOSFETs, between the gate and the semiconductor channel to create a “negative capacitance” effect [APPLEBY 2014; CATALAN 2015] in order to obtain a subthreshold swing below $<60\text{mV/dec}$. [LEE 2017; LI 2015]. The negative capacitance [SALAHUDDIN 2008] allows one to boost and increase the surface potential for a same voltage applied to the gate, allowing obtaining a subthreshold swing smaller than 60mV/dec . (see Eq. 10), being C_{FE} the negative ferroelectric capacitance. The fact that the surface potential increases faster than the gate voltage allows creating a large amount of carrier and a higher current compared to the MOSFET.

$$SS = \left(1 + \frac{C_s}{C_{FE}(< 0)}\right) \times \frac{kT}{q} \quad \text{Eq. 10}$$

Sub-20-mV/decade Subthreshold Slope have been demonstrated [KO 2017]. One of the challenges of this negative capacitance transistor is that sometimes the ferroelectric material stays polarized inducing a shift on V_{TH} and a hysteresis effect in the I_D vs V_G curve. Consequently, a big challenge resides in the identification of appropriate materials (oxides and ferroelectrics) for the best swing with minimal hysteresis. Even if recently was reported a small hysteresis window with V_{TH} shift below 0.1V see Fig. 18. Integrating high-quality single-crystalline ferroelectric oxides on silicon stays a challenge [MIKOLAJICK 2018].

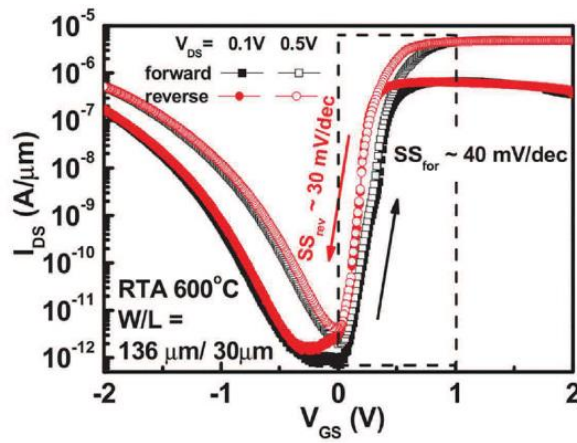


Fig. 18: Transfer characteristics (I_{DS} vs V_{GS}) of FE-HZO FETs with a hysteresis window < 0.1 V for 600°C annealing [Lee 2015].

1.5.2 BEYOND CMOS EMERGING NON-VOLATILE MEMORIES

Usually, systems-on-chip for low-power applications separate volatile and non-volatile memories in two different macros. The nonvolatile macro is made using nonvolatile memories NVMs such as Flash. NVMs generally require a high voltage and have high dynamic power consumption during read/write operations. Usually NVMs are slow and have limited endurance. However, the limited endurance is not a showstopper for the NVs macro. Especially because of the FLASH serial peripheral interface SPI which resulting on sequential access of data, relaxing endurance requirements. The positive aspect of these two macros approach is that NVMs enables to switch off the power supply and to suppress standby power (e.g. in a mobile chip), thereby extending battery life without a loss of data. The negative aspect is the long power off-on time that requires power intensive task caused by the transfer operation between the two macros limiting the battery lifetime. The volatile macro is usually made by Static Random Access Memory SRAM that is the best candidate to achieve high performance with low voltage and high endurance. In fact, SRAM provides on-chip storage, high speed read/write operations and a low minimum supply voltage (e.g. 5.36GHz for 1.115V @7nm [CLINTON, 2018]). Unfortunately, SRAMs are unable to preserve data when the power supply is turned off.

Nowadays, many applications are very data intensive, requiring both real-time and offline manipulation of great amounts of data. Extended battery life and rapid, reliable power on/off operations are very important for mobile applications (e.g. smartphone, smartwatch). Most of the time, those applications have relaxed workload, but need a high performance burst right after being powered up. In addition, many wireless electronics or bioelectrical devices and sensor applications

are usually in stand-by or in deep sleep mode and needs to quickly activate to perform few necessary tasks before returning to an inactive state. Dynamic voltage scaling (DVS) is a popular approach to suppress active-mode standby power and dynamic power by adjusting the operating voltage but it does not completely suppress the stand-by-power consumption Fig. 19 left. Moreover, when the standby time is long the standby power becomes significant Fig. 19 right. The two-macro solutions eliminate the need of stand-by power but create the need of additional power to transmit the data.

From here the idea or the necessity to combine high performing SRAM with NVMs having ultra-low standby power, obtaining NV-SRAM. The main idea is to integrate the NV element at the standard cell level, in such a way that the store and restore operation can be done bit-to-bit, which allows lowering store and restore time, memory latency time, and avoiding data retention failure due to powering off the volatile memory.

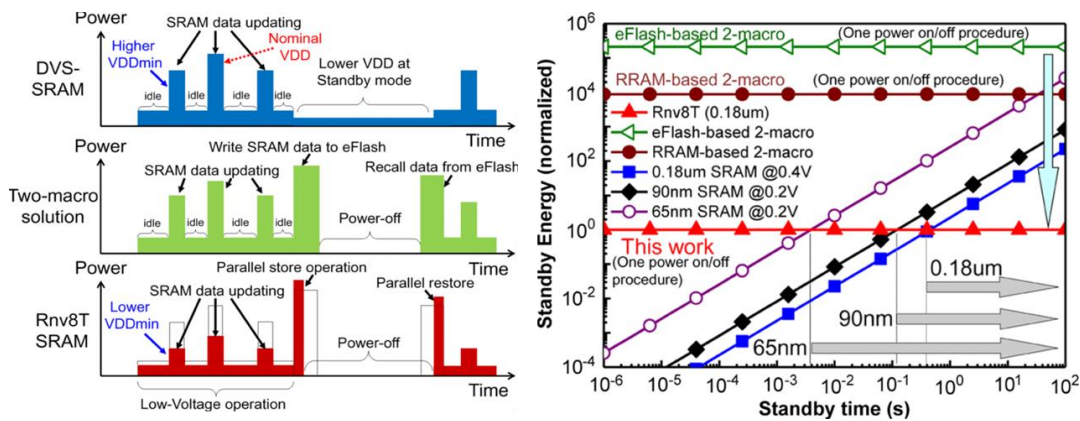


Fig. 19: (Left) Power consumption of Dynamic voltage scaling SRAM, two macro solutions, and NV resistive SRAM and (Right) Comparison of standby mode energy consumption [Chiu 2012]

Many Beyond CMOS NV memories could be used for NV-SRAMs, such as magnetic spin transfer torque memory (STT-RAM), phase change memories PCM and FeRAM which are prototypical. But there are also emerging beyond CMOS NV memories that need further investigation such as Resistive RAM (RRAM) (*e.g.* Oxide based RAM OxRAM, Conductive Bridging RAM CBRAM) and nanoelectromechanical NEM switches. Example of Resistive NVSRAM cell can be found in [CHIU 2012] and example of RRAM based ternary content addressable memory (TCAM) can be found in [GROSSI 2018]. Within the scope of this thesis, NV NEMS/SRAM and TCAM cells have been investigated and fabricated.

1.5.2.1 RESISTIVE RAM (RRAM)

In general, the term Resistive RAM includes all the memories that are made by two electrodes that sandwich a resistive layer. There are two main families of RRAM:

The Oxide RAM (OxRAM) and the Conductive Bridge RAM (CBRAM). OxRAMs are based on filamentary switching made by oxygen vacancies while CBRAM are based on metal ion filamentary switching, as shown in Fig. 20 (left). The standard operating behavior of a RRAM cell is composed of three different states. A highly resistive state, which corresponds to the resistance state just after processing, a low resistive state (LRS) or written state, that occurs once a path between the two electrodes into the resistive layer is created (filament formation) and a high resistive state (HRS) that occurs once the path is dissolved (filament dissolution). The conductive path or conductive filament is created at first with a forming operation at high voltage. RRAM can be also classified by their switching mechanism. In fact, there are two main switching modes: the unipolar and the bipolar modes. The unipolar switching direction depends on the amplitude of the applied voltage but not on its polarity, meaning that SET and RESET can happen on the same polarity. On the contrary, for bipolar switching the SET can only occurs at one polarity of the applied voltage and RESET to the reverse one Fig. 20 (right). OxRAM Bipolar switching can be achieved with 100uA compliance current, after initial forming at $\sim 2.8V$ [AZZAZ 2015]. TaOx and HfOx are the most common switching materials used for OXRAM. For the CBRAM the filament or conductive bridge is usually made by Cu/Ag ions, rather than oxygen filament (e.g. of schematic and TEM cross section of OxRAM in Fig. 21 a) and b)).

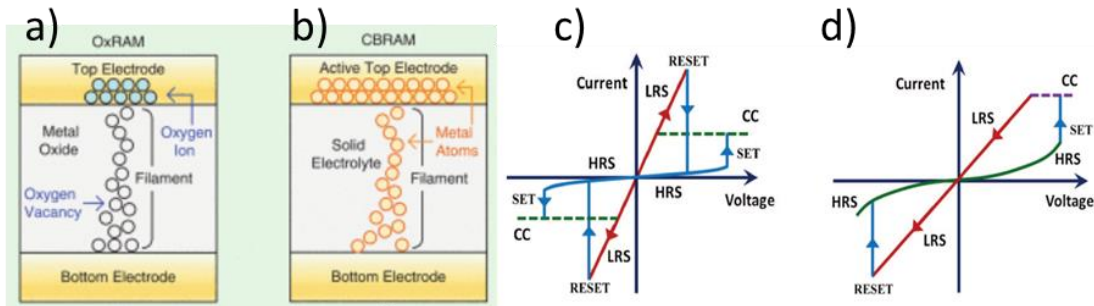


Fig. 20: (Left) A schematic RRAM based on oxygen vacancies in the filament in the oxide, referred to as OxRAM (a). RRAM based on metal ions diffusion from active electrode to form conductive bridge in solid electrolyte, referred to as CBRAM b). Adapted From [Yu, 2016]. (Right) Schematic of I-V curves showing the two RRAM modes of operation: (c) Unipolar and d) Bipolar adapt from [Grossi 2016].

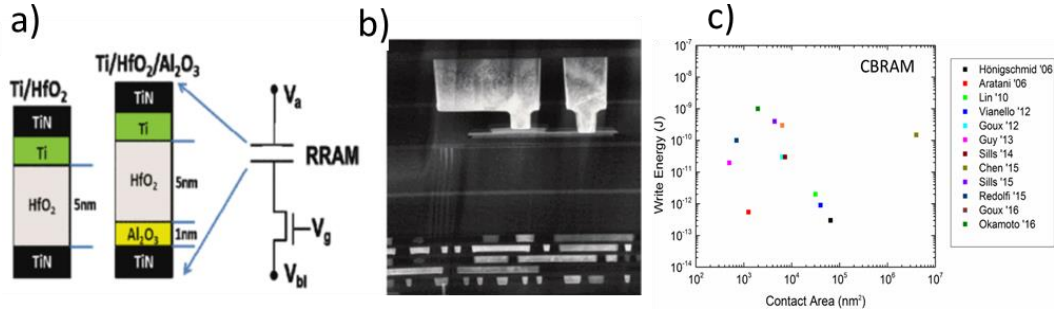


Fig. 21: (a) Schematics of the RRAM cells integrated into a 1t1R structure (b) TEM cross section of the cell from [Azzaz 2015] integrated in 28nm CMOS technology. (c) CBRAM Write energy vs Contact area Adapted from [Wong 2018]

This kind of memory needs a selector device to avoid sneak current problems and to give current compliance. In fact, a permanent dielectric breakdown during the Set process could occur if any current compliant is imposed. A big advantage of Re-RAM technologies is that they have the potential to achieve a $4F^2$ cell, and they can be stacked in 3D architecture to enhance memory density that positions them very well for scaling. Switching from one state to the other requires 10 ns, and energy roughly on the order of 0.1–5pJ can be assumed for programming. Retention and endurance can vary over a wide range depending on the material choice for the stack. OxRAM endurance is $>10^6$ with value of high resistance $10^5 \Omega$ and low resistance of $10^3 \Omega$ [Azzaz 2015]. For the RRAM a trade-off between write energy and retention is necessary, because higher writing energy gives filament robustness. Hence typical writing voltage is around 2V and writing current between 100 μ A–200 μ A for 100ns, obtaining a typical writing energy of 20pJ–40pJ. To read the memory the current is around 1–10 μ A, $V_G \sim <2V$ [Jurczak, 2015], obtaining a typical reading energy of 0.1pJ–1pJ. See Fig. 22 for reference.

		A. CBRAM	B. OxRAM	C. OxRAM - Interface
Storage Mechanism		Electrochemical filament formation	Valence change filament formation	Change in interface
Feature size F	Best projected	<5nm [A1]	<5nm	<10nm [C1, C2]
	Demonstrated	20 nm (GeSe) [A2]	3 nm [B1]	30nm [C3]
Cell Area (2D)	Best projected	4F ²	4F ²	4F ²
	Demonstrated	4F ² [A3]	4F ²	4F ²
Read Current	Best projected	not available	not available	n/a
	Demonstrated	100 pA [A4]	~ 100 pA [B2]	0.1μA [C3]
W/E time	Best projected	<1ns	<1 ns	<10ns [C1, C2]
	Demonstrated	< 1ns [A5]	100ps [B3] 85ps (AlN) [B4]	10ns [C1, C2]
Retention Time	Best projected	>10yr	> 10yr	3yr @ 55°C [C2]
	Demonstrated	>10 yr (1000hr 200°C) [A6]	> 10 yr (50 days @ 250°C) [B5] 105 @ 240°C [B6]	180 days @ 110°C [C3]
Write Cycles	Best projected	> 10 ¹¹ [A7]	>10 ¹²	>10 ¹² cy [C5]
	Demonstrated	10 ¹⁰ [A7]	10 ¹² [B7], [B8]	10 ¹² cy [C5]
Write operating voltage (V)	Best projected	<0.5V	<1V	n/a
	Demonstrated	0.4V [A8]	110mV [B9]	<3V (pulsed) [C1]
Read operating voltage (V)	Best projected	<0.1V	0.1 V	n/a
	Demonstrated	0.05V [A9]	0.1	0.1V [C3]
Write energy (J/bit)	Best projected	not available	0.1 fJ [B10]	n/a
	Demonstrated	0.1-10pJ [A10]	3.8fJ SET, 20fJ RESET [B11]	<1pJ [C1, C2, C4]

Fig. 22: [IRDS 2017] Reports RRAM Demonstrated and projected parameters.

1.5.2.2 NEMS

With Nanoscale electromechanical switches (NEMS) it is possible to controls the presence (or absence) of current by making (or breaking) a physical contact between two conductive electrodes. The two electrodes are normally separated by an air gap, allowing achieving virtually “zero” leakage current and abrupt switching. Fig. 23 shows typical I-V curves, where V_{PI} is the voltage to be applied to pull the movable electrode into the contact electrode and V_{PO} is the voltage at which the contact is broken. The NEMS typically have a hysteresis that depends on the design, contact material, contact surface and cycling. Because of their steep slope and zero leakage current there is a high level of interest on using NEMS for ultra-low power applications [LIU 2017A; OSOBA 2018; XU 2014A; BEIU 2014].

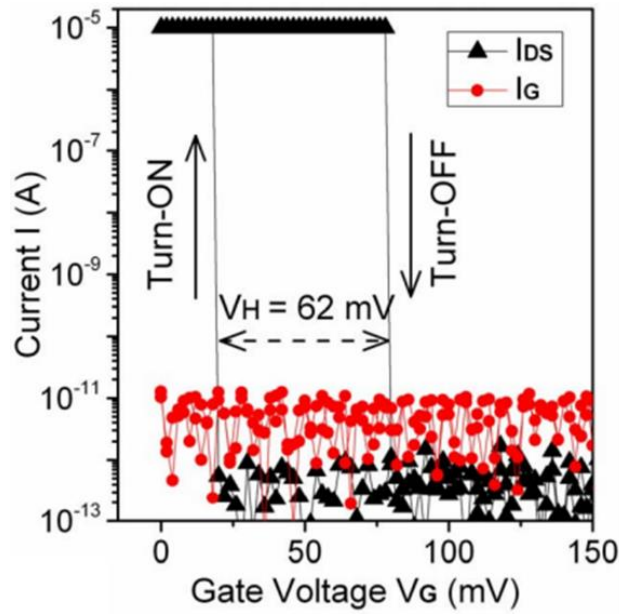


Fig. 23: Typical I-V transfer characteristic of a NEM switch: source [Qian 2017a]

Various actuation mechanism can be employed for mechanical actuation such as electromagnetic, electrothermal, piezoelectric and electrostatic. For the scope of this thesis only the electrostatic mechanism will be detailed but here a brief explanation of the various mechanism will be given.

1.5.2.2.1 ELECTROMAGNETIC

An electromagnetic switch uses magnetic force to actuate a movable electrode to make a contact, see example in Fig. 24. The movable electrode in response to the magnetic field moves, and once the magnetic field is removed (or reversed the magnetic polarity), the movable element goes back to its original position (thanks to the restoring spring force). Most of the electromagnetic switches employ external permanent magnet [Wu, 2010], which is a disadvantage because of its incompatibility with CMOS ICs. Otherwise some electromagnetic relays employs micro-coils to create electromagnetic field to move the element [GLICKMAN 2011]. However, this kind of relay is not suitable for low-power applications because the micro-coils requires high static power dissipation [DELEONIBUS 2018].

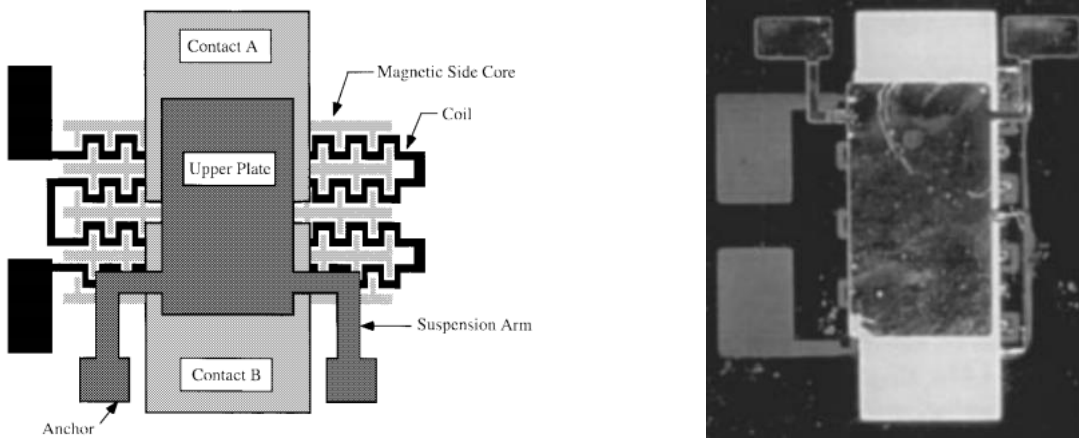


Fig. 24: Schematic view (left) and photograph (right) of an electromagnetic cantilever microrelay [Taylor 1998]

1.5.2.2.2 ELECTROTHERMAL

An electrothermal relay actuates through joule heating effect caused by passing current through a conductor and inducing thermal expansion and contraction of a movable element.

The thermal actuator is generally made by a single material [UDESHI 2004], with a specific geometry shape that allows elongation [Li 2010B], rotation [Dellaert, 2015] or bending through a preferential direction.

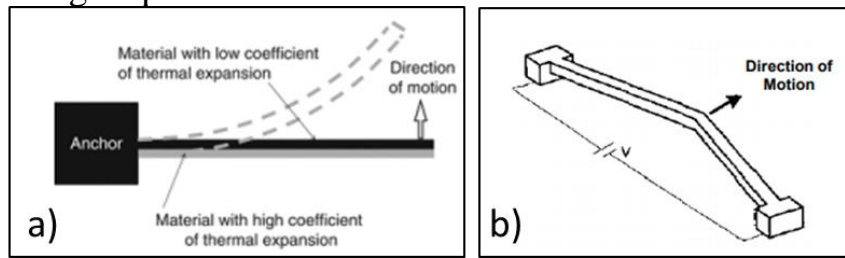


Fig. 25: Examples of thermal actuator designs, including bimorph (a) [Riethmuller, 1988] (b) geometry-based (V-shaped) [Girbau 2003]

However, such a switch can also be made by stacked layers of different materials, having different expansion coefficients (bimorph). The different thermal expansion between the two layers due to temperature increment results in a bending moment creating vertical displacement. This kind of relay can provide a very large contact force with low voltage; however, the power consumption and the actuation speed are higher than with respect the other types of relay. Examples of electrothermal actuator are represented in Fig. 26 and Fig. 27.

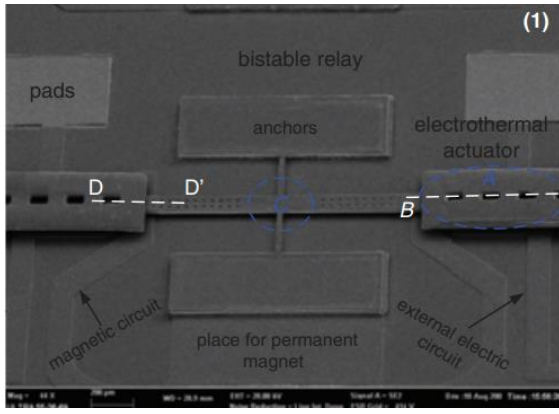


Fig. 26: SEM picture of magnetostatic and electrothermal MEMS switch [Wu 2010]

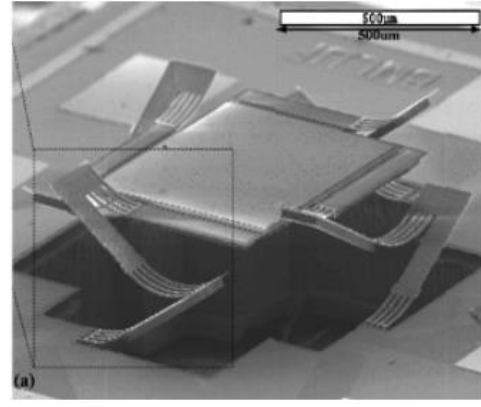


Fig. 27: SEM pictures of an electrothermal vertical displacement of a micro-mirror [Wu 2008]

1.5.2.2.3 PIEZOELECTRIC

Piezoelectric relays exploit piezoelectricity to achieve amplified deflections to make/break contact. Piezoelectricity is the linear coupling between electric polarization (P_{ij}) and strain (ϵ_{ij}) and stress (σ_{ij}). In Fig. 28 is shown an example of piezoelectric relay.

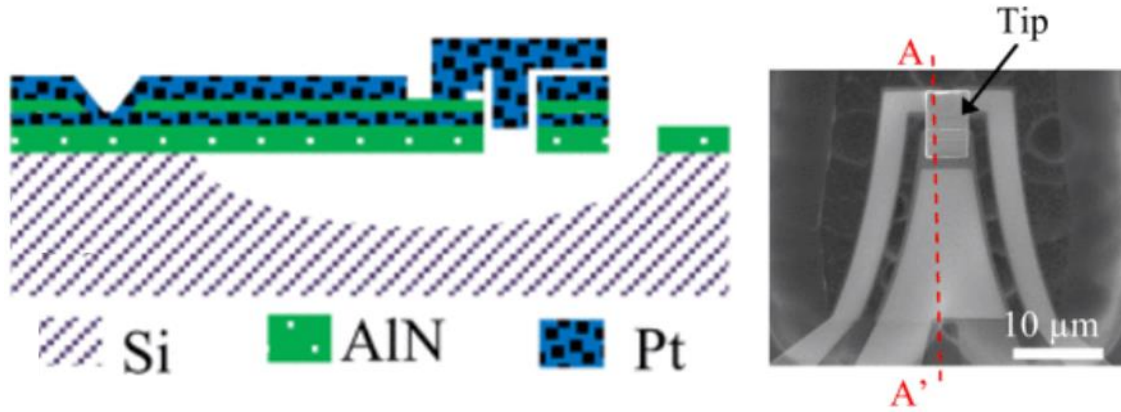


Fig. 28: Schematic and SEM images of the triple beam NEM piezoelectric switch [Zaghloul, 2013]

The mechanical deformation (induced stress or strain) is obtained while applying a certain electric field. The deflection direction is usually perpendicular to the electric field. The beam deflection amplitude is linearly controlled by the magnitude of the applied electric field across the piezoelectric layers, allowing very low actuation voltages [CHEN 2015]. ~10 mV gate voltage swing, with body bias voltage applied are reached in [Zaghloul, 2013]. The hysteretic switching behavior is only due to contact surface adhesive force, such that the hysteresis voltage can be very small (~8mV). Typical material used for MEM switch applications include ZnO, AlN and lead zirconate titanate (PZT) [Trolier-McKinstry, 2004]. Compared

to electrothermal switches this relay technology offers better switching speed ~ 1 μ s, and lower active power consumption [CHEN 2015]. However, to make them compelling to IC applications, piezoelectric relay should overcome the following challenges. a) Scaling: Lateral dimension scaling to increase device density at the scope of augmentation of functionality per area. Piezoelectric film thickness scaling to achieve low actuation voltage while maintaining low leakage current through the piezoelectric film. b) Maintaining stress uniformity: it is crucial to maintain the same stress profile across many devices and same stress gradient across the multilayer thin film stack of piezoelectric actuator to achieve a uniform actuation voltage [DELEONIBUS 2018].

1.5.2.2.4 ELECTROSTATIC

An electrostatic relay actuates thanks to capacitive effects. When a voltage difference between a fixed (gate) and a movable electrode (source) is applied, the charges of opposing polarity create an electrostatic force. By charging and discharging the capacitor, the movable electrode is attracted (by electrostatic force) or pushed away (by elastic force). If the potential value applied to the gate and the source corresponds to the pull in voltage V_{PI} then the source is able to pull into the contact electrode (drain). Source and drain are normally separated by an air gap, allowing zero off leakage current (zero static power consumption). The electrostatic force is not linearly dependent on the gap between gate and source, while the elastic force is. For this reason, when operating in pull-in mode (*i.e.* the gap between drain and source is 1/3 the gap between source and gate) electrostatic relay has a built-in hysteresis Fig. 29. In addition to this “built-in” hysteresis in pull-in mode, adhesion forces appear upon contact, counteracting the spring restoring force. This hysteresis is a drawback for logic switches, while it is an advantage for nonvolatile switches, allowing increasing CMOS functionality while adding a memory element to it. Electrostatic switches are the most used for mechanical switches because they can be fabricated above CMOS circuitry. Electrostatic nano-relays can achieve relatively fast switching speed (~ 100 ns). Moreover, NEMS shows promises for future low-cost, ultra-low-power electronics applications.

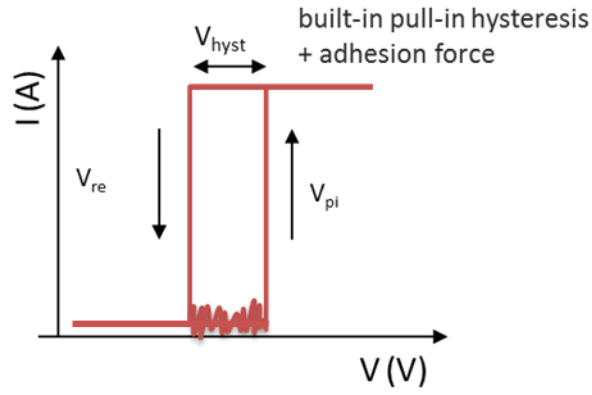


Fig. 29: Hysteretic behavior showing the pull in voltage and the release voltage that is always smaller than V_{pi} .

However, some challenges still need to be approached, such as the variability and reliability of the contact surface and consequentially the reliability and magnitude of contact resistance. For digital logic applications, the ON-state resistance (R_{ON}) of a relay can be as high as $\sim 10\text{k}\Omega$ without significantly affecting circuit performance [LIU 2012]. For NON-Volatile application, the contact resistance needs to be lower than the resistance given by the CMOS technology (see Chapter 3).

1.6 THESIS GOAL

This dissertation aims to address the promises and challenges associated with Non-Volatile NEMS relay electrostatically actuated. It presents the fabrication of monolithically co-integrated NEMS above CMOS technology with the goal of achieving ultra-low power consumptions. The originality of this thesis work resides in exploring the NEMS as nonvolatile element to increase CMOS functionality together with energy efficiency.

The details of the chapters are as follows:

Chapter 2 gives overview of the NEM state-of-art, it introduces mechanical and electrical behavior of NEMS relays. It gives an extensive review of the suitable materials used to fabricate NEM relay device. It investigates about NEM relay challenges such as high and/or variable contact resistance, mechanical delay, and scaling limits. It gives general information about the issues that are still blocking the development of NEM relay technology.

Chapter 3 covers design and modeling for non-volatile NEM relay. In particular, it deals with design studies such as: relay configurations, actuation type, and dimensioning of the relay parameters. The presented study aims at optimizing NEM relays in terms of reliability and performance and it is supported by simulations and modeling. Design strategies are presented in order to obtain a larger working margin that is robust to pull in voltage shift and variation due to cycling and/or pull in voltage variability due to fabrication. The chapter continues with a study that gives design guidelines for Non-Volatile NEM relay. The particularity of this study is that it considers the incertitude of adhesion force value and the possibility of catastrophic pull in (beam contacting the gate), resulting in a more robust approach. In the chapter two parentheses are opened; One to investigate volatile NEM relay scaling in terms of contact stability and contact resistance. The other parenthesis investigates the actual advantages of having air gap in the CMOS interconnections, to simplify the fabrication and release process of NV NEM relay. The chapter end with simulation studies of NV NEM relay implemented with SRAM using MEMS+ plugin for Cadence.

Chapter 4 discusses the fabrication of hybrid NEM/CMOS technology. It explores the feasibility and the interest of the co-integration between CMOS technology and NEM relays using monolithically 3D integration. In particular it presents the process flow to fabricate in-plane actuated relay defined on a mono-crystalline Silicon layer. This fabrication method is obtained by doping and annealing the mono-crystalline Si layer of an SOI wafer, followed by wafer bonding between a commercial CMOS technology wafer with functional circuits and the SOI wafer. The chapter ends with the presentation of electrical results of the fabricated technology.

Chapter 5 summarizes the key results and gives perspective for future research.

In the annexes two different fabrication approaches for monolithically integration of NEM relay with CMOS are explored: i) NEM relays defined on a polycrystalline Silicon layer above a commercial CMOS technology wafer. ii) Vertical defined NEM relays made by multi-levels of metal interconnections of a commercial CMOS technology wafer (ST 130nm). The first approach is obtained by depositing amorphous silicon p-doped in situ above a CMOS wafer and followed by laser annealing re-crystallization and dopant activation.

2 NEMS RELAY: STATE OF ART

2.1 INTRODUCTION

Technological development of Silicon and in general CMOS device fabrication technology has affected positively the development of micro/nano-electro-mechanical system. That is why, Nano-electro-mechanical (NEM) relays have been considered as good candidate for ultra-low power consumption microelectronics applications. From now on, on this manuscript the term NEM relay/switch will refer only to micro/nanoelectromechanical switches that are turned OFF and ON by making and breaking a contact through electrostatic actuations.

Many studies regarding the use of NEMS as logic switches to replace or help MOSFET have been published *e.g.* [KAM 2015]. At the same time, owing to the NEMS relay intrinsic hysteresis NEM switches have also been considered as good candidates for Non-Volatile applications and memory in computing [Xu 2014; Kim 2015; Usai 2017; Qin 2018].

The main opportunities offered by NEM relays are:

- 1) Zero off-state leakage.
- 2) Abrupt switching (ideally zero subthreshold swing).
- 3) Infinite ON/OFF current ratio.
- 4) Ability to operate in harsh environment.
- 5) Non-Volatility depending on the design.

Despite the fact that simulations show low actuation voltages, only few experimental results effectively demonstrated sub 1-V actuation voltages [Zaghloul 2014; Osoba 2016; Qian 2017] exploiting the body-bias technique and adding an anti-adhesion coating. However, the body-bias technique requires the presence of an additional terminal, which increases the overall footprint.

NEMS relay contact stability remains a reliability concern. In fact, over many switching cycles, a progressive alteration of the morphological and chemical properties of the contacting surfaces may cause a drift in the contact resistance and in the adhesion forces [MUHLSTEIN 2001].

The main challenges for NEMS relays that are shared by both logic and memory applications are:

- 1) Achieving low Pull-In voltages to assure low switching power (active power).
- 2) Improving endurance (cycling) with stable OFF/ON resistance over the life of the switch (*e.g.* $\sim 1 \times 10^{14}$ switching cycles for logic application, $\sim 10 \times 10^3$ for Code macro and $\sim 10^6$ for data macro and Nonvolatile applications in automotive).

- 3) Fabricating NEMS with CMOS-compatible process, especially in case of hybrid CMOS/NEMS applications.
- 4) Scaling the footprint and the actuation speed (below 10ns).

This chapter deals with NEMS relay State of the Art, describing NEMS relay typical design and orientation type and reporting suitable materials for all the NEM switching parts such as: structural layer, sacrificial layer, contact layer.

Contacts propriety will be discussed in 2.2.5, explaining the NEM switching contact domain and contact resistance. Relay endurance, switching speed and scaling will also be approached.

Explanation about the main issues that interfere with reaching high endurance will be given in 2.2.6, additionally with the relay endurance state of the art. Section 2.2.7 and section 2.2.8 will deal with relay mechanical delay and scaling limits.

Section 2.2.9 deals with why Logic NEMS relays are today a good candidate for computing in a harsh environment. In fact, NEMS relays have been successfully tested at up to 500 °C [LEE 2010C]; therefore NEMS relays could be suitable for many applications: *e.g.* automotive, aerospace, or geothermal exploration. In fact, those applications need a microcontroller that could operate at very high temperatures. MOSFET transistors operating at high temperature are subject to electrical failure due to material electrical properties degradation.

Finally, we will list the applications that so far have been targeted for NEMS relays, highlighting the advantages that NEMS relay will give to those applications.

To study NEMS relays it is necessary to include considerations from various disciplines such as mechanical engineering, material science and integration engineering.

2.2 NANOELECTROMECHANICAL SWITCHES

2.2.1 DESIGN

In general, A NEM switch has at least three terminals: The source (input), the drain (output), the gate (actuation terminal) and sometimes the body. The body allows for having a reference potential for actuation besides the source. The source and the drain are separated by an air gap. Some relay designs have more than one electrode intended for the actuation, the source or the drain. Terminology such as “drain”, “source”, “gate” and “body” is used to make analogies to their MOSFET counterpart [NATHANAEL 2012] see Fig. 30 for reference.

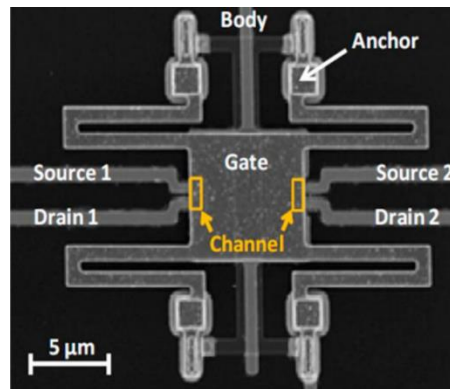


Fig. 30: SEM plan view of a 6T relay (1 Gate, 1 body and 2 sources and 2 drains) of a logic relay [Liu 2012]

The relay design is strongly influenced by the target application. Therefore, its design can vary in terms of number of terminals, structure orientation, actuation direction and fabrication process. Additionally, each research group has its own expertise in relay design. For example, logic relays with cantilever beam designs can be found in [CHONG 2009; JANG 2008] and clamped-clamped beam designs can be found in [CZAPLEWSKI 2009; KAM 2009].

Nanoelectromechanical relay can have a lithographically-defined gap with an actuation area partially defined by deposition (see *e.g.* in Fig. 31): [GROGG 2014; MUÑOZ-GAMARRA 2014; PARSA 2011], or a gap defined by deposition with a lithographically-defined actuation area *e.g.* in Fig. 31: [PAN 2015; RAMEZANI 2014]. Multiple NEM relays structure orientation exists such as lateral, vertical or seesaw (*e.g.* in see Fig. 32).

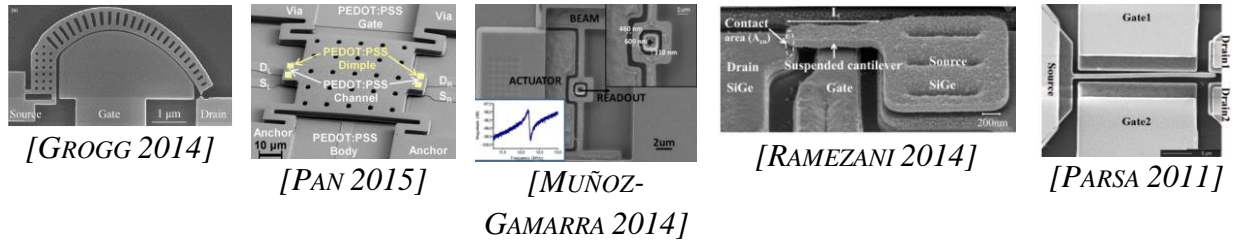


Fig. 31: Example of various NEM relay design type and structural orientation

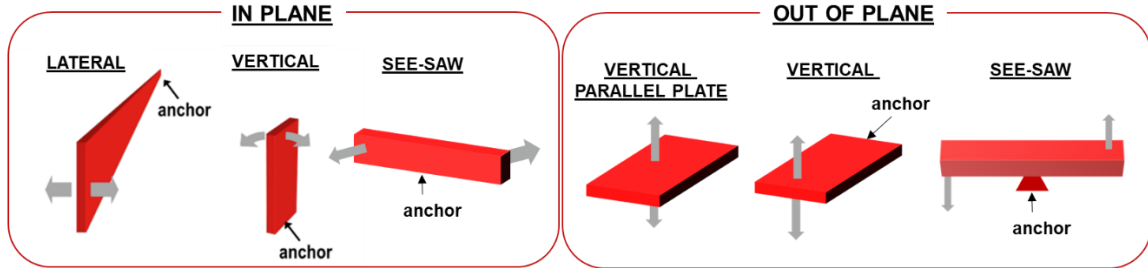


Fig. 32: Options for NEM relay structural orientation; Figure adapted from [Liu 2017a]

Commonly, NEMS relays are vertically (out of plane) or laterally (in plane) actuated. The vertical actuation is when the relay moves or tilts out of the plane of the wafer surface see literature examples Table 1, and lateral actuation is when it moves or tilts in the plane of the wafer surface see literature examples Table 2. In the tables below w_b is used to indicate the beam width and L_b the beam length, F^2 is used to indicate the device footprint.

Table 1: Vertical Actuated Relays State of Art:

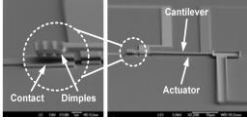
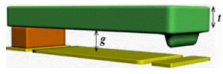
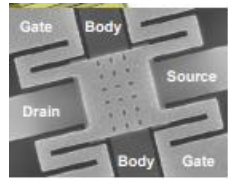
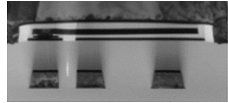
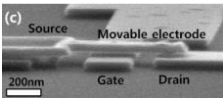
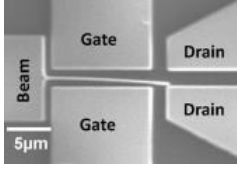
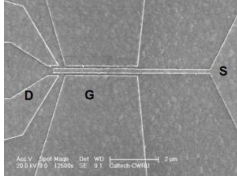
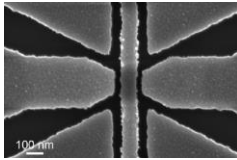
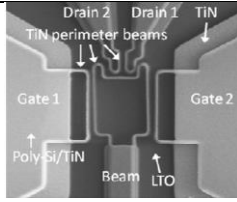
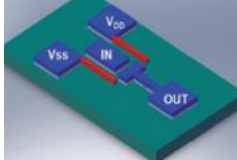
<i>REF</i>	<i>Figure</i>	<i>Terminals</i>	<i>Nominal F^2</i>	<i>V_{DD}</i>	<i>Applications</i>
[Do 2013]		3	$F^2=9*10^{-12}m^2$ $L_b=18\mu m$ Contact dimple width=500nm	9V	Memory (with vibration reset)
[Kim 2011]		3	-	66V	Logic (suppressed bouncing)
[NATHANAEL 2012]		4	$F^2=2.622*10^{-9}m^2$ From anchor to anchor vertical =40 μm From anchor to anchor lateral=70 μm	8V to 2V (Tuned by body bias)	Logic
[JOSHI 2010]		4	$L_b=5\mu m$	4V	Logic
[LEE 2009A]		3	$F^2=4*10^{-13}m^2$ $L_b: 700\text{ nm};$ $w_b*3:200\text{ nm}$	8V	Logic (liquid packaging)

Table 2: Lateral Actuated Relays State of Art:

REF	Figure	Terminals	Nominal F^2	Pull-In	Applications
[PARSA 2013]		5	$F^2=6*10^{-11}\text{m}^2$ $L_b=16\mu\text{m}$ $w_b*6+\text{gap}=600\text{nm}$	7V	Logic (Platinum coating to improve Resistance)
[LEE 2009B]		5	$F^2=2*10^{-12}\text{m}^2$ $L_b=5\mu\text{m}$ $w_b*6+\text{gap}=4\mu\text{m}$	-	Logic (high temperature)
[CZAPLEWSKI 2009]		5	$F^2=1.6*10^{-12}\text{m}^2$ $L_b=4\mu\text{m}$ $w_b*6+\text{gap}=400\text{nm}$	6V	Logic (co-integration with CMOS)
[LEE 2010A]		5	$F^2=1*10^{-10}\text{m}^2$	34V	Logic (Titanium as structural layer)
[LEE 2010C]		3	-	6V	Logic (high temperature)

2.2.2 NEM RELAY ACTUATION MECHANISM

In this paragraph, the NEM relay actuation mechanism is explained, taking as an example a three-terminal relay defined by a movable part, for instance a cantilever beam (source-input) shaped electrode, and two fixed electrodes: the gate (control) and the drain (output). When a bias V_{DD} is applied between gate and source, an electrostatic force F_{ele} is produced, proportional to the actuation area A and following the inverse square of the separation distance between gate and source [USAI 2017A]. F_{ele} is competing with the spring restoring force F_k , which itself has a linear dependence on the beam displacement. As a consequence, beyond a critical displacement of the beam, the electrostatic force always overwhelms the spring restoring force, causing the beam to “pull-in” and contact the drain electrode, creating an electrical connection see Fig. 33. The bias corresponding to the critical displacement is called Pull-In Voltage (or V_{PI}). If the adhesion force F_{adh} between the two contacting surfaces (source-drain) is larger than F_k , when V_{DD} is removed, the movable electrode retains its position (*i.e.*, it is non-volatile).

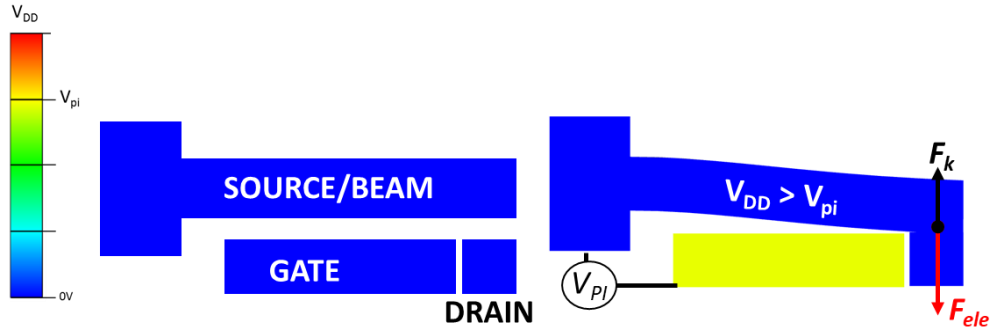


Fig. 33: Schematic top view of electromechanical cantilever beam relay, showing the main design variables, and introducing the forces. Left side, shows relay neutral position ($V_{DD}=0$). Right side (Pull-In position $V_{DD} \geq V_{PI}$)

If F_k is larger than F_{adh} , the source detaches from the drain at a voltage corresponding to the release voltage V_{RE} . V_{RE} is always lower than V_{PI} because of the nonlinear nature of F_{ele} , and because of presence of F_{adh} between the two contacting electrodes, causing a hysteresis effect. It is possible to define the hysteresis voltage as $V_{HY}=V_{PI}-V_{RE}$. This property is very useful in applications that need non-volatility and it is an issue for logic switches.

In the literature, there are papers which focus on reducing the contact adhesion force of NEM switches for logic application *e.g.* [OSOBA 2016], and there are papers which focus on having the right amount of adhesion margin to allow non-volatility and re-actuation *e.g.* [USAI 2017B]. Table 3 shows the relay needs in terms of voltage, size, hysteresis and RC delay in case of logic and memory applications.

Table 3

	Voltage	Size	Hysteresis	RC delay
Logic	It needs to be as low as possible, sub 50mV demonstrated thanks to body bias and contact coating [OSOBA 2016]. Today CMOS technology $V_{DD} \leq 0.9V$	It needs to compete with MOSFET size	It needs to be as low as possible tending to 0	The mechanical delay is predominant
Memory	For embedded memory the voltage is given by CMOS circuitry. For standalone memory it needs to be low as possible	Comparable to Standard Cells (e.g. SRAM size) or Comparable to other 1R-1T Non-Volatile Memory cells	The hysteresis margin has to be semi-infinite to assure non volatility if power OFF, without permanent stiction	The relay contact resistance has to be as low as possible to reduce the writing delay and for some application to assure the re-writing operation (see Chapter 3)

2.2.3 MATERIALS

To identify suitable materials for NEMS relays it is ideal to divide the relay by its constituent parts: structural layer, sacrificial layer, coating layer and contact layer (which might be the same or different from the structural layer) see Fig. 34. Each of these layers has proper requirements in terms of physical, chemical and electrical proprieties.

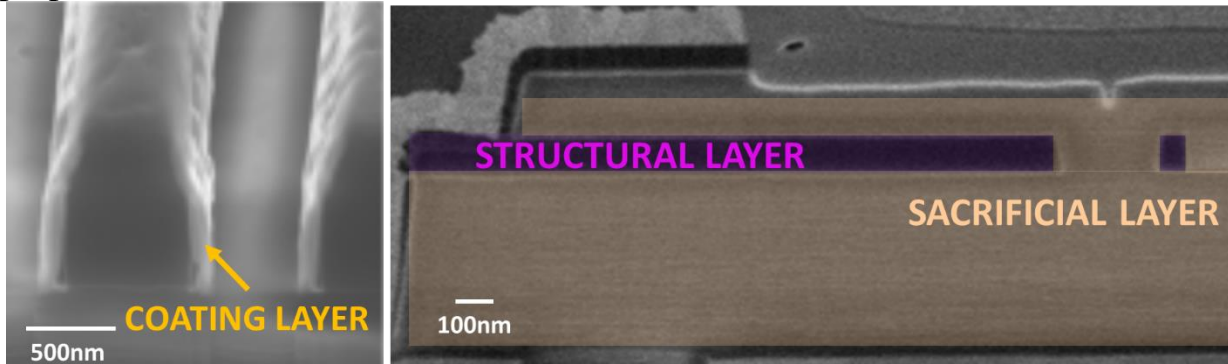


Fig. 34: Left) Cross-sectional SEM picture of Lateral actuated NEMS, highlighting sidewall Coating Layer, figure adapted from [Parsa 2013]. Right) Cross-sectional SEM picture of Lateral actuated NEMS, highlighting structural layer and sacrificial layer.

2.2.3.1 SACRIFICIAL LAYER

The sacrificial layer is used for the fabrication process of the NEM structure. The purpose of this layer is to create air gaps between the electrodes, and it might be needed to give support to the structural layer. The air gaps are needed to create electrical isolation between the electrodes and to allow for the movement of the movable electrode. There are several methods for depositing this layer, such as physical vapor deposition (PVD); chemical vapor deposition (CVD); epitaxial growth, depending on the size and the need of conformity and on the release process method. After processing the NEM relay this layer is going to be etched away. The chemistry used to etch this layer has to be highly selective in order to avoid the structural layer etch. Therefore, the choice of the sacrificial material type is highly influenced by the structural material and vice-versa. The sacrificial layer needs to be uniformly and conformally deposited, because it needs to fill the small gaps ($<100\text{nm}$) between the electrodes. The conformality helps also with the release process, allowing for a more controlled etch. If there is a need of CMOS co-integration, the sacrificial layer should be deposited at low temperatures. In case of in-plane actuated relay, the sacrificial layer needs to fill the space between the electrodes. Therefore, the deposition needs to be highly conformal to avoid leaving voids that will create inhomogeneity during the release process. In case of out-of-plane actuated relay, the thickness of the sacrificial layer determines the gap between the gate and the source. Meaning that the actuation voltage is influenced by the sacrificial layer thickness. Therefore, the deposition rate needs to have high

controllability and uniformity. As the relays scale down, the importance of the uniformity of the deposited thickness increases making critical being able to control the deposited thickness. The most used sacrificial layer in MEMS and NEMS is the SiO_2 (with Si as structural material). SiO_2 can be selectively removed in wet or vapor HF. It can be deposited via low-pressure chemical vapor deposition with good uniformity and conformity and low temperature [NATHANAEL 2012]. Another sacrificial/structural couple of used materials in MEMS is Si-SiGe. [CAUBET 2006; DURAND 2008].

2.2.3.2 STRUCTURAL LAYER

The structural layer is the layer in which the NEM structure is defined. It influences the actuation voltage value through its Young Modulus (E), and the structural layer mechanical stability influences the reliability and endurance of the NEMS. Especially for contacting relays, it is important for its structural material to be robust against fracture and fatigue. And in general, to be resistant to the chemistry that etches the sacrificial layer. Furthermore, it has to be deposited/grown uniformly and conformably and it has to have low residual stress and strain gradient, otherwise after the release the structural layer could buckle. Buckling of the movable electrode could cause an actuation/contact gap enlargement resulting in the need of a higher actuation voltage or could cause collapse of the movable electrode into the substrate or into another electrode resulting in relay fabrication failure. Finally, if the structural layer is not conductive it should be combined with a coating conductive material.

Silicon and its compounds are the most widely used materials for the MEMS structural layer. The main reason is that Silicon is dimensionally stable, and the fabrication techniques are well established and well documented. Additionally, monocrystalline Silicon has a good mechanical stability; it can be integrated with conventional electronics. It has a relatively high melting point (1400°C) and a low thermal expansion coefficient and it has low residual stress $\sim 1.4\text{MPa}$ [VRINCEANU 2002]. In the case of Silicon-On-Insulator SOI wafers, widely used in the MEMS technologies, the residual stress cannot be higher than 12MPa in order to avoid adverse effect on the bonding interface [CHER MING TAN 2003]. If highly doped it can be used as a conducting material. The problem in using monocrystalline Silicon NEMs relays that are monolithically integrated above CMOS circuitry is that the entire fabrication process flow requires many steps, including wafer bonding. A solution of that problem is to co-integrate monocrystalline NEM and CMOS using co-planar fabrication process but at the expenses of the occupied area (see Chapter 4). Another low point for contacting NEM relay is that Si tends to oxidize very fast, which is a problem for the contact resistance point of view. Other Silicon compounds such as Polysilicon or recrystallized amorphous silicon can be deposited at low temperature, making the co-integration process easier but the fast

oxidation remains a problem. However, poly-Si residual stress is higher than the mono-crystalline Si. Many other materials have been considered as structural layer and their main advantage and disadvantage are listed in Table 4.

Table 4: Frequently used materials for NEMS relay structural layer:

Material	Advantages	Drawbacks	Research Group
Mo	Ideal contact properties: hardness, low wear-out over cycles, no native oxidation, and excellent thermal stability	Large Young's modulus =329GPa (intrinsically high Pull-In voltage)	VIRTUS, Singapore [Do 2013] A*STAR, University
Si	Highly used in MEMS, CMOS compatible, Low residual stress	Oxidizes fast, high wear-out	IBM [GROGG 2013]; Stanford [AKARVARDAR 2007A]; CEA LETI (this research work)
Poly Si	Low residual stress, negligible strain gradient can be achieved after release or with annealing.	Oxidizes fast, high wear-out. The reduction of the stress gradient after releasing could increase the actuation gap increases the actuation voltage.	Texas Instrument [PARSA 2013]; UC Berkeley [KAM 2009A]
Poly SiGe	Highly used in MEMS, CMOS post-processing compatible, easy to release with dry etching if sacrificial layer is Si.	Oxidizes fast, high wear-out	[RAMEZANI 2015] IMEC UC Berkeley [Osoba 2018]
SiC	Low thermal expansion, High thermal conductivity, High hardness 9-9.5, Excellent thermal shock resistance	Oxidizes fast, structure with many dislocation defects.	CWR University [HE 2013]
CNT	Possibility to achieve small footprint and reduced area	Difficult to integrate	[LEE 2004]; EPFL [CAO 2012]
W	Hard refractory metal, good mechanical resistance to wear and tear (hardness	Oxidizes rapidly	University of Barcelona [RIVEROLA 2016]

	7.5-8), high melting point		
Cu	Highly conductive, CMOS compatible (BEOL relay)	Relatively high adhesion force, oxidizes fast, low hardness	Cea Leti / UC Berkeley; Barcelona University <i>[MUÑOZ-GAMARRA 2016]</i>
TiN	High melting point, extreme hardness and excellent chemical and thermodynamic stability	Likely to be etched in vapor HF	KAIST <i>[LEE 2009A]</i>

2.2.3.3 CONTACT AND COATING LAYER

The contact electrode material can be the same material as the structural layer [Ramezani 2014], or it can be a coating layer added to the structural layer [Parsa 2011; Shavezipur 2013], or it can be a completely different material from the rest of the relay [Nathanael 2012]. Either way, the main goal of this layer is to control the current flow, and to assure that the same amount of current will be able to flow for many ON/OFF switching cycles. Hence, it has to have low contact resistance, without too much drift or shift of the ON initial resistance value. Linked to what written above is that the contacting material has to be resistant to wear and to plastic deformation (*i.e.* high hardness). It has to be resistant to the sacrificial layer etching chemistry. It has to have a conformal and uniform deposition process, and it has to be possible to process it at low temperature for CMOS co-integration. Detailed information about contacting material (Ir, Pt, W, Ni, Cr, Ti, Cu and Al) for relays can be found in [Chowdhury 2013].

2.2.3.3.1 COATING LAYER

The coating material must be resistant to corrosion to avoid oxide formation and it should be a hard metal to increase the device endurance with an adequate low resistance. It should be resistant to the chemicals used to release the relay. For example, tungsten W is a hard metal, meaning that it can increase the device endurance; however, it is etched away in XeF_2 chemistry (silicon etchant). A good contact coating material can be platinum (Pt) because it can sustain high current conduction for long periods of time and it has low contact resistance [LIU 2010] and can endure many contact cycles due to its high hardness. Pt is also resistant to liquid and vapor hydrofluoric acid (HF) and XeF_2 , which make the process more flexible in terms of structural and sacrificial layer material choices. [PARSA 2013] shows that Pt sidewall coating reduces the on-resistance as low as $3\text{k}\Omega$. Another promising contact material is the Ruthenium Ru and its oxide RuO_2 which is only 7 times less conductive than Ru [PESCHOT 2015]. 10^{10} cycles without failure were demonstrated for a relay having RuO_2 as contacts material. However, Pt and Ru are difficult to integrate for a NEM above CMOS monolithic process. In addition, adding a coating layer could modify the stress gradient of the release structure causing buckling.

In some cases, especially for logic applications, it is possible that the coating material main goal is to reduce the adhesion surface energy [OSOBA 2016] even if it augments the contact electrode resistivity. Usually the materials used to decrease the surface adhesion decrease the contact electrical conductivity. Therefore, the choice of using whatever type of coating layer material depends also on the target applications. For instance, logic switches must have essentially zero hysteresis; this requirement often overcomes that of high electrical conductivity.

Table 5: Frequently used materials for NEMS relay contact and coating layer:

Material	Advantages	Disadvantages	Research Group
TiN	High melting point, extreme hardness and excellent chemical and thermodynamic stability	Difficult to integrate, likely to be etched in vapor HF	KAIST[Lee 2010]; IMEC [RAMEZANI 2015] [LEE 2010A]
Pt	Very high corrosion resistance, high melting point, stability in air and oxidizing environments, resistant to plastic deformation, hard	Hard to integrate	Texas Instrument [PARSA 2013]
a-C	Low adhesion, stable R_c value. Lower surface energy, CMOS compatible	High resistance	IBM with EPFL [GROGG 2014]
Ru	High melting temperature; stability in air and oxidizing environments; high. Hardness 6–7; stable in air and resistant to most corrosive agents. Good cycling demonstrated [CZAPLEWSKI 2009]. Resistance to oxidation	Difficulty in process integration, residues problem, adhesion problem (can be partially solved by inorganic sacrificial layer)	Sandia National Lab [CZAPLEWSKI 2009] UC Berkeley [CHEN 2014]
Au	Low-contact resistance	Low hardness 2.5–3; Low melting point, not CMOS compatible (highly diffusive)	University of Southampton [LEWIS 2015]
TiO ₂ /RuO ₂ /WO _x	Helps alleviating stiction	It limits the current	UC Berkeley [SAHA 2017A]
PFDTES	Reduce adhesion	It limits the current	UC Berkeley [OSOB 2016]
W	Low-contact resistance	Oxidizes rapidly	UC Berkeley [LIU 2012];

2.2.4 METAL INTERCONNECTION AIR GAP FOR NEM RELAY: ARE THEY GOING TO SIMPLIFY THE RELEASE PROCESS?

The advantage of fabricating vertically oriented BEOL NEMs relays (*e.g.* Fig. 35) is that for a given footprint, by using multiple metal layers, it is possible to form a larger relay actuation electrode, hence a larger actuation area, resulting in a lower actuation voltage [SIKDER 2017]. The presence of the interconnect air gap in the more advanced node is anticipated to be an advantage for the fabrication of those type of relays. In fact, the air gap could be used to pre-define the gap between the actuation electrode and the movable electrode and the gap between the contact electrode and the movable electrode, reducing the release step complexity. Additionally, if a common output electrode is used, the “channel” could be isolated by the movable electrode by the presence of the intermetal dielectric.

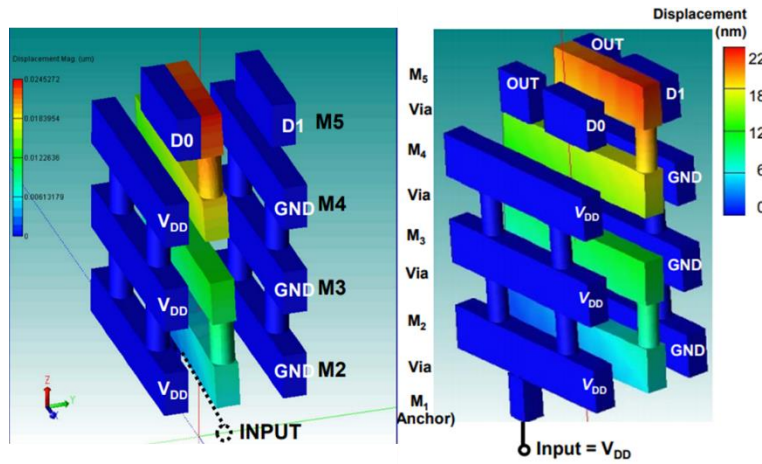


Fig. 35: Example of two BEOL NEMS structure using 5 interconnect layers: M1 as the anchor, M2-4 as the actuation layer, and M5 as the contact layer. The left figure shows a 5-terminal SPDT structure from [LIU 2017A]. The right figure shows a 6-Terminal SPDT structure with a common output electrode from [XU 2014A]. The color scale indicates physical displacement.

2.2.4.1 INTERCONNECTION STACK AIR GAP FABRICATION

In order to further reduce the capacitance between neighboring copper lines in interconnect structures, the introduction of cavities (*i.e.* air gaps) has been proposed [WANG 2004] and chips having multi-layer high-volume air gap interconnection stack have been fabricated [Natarajan 2014; Fischer 2015], see example in Fig. 36 Left.

There is more than one integration approach to obtain an air gap between conductive layer [Hoofman, 2006]. One approach starts after damascene Cu formation: Fig. 36 right (a). A cap layer can protect the metal, a lithography/etch step is introduced to define the cavities (as trenches) between conductive layer Fig.

36 right (b-c). The etching can be done by wet process or by plasma etching process. Then, a conformal protective layer is deposited followed by a non-conformal inter-metal dielectric deposition Fig. 36 right (d). Thanks to the non-conformal dielectric deposition, the air gap is formed automatically. The protective layer improves electrical performance and reliability of the interconnect stack and helps in controlling the final air gap shape [Gosset 2003a]. Finally, a chemical mechanical polishing process is used to planarize the dielectric interconnect stack.

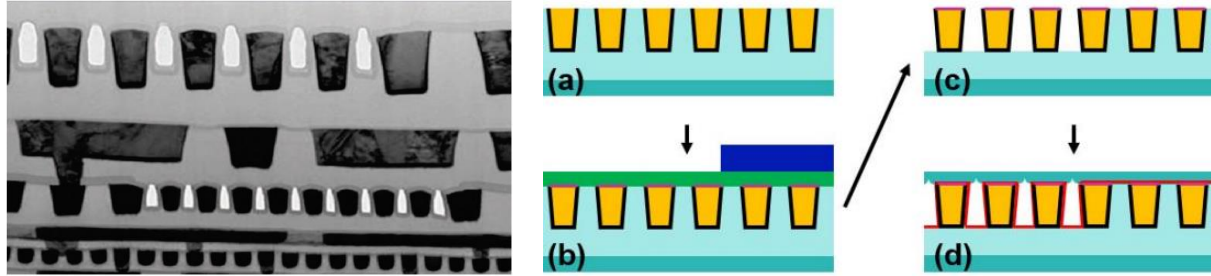


Fig. 36: Left) Multi-layer air gap using non-conformal CVD process, the air gap and the air gap protection layer are well visible. Figure from [Fischer 2015]. Right) Air gap integration flow with conformal dielectric liner and dielectric cap from [Penny 2017].

The air gaps need to be as narrow as possible to reduce the risk of via-misalignment, and at the same time to be as wide as possible to obtain the best low-k performance. Therefore, the cavities occupy only a portion of the intralayer region [PARK 2008]. Additionally, the closure of the air gap needs to be inside via level to prevent metal contamination insides the air gaps during the integration of the above metal layers [GOSSET 2003B]. An issue with this approach might be the additional cost of the lithography necessary to define the trenches to make the air gaps.

Another approach is to place between the copper lines where the air cavities are required, a sacrificial layer [Kohl, 2000]. Once the copper lines and the interconnect stack are completed, the sacrificial layer is removed. The etching can be done by wet etching, thermal or reactive plasma. The weakness of this approach is the low mechanical structural strength and often-low thermal budget (below 200°C) to avoid sacrificial layer etching during the interconnection stack processing. To assure mechanical stability it is better to avoid stacked connected air gap in successive metal levels. Moreover, above the sacrificial layer a hard mask that covers the air gap is added, allowing improving mechanical stability of the stack. This hard mask also acts as etch stop layer when opening misaligned vias (*i.e.* a protective cap).

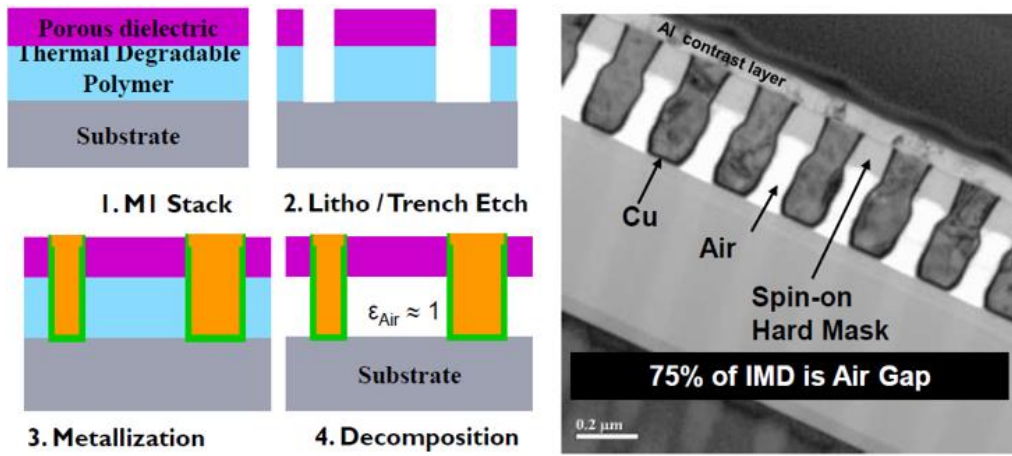


Fig. 37: Multi-layer air gap using sacrificial materials, the air gap and the air gap hard mask are well visible. Figured adapted from [Daamen 2015]

2.2.4.2 AIR GAP RELIABILITY ISSUES

A slight misalignment can happen while patterning the trenches to define vias. However, a mislabeled vias in an air gap cause metal contamination (see *e.g.* in Fig. 38), which increase leakage and intermetal capacitive coupling. Therefore, the presence of air gap directly adjacent to the via landing metal are to be avoided.

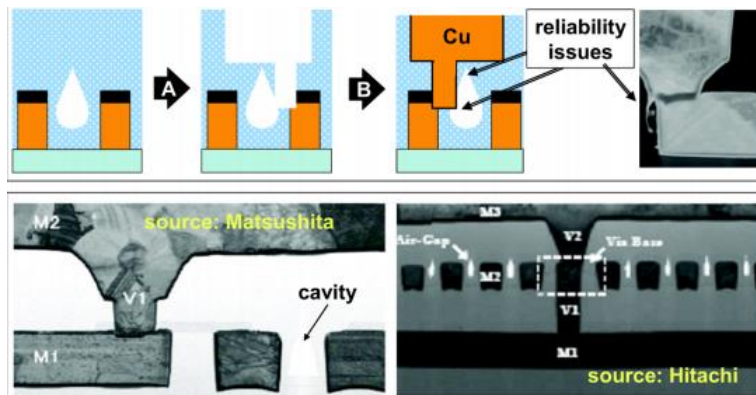


Fig. 38: Reliability issues caused by presence of air gap next to via: mislabeled via due to misalignment that can cause metal contamination on the air cavity, increasing leakage and capacitive coupling [Daamen 2007].

In case of air gap definition by non-conformal deposition approach, after dielectric trenches etching, a protective layer is deposited to protect the copper and reduce reliability issues caused by the non-conformal dielectric deposition. Similar the case of air gap defined by sacrificial layer approach. The air gap is defined below a hard mask layer, which protects the metal lines and increases mechanical stability of the entire stack.

2.2.4.3 CONCLUSION

If the commercial air gap process is used, no air gap can be defined between the actuation electrode and the movable electrode. In fact, compact vertically oriented relay structure needs to have vias connecting the stacked metal line used for the actuation electrodes and for the movable electrode. Additionally, in the commercial air gap process there is not air gap definition on via layers.

If the commercial air gap process is modified and air gap are defined between the actuation electrodes and the movable electrode, the relay release process should include: i) a dedicate release step to get rid of the protective layer and the remaining intermetal dielectric between void and metal ii) a dedicate release step to get rid of the hard mask.

In summary, the presence of the air gap in the interconnection stack does not imply an easiest fabrication process for the vertical oriented nanoelectromechanical relay.

2.2.5 NANO-CONTACT

MEMS and NEMS contacting surfaces seem to be very smooth but usually either the movable electrode (source), the contact fixed electrode (drain) surfaces presents both a roughness value of some nanometers (Fig. 39 Left). This implies that when the source pulls into the drain, only a few number of asperities makes contact, dramatically reducing the actual contact area from the nominal contact area Fig. 39 right (*i.e.* asperity regime contact).

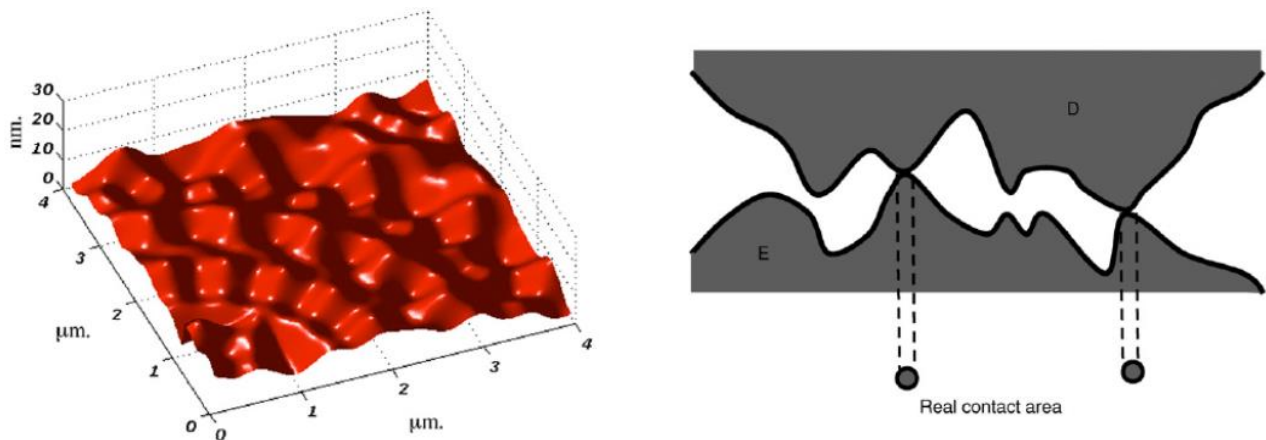


Fig. 39: Left) Example of surface topography having peak-to-valley of 12.8 nm and a root mean squared RMS of 7.2 nm. Right) Schematic of the surface asperities of the movable contact electrode (top) D and fixed electrode contact (bottom) E. Asperities-contacts are established when the two surfaces touch. Both figures from [Rezvanian 2007]

Typical dimension of asperities; range from tens to hundreds of nanometers. Nanoelectromechanical relay which have a nominal contact length and width of less than hundreds nanometer could have a single asperity electrical contact. The surface topographies of the contacting electrodes can be examined with an atomic

force microscope (AFM), which can give information about the asperities size with a resolution of $\sim\text{nm}$ [LUMBANTOBING 2004], and also of the asperities height, or density [BLACKMORE 1998]. However, to get a real insight on the contact surface, the number of the asperities present on the two surfaces, the position of the asperities and the deflection of the movable electrode need to be considered. Indeed, it might be possible that only a part of the asperities is contacting. Furthermore, to better get indication on the spot size of the contacting asperities the contact load at the contact interface needs to be taken into account. The surface contact is established by means of an electrostatic closing gap, driven by a fixed actuator electrode. To take into account the actual effect of the contact load at the surface interface it might be necessary an energy analysis to accurately determine the impact effect [LUMBANTOBING 2004] that could cause asperities deformation [YAN 1998]. In fact, this type of mechanical system exhibits instability at a critical distance beyond which, the movable electrode is abruptly pulled into contact [BAO 2000]. In the analysis of the contact surface, one also need to take into account the fact that Nanoelectromechanical relays have to be switched ON/OFF for many cycles. During cycling, the two contacting surfaces are likely to undergo degradation that causes a topography variation of the original contact surface and as well a variation of the chemical proprieties (*e.g.* oxidation). Therefore, the initial analysis is only a start for understanding the contact theory.

The relay contact resistance is influenced by material proprieties of the contact surfaces and it is limited by the contacting asperities [KAM 2009B]. More details on the contact resistance will be given on section 3.4.

In conclusion, the electrical contact between two rough surfaces is a complex physical interaction between thermo-mechanical deformation, current flow and heating between the contacting surfaces, which can only be determined probabilistically.

2.2.6 ENDURANCE

The endurance of a device is an important parameter for IC applications, and it is a requirement to reach before mass-producing a device. The industry standard requirement is a least 10 years of lifetime. The endurance of a NEMS relays highly depends on how many times the relay needs to be switched (*i.e.* to make and break a mechanical contact) and the mechanical wearing is one of the most disruptive phenomena for NEM relays.

[NATHANAEL 2012] estimates a benchmark for relay-based microcontroller for embedded applications operating at 100 MHz with 0.01 average transition probability: the relay should be able to undergo $\sim 3 \times 10^{14}$ on/off cycles over 10 years. In case of a nonvolatile memory, the endurance requirements are a little more relaxed. In fact, nonvolatile memories in general tend to have limited number of writing before starting to degrade. For example, PCRAM or STT-RAM have endurance of 10^6 - 10^9 and 10^{12} writes [LU 2018]. The endurance issues can be mitigated by adding to the memory elements a selector which will activate the memory only when strictly needed. In case of NEM relay, instead of using a selector, it is possible to modulate the signal pulse period. In fact, the movable electrode pulls into the contact electrode after a certain mechanical delay t_m , the relay to be activated would need to have a signal pulse with an offset larger or equal to t_m .

The NEM endurance is influenced by which type of switching mode is used. The relays are usually tested under two different types of relays switching condition: the cold switching and the hot switching mode. The third switching type is the mechanical one, but in this paragraph only cold and hot switching have been considered. During the cold switching, the relays does not carry any signal when is closed and opened. Therefore, no current flows when the switch is closed, and no current is interrupted when the relay is opened. In contrary, during the hot switching test, the relay is closed and opened carrying a signal. Therefore, when the switch is closed a current is free to flow, and when the switch is opened a current is interrupted. The abrupt interruption of current might cause arcing effect.

Testing in cold switching mode has the advantage of longer switching life, with respect the hot switching mode. In memory applications, the cold switching mode is more likely. In fact, a 5-terminals switch (single pole-double throw) can be activated to the side of the grounded actuation electrode (for example left) if its voltage is high ($V_{DD} > V_{PI}$), and to the sided of the biased ($V_{DD} > V_{PI}$) electrode (*e.g.* right) if it is grounded while the contact electrode can be left floating.

The main failure mechanism that can cause poor endurance in relays are typically associated to contact or structure layers.

Below a list of common degradation/failure mechanism adapted from [MERLIJN VAN SPENGEN 2003] and [VINCENT 2011]:

Structural Material:

1. **Fracture** especially of the structural material caused by:
 - overload fracture
 - fatigue fracture
2. **Creep** (*i.e.* gradual deformation under stress) due to:
 - Applied stress
 - Intrinsic stress
 - Thermal stress

Structural and Contacting Material:

3. **Stiction** due to:
 - capillary forces (presence of humidity layer between the two surfaces)
 - van der Waals molecular forces (adhesion between two extremely smooth surfaces, the action range is about ~ Angstrom)
 - solid bridging
4. **Wear** such as:
 - adhesive (bonding between the contacting solid surfaces, causing material transfer or material loss from either contacting surface)
 - abrasive (removal of material from a surface by a harder material)
 - corrosive (surface degradation due to the environment *e.g.* oxidation)
5. **Pitting** of contacting surfaces (localized corrosion that leads to small holes in the metal)
6. **Welding** of contacting surfaces due to:
 - Thermal effects (the two surfaces can locally weld because of the heat caused by the current flowing between the two surfaces)

[LEE 2009A] presents a 3 terminal out-of-plane actuated relay using TiN for the structure and the contact layer. They showed 40% of improvement for switching cycles in insulating liquid media compared to air. The liquid allows to slow down the contact degradation and to reduce the stiction force at the expense of a slower response time. The device is considered no longer operable when the

contact resistance becomes too high with respect the initial value. However, even in liquid the switching cycles is still low ~50 cycles (cold switching).

[CHEN 2012C] presents a 4-terminal out-of-plane actuated relay using W contact electrodes. The W contact electrodes encounters surface oxide formation that degrades the contact resistance. The paper shows 10^6 ON/OFF hot switching cycles (for 98% of duty cycle with a period of $1000\mu\text{s}$) and 10^8 ON/OFF hot switching cycles for 50% of duty cycle with a period of $40\mu\text{s}$ before the contact resistance starts to degrade. In the ON state, the surface degrades due to Joule effects and more the switch stays ON more the volume of heated material increases requiring more time to cool after the relay is switched off. The resistance degradation of long OFF state is due to the relay contacting surfaces being exposed to the residual oxygen, and the oxygen accumulations increases with the exposure time. Therefore, the best endurance performance was found in 50% duty cycle and high frequency (25 kHz).

[PARSA 2013] presents a 5-terminal polycrystalline silicon in plane-actuated relay with platinum coating for the contact surface. They show that after 10^8 hot switching cycles the relay has the same pull-in voltage at the first cycle but, higher contact resistance and higher pull-out voltage. These two degradation mechanisms are due to the modification and changes in the contacting surfaces caused by electrical and mechanical stress.

[HE 2013] presents a 3 terminal Silicon Carbide in plane-actuated relay. The paper shows more than 14000 hot switching cycles without failure (almost same Pull-In Voltage and slightly higher Pull-Out Voltage with one decade of ON current degradation).

[CZAPLEWSKI 2009] presents a 5 terminal Ruthenium in plane-actuated relay. The paper shows 2×10^6 cold switching cycles, after that the relay encountered permanent stiction. In the paper, there are not information about the electrical performance during the switching test.

[GROGG 2014] presents a 3 terminal in plane-actuated relay having as structural material Si, and as coating material for the contact surface amorphous Carbon. The paper shows an endurance of $\sim 10^8$ cycles in hot switches conditions. They observed an initial higher value of the ON resistance, which becomes stable only after $\sim 10^2$ cycles. The variation and decrement of the ON resistance in the first 100 cycles is probably due to the formation of multiple filaments in the contact tip, which contact the previously formed filament in the same location.

[OSOBA 2018] present a 6 terminal out-of-plane actuated having a structural layer polycrystalline silicon germanium and tungsten and perfluorodecyltriethoxysilane (PFDTES) as contact material. The paper shows that the fabricated relays are functional for at list 100 cycles in hot switching condition. The pull in voltage stays very stable for the first 20 sweeps, with a minimum pull out voltage variation of

(~20mV), probably due to variation in the actual contact area for each switching cycle.

[RIVEROLA 2018] presents a 5-terminal in plane-actuated relay made by using the tungsten via layer of 0.35 μ m commercial CMOS technology. The relay is swept 2800 times showing stable and repetitive V_{PI} with a maximal variation of the original V_{PI} 0.5V. However, there is an evolution of the contact resistance, it sweeps up from 1M Ω to 100M Ω in the first 200cycles. Giving a confirmation of the continuously formation of the native oxide at the contact interface during cycling.

Here below for an easy and immediate visualization the summary of the endurance value cited above.

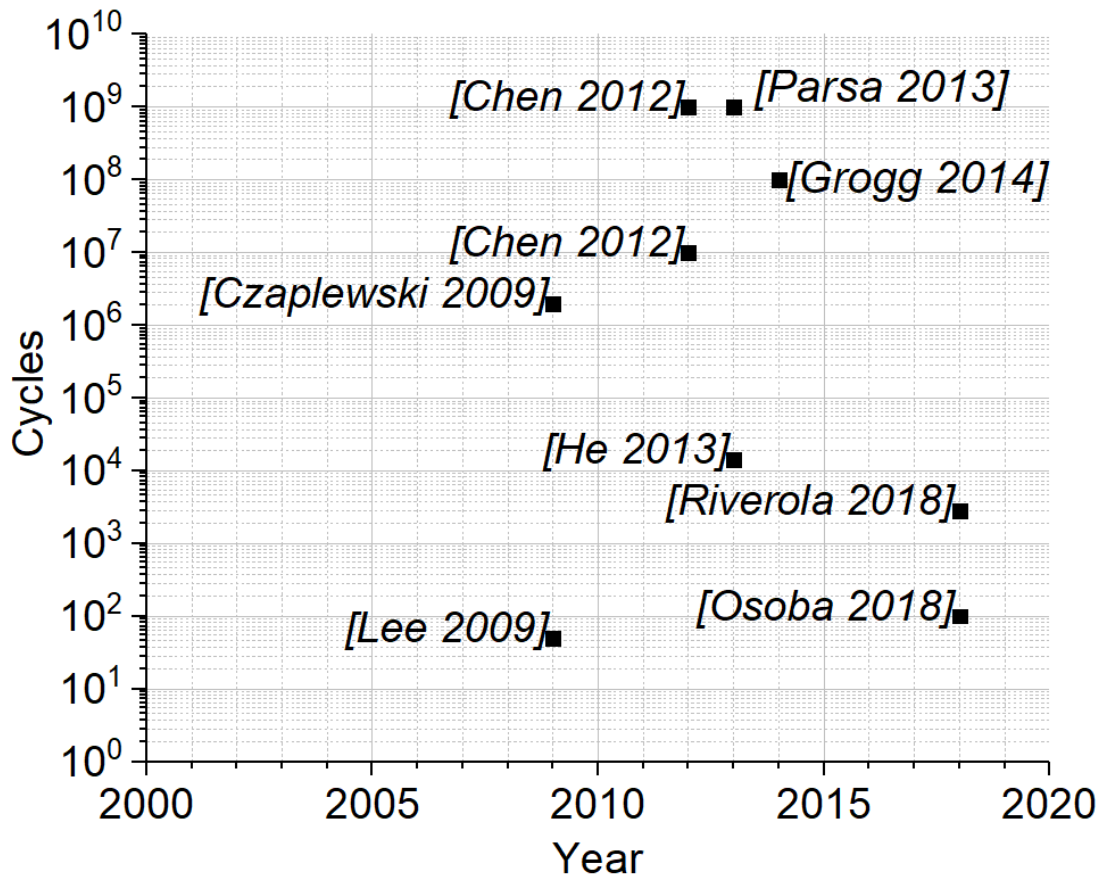


Fig. 40: Endurance State of ART: Cycles vs year of publications of NEM relays. The NEMS relays reported on this graph vary by design type, structural or contacting material and actuation type, Cycling is performed in hot switching conditions, except for Lee 2009 and Czaplewski 2009

2.2.7 MECHANICAL DELAY

Dealing with circuits made by nanoelectromechanical relay requires taking into account two main delay components: the mechanical delay and the RC delay. After applying a bias corresponding at least to the pull-in voltage value, the movable structure moves from its suspended position into the contacting position, and this time is referred as mechanical delay (t_m). t_m is inversely proportional to the natural frequency of the movable electrode. Therefore, it depends on the movable structure spring restoring force and, on its mass, and it depends on the applied voltage.

$$t_m \propto \frac{1}{f_0} \frac{V_{PI}}{V_{act}} = 2\pi \sqrt{\frac{m}{k}} \frac{V_{PI}}{V_{DD}} \quad \text{Eq. 11}$$

For instance, an ultra-scaled cantilever-beam relay design with a graphene structure will have a mechanical actuation time of $t_m \sim 10.3\text{ns}$. Having the young modulus $E_g \sim 2.5\text{TPa}$, the density $\rho = 2267\text{Kg/m}^3$, beam length $L_b = 300\text{nm}$, beam thickness and width $t = w = 7\text{nm}$, actuation gap $= 8\text{nm}$ and $V_{DD} = V_{PI}$.

The RC delay is the capacitive charging time of the load capacitance through a resistance, which usually is the relay R_{ON} . Depending on the relay contact material and on what the relay is driving, it can happen that the mechanical delay is several order larger than the RC delay making NEM relay limited by the mechanical delay [Qian 2018].

In Chapter 3, the delay will be explained considering the overall switching delay of the circuit. It will be taken into account the circuit layout and the relay design, the circuit operation (writing/reading/rewriting) and the relay materials. In fact, mechanical motion tends to dominate the switching delay of a single relay. However, the overall circuit switching delay is affected by the electrical delay as well.

2.2.8 SCALING LIMITS

NEMS relay should be scaled to reach minimum actuation voltage hence minimum switching energy, minimum actuation delay and minimum footprint, respecting the following condition and limitation:

- High device density
- Process technology limitation
- Contacting surface adhesion forces limitation

In order to switch on a NEM relay a sufficiently large voltage V_{DD} is required to generate an amount of electrostatic force F_{ele} larger than the spring restoring force F_k . Therefore, to reduce V_{DD} , F_k has to decrease:

$$F_k = k g_d \quad \text{Eq. 12}$$

In general, reducing the strength of the spring restoring force can be achieved by minimizing the structural stiffness (k) of the NEM relay or by reducing the contact gap (g_d). Minimizing (k) or reducing g_d allows the relay to be actuated with less amount of electrostatic force F_{ele} , therefore with a smaller V_{DD} and a smaller switching energy ($\propto V_{DD}$).

For lateral actuated relay the minimum g_d is given by photolithography resolution and for vertical actuated relay by deposition method capability and release process. However, the adhesion force limits the minimum reachable F_k . In fact, volatile relay (used mostly for logics) needs to obey the following condition:

Volatile Relay

$$0 < V_{re} < V_{PI} < V_{DD}$$

$$F_{adh} < F_k < F_{ele}$$

This means that the surface adhesion forces dictate the minimum allowable value for F_k . The surface adhesion forces limit the spring restoring force both in the minimum achievable gap and in the minimum value for the spring constant k . For a given k , after a certain gap, the surface adhesion forces are stronger than the spring restoring force, and the movable electrode collapse into the fixed electrode. Fig. 41 shows the minimum achievable gap before the adhesion force overcomes the spring restoring force for a k value of 0.1 N/m. The curves are plotted considering a cantilever beam Si relay of beam length L_B width w , thickness t and Young's modulus of E and Hamaker constant of $18,65 \cdot 10^{-20} \text{J}$ [LEITE 2012]. In this graph, the only adhesion force under consideration is the only Van der Walls force.

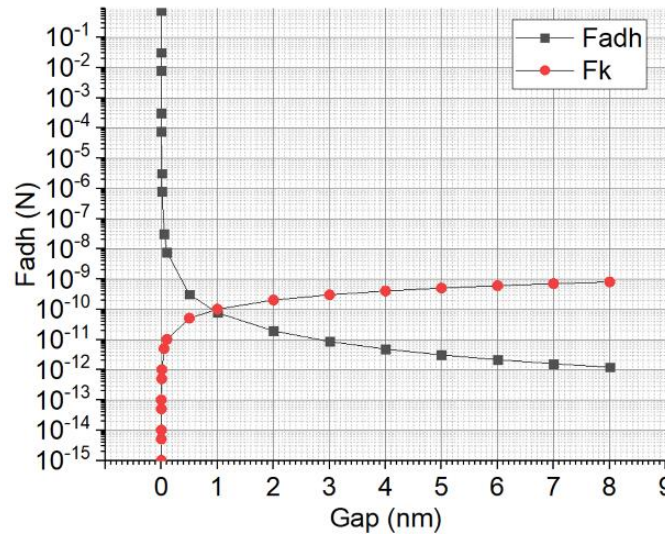


Fig. 41: After a certain gap ($\sim 1 \text{nm}$), for a given beam stiffness ($L_b=100 \text{nm}$, $t=w=7 \text{nm}$, $E=169 \text{GPa}$) the adhesion force F_{adh} overcomes the spring restoring force F_k .

Another way to reduce the switching energy is by increasing F_{ele} without increasing V_{DD} . To obtain a stronger F_{ele} (Eq. 13), it could be possible to decrease the gap g_g between the actuation electrode and the movable electrode, it could be possible to

increase the actuation area A_{ACT} , or it could be possible to use the body bias technique.

$$F_{ele} = \frac{1}{2} \frac{V_{DD}^2 A_c \epsilon_0}{g_g^2} \quad Eq. 13$$

The reduction of the g_g increases quadratically the electrostatic force resulting in a need of a smaller actuation voltage, therefore a smaller switching energy. However, to avoid the movable electrode collapsing into the actuation electrode (catastrophic pull-in CPI), the g_g should be larger than g_d .

The actuation area A_{ACT} size is constrained largely by the requirements of high device density (*i.e.* small footprint). A solution to increase the actuation area without area penalty is to prefer vertically oriented compact relay design such as [SIKDER 2017].

The third option or possibility is to use the body-bias technique. This technique has been proposed, especially for logic relay (*i.e.* volatile) [CHEN 2012B; NATHANAEL 2009; QIAN 2015; RANA 2014]. Body bias exploits a fourth terminal negatively or positively biased depending on the switch polarity, which pre-actuates the movable electrode allowing to turn on the switch at lower V_{DD} . After release, it can happen that the residual stress present on the structural layer of the relay causes a variation of the nominal actuation/contact gap and the body bias is very useful to adjust this gap without increasing the nominal V_{DD} . The applied body bias voltage V_B allows pre-actuating the relay, resulting in a need of lower V_{DD} to obtain the same amount of F_{ele} as without body bias and larger V_{DD} . In general, lower switching energy is obtained for value of pull-in mode operation *i.e.* $g_d > (g_g/3)$ [KAM 2011]. However, smaller switching delay is obtained for non-pull in mode operation. Therefore, there is a trade-off between optimal switching energy and switching delay. [QIAN 2015B] shows that thanks to body bias, it is possible to minimize switching energy of a relay over a range of switching delay. It means that a conventional switch (without body bias) has higher delay and higher V_{DD} with respect the one exploiting the body bias technique.

Despite that, there is a limit on the amount of body bias that can be applied, hence there is a limit on the minimum V_{DD} . The movable electrode of volatile switches does not go back to the original position until the applied voltage has reached a value of V_{RE} that is smaller than V_{PI} . This pull-out voltage is influenced by the adhesion force and by the hysteretic behavior typical of a device operating in pull-in mode. V_B is limited by the release voltage V_{RE} *i.e.* $|V_B| < V_{RE}$ to avoid permanent stiction. Therefore, the minimum pull-in voltage is also limited by the pullout voltage (or release voltage).

[DADGOUR 2011] studies the impact of relay size scaling in the electrical contact, analyzing the electrical and mechanical proprieties of mechanical contacts.

The paper highlights how scaling degrades the contact performance and reliability of such contacts. They observed that the contact resistance increases faster (exponentially) than the predictable increment caused by the only contact shrinking (*i.e.* quadratically). The nominal contact area is reduced; however, the actual contact area can be reduced even more. In fact, scaling the contact gap reduces significantly F_{ele} resulting in a smaller applied contact load. If the contact material is very hard, this effect could cause an actual contact area much smaller than the nominal shrunk contact area. This not only increases the contact resistance but also could be a problem for non-volatile application, especially if the adhesion force generated by the asperities spot in contact is lower than F_k . Additionally, for aggressively scaled logic NEMS, the contact adhesion forces could be predominant with respect the spring restoring force, resulting on permanent stiction of the movable electrode into the contact electrode. Moreover, the huge increment of the contact resistance for scaled contact could make the nanoelectromechanical switch RC delay predominant with respect the mechanical delay. In conclusion, relay scaling could be limited by contact performance and reliability degradation.

2.2.9 HARSH ENVIROMENT

A harsh environment can generally be described as an environment for which an entity is not intended. Given the vast number of electronic materials available, an environment that may be harsh for one application could be ordinary for another application. However, for most of the materials used in electronics it exists a common range of factors that may be used to characterize harsh environments. These factors include temperature, chemical, radioactive, and mechanical stresses.

Operating temperatures outside the range -40°C to 150°C are generally considered harsh. Usually the temperature considered is the environment temperature, which is because the heat source comes more frequently from the environment rather than self-heating. In some special cases such as high-power dissipation or very dense packaging, self-heating needs to be considered and added to the environment temperature. Usually in electronics, 150°C is the typical temperature from which performance degradation starts to appear. Even if some emerging technology and some commercially available industries products are capable of working at temperature of $210\text{-}225^{\circ}\text{C}$. Applications that needs high-temperature electronics and sensors are down-hole oil and gas industry [WATSON 2015], avionics and aerospace [BEHBAHANI 2012] and automotive [JOHNSON 2004].

Most semiconductor electronic components are susceptible to radiation damage see Fig. 42. However, radioactive stress is particularly present in outer space and high-altitude flight, around nuclear reactors, particle accelerators, and even in nuclear accidents or nuclear warfare. Radiation stress is caused by the interaction of the radiation (mostly photons and particles) with the materials composing the

electronics. The two main fundamental damage mechanisms due to radiation stress are: Si lattice displacement and Ionization effects. The first is caused by the impact of heavy particles with the Si atoms. The impact can change the arrangement of the atoms in the crystal lattice or can cause local annealing causing loss of transistor gain. Ionization effects especially on SiO₂ caused by charged particles gives transient effects that cause soft and hard errors, the energy absorbed by the ionization liberates charge carriers, which diffuse or drift to other locations, leading to parasitic fields [SPIELER 1997].

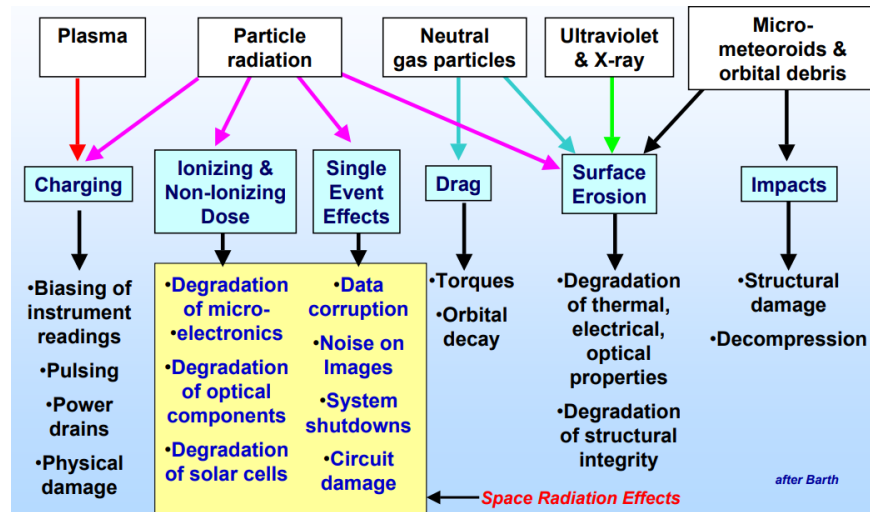


Fig. 42: Description of space environment and related effects in electronics from [Kenneth 2004].

In contrast with MOSFET, NEMS relay are anticipated to have high tolerance to high/low temperature and an inherent resistance transient radiation (single-events and photo-currents) [ARUTT 2017]. Especially Silicon Carbide (SiC) NEM relays are good candidate for harsh environments because SiC is an high temperature resistance material [MÜLLER 1994]. Case Western University demonstrated SiC NEM switch with reliable switching at 500°C having a pull in voltage of 11 V [HE 2013]. The device had an endurance of 14,000 cycles under “hot-switching” conditions.

2.3 TARGET APPLICATIONS

2.3.1 LOGIC NEMS

For digital logic, the function of the NEMS relay is quite similar to the MOSFET except for some characteristics.

The MOSFET can operate in three main regions: linear (triode), cut off and saturation (quadratic), hence it can act as a current source, variable resistor and high-impedance element. On the other hand, the NEM relay can operate in two: open circuit (OFF) or fixed resistor (ON).

They also differ in dynamic switching behavior. For the case of n-MOSFET, once the applied voltage exceeds the Threshold voltage, the CMOS begins to charge/discharge the output capacitance with a delay given by the MOSFET on-resistance R_{ON} and the load capacitance: $t_{RC}=R_{ON}C_L$. On the other hand, the NEM switch delay is dominated by the time it takes to mechanically displace the movable electrode and pull it into contact with the drain (*i.e.* mechanical delay) plus the additional RC delay (which is proportional to the relay contact resistance R_C).

The voltage difference applied between the actuation electrode and the movable electrode can be positive or negative, in fact, the two electrodes are like the plates of a capacitor and the electrostatic force between them is always attractive. This means that a single relay can be used both as N-type and as P-type relay. Then, two relays can be wired together to form, for example, an inverter Fig. 43 Left. In the inverter, only a relay is ON at a time, resulting in zero crowbar current [FUJIKI 2014].

Additionally, both the inverter stage and buffer stage can be implemented using only a stage made by two relays, instead of the case of the CMOS circuit in which the buffer is made by two stages (4T). In fact in 4T relays the source and drain voltages are independent from the input signal (I or B) and both relays are equally good at driving the signal to ground or to supply [KAM 2015]. The NEM relay buffer delay will be the same as the single inverter delay, (*i.e.* mechanical delay + relay RC delay), while the buffer made by the MOSFET will have a delay comparable of two inverters. In summary, optimized relay-design, having all mechanical movement happening simultaneously can and should be arranged [CHEN 2008].

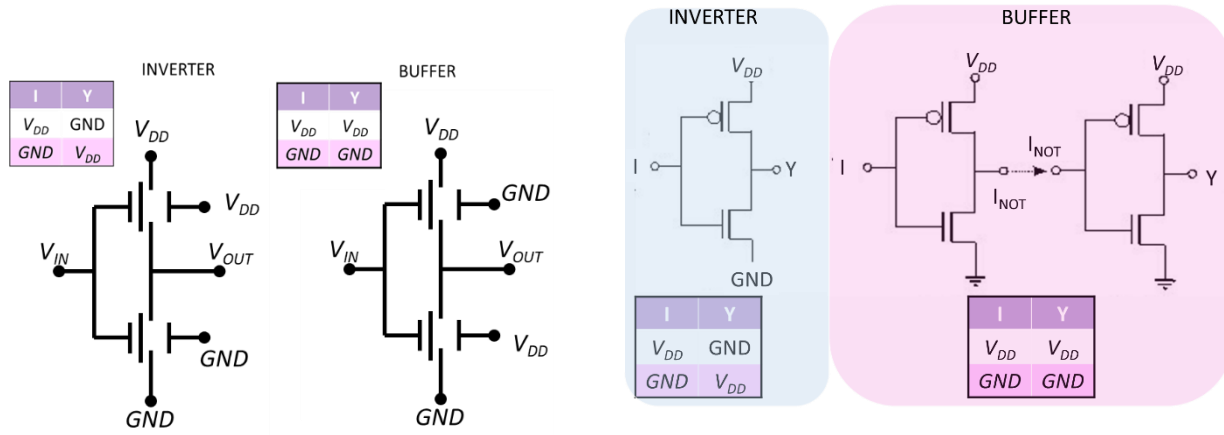


Fig. 43: Logical equivalent logic NANO-relay (Left) and CMOS (right) circuit.

Many integrated circuit design with NEM relay have been proposed by [CHEN 2008]. In general, the use of the NEM relay is to extend the energy efficiency beyond the CMOS limitation due to the sub-threshold leakage of the already existent digital system such as Digital Analog Converter or Analog Digital Converter. More information about design complementary NEM for logic gate and digital application can be found in [AKARVARDAR 2007; JEON 2012; JEON 2010; KAM 2011], DC-DC Converter in [MANOHAR 2012], Data Converter [LI 2018] and for sequential logic [VENKATASUBRAMANIAN 2013].

2.3.1.1 NANO-ELECTRO-MECHANICAL RELAYS FOR FPGA ROUTING

FPGA enables low design costs and quick turnaround (*i.e.* small time between the submission of a program/task and return) [KUON 2007]. NEM relays are promising candidates for programmable routing in Field-Programmable-Gate Arrays (FPGAs) because of their zero leakage and potentially low on resistance [CHEN 2010]. The low relay endurance is not a big problem for FPGA routing, because FPGA routing relays are usually subjected to a limited number of reconfigurations (~ 500) and since the programmable routing switches does not change states after configuration, the NEM relay mechanical delay does not affect the FPGA performance [KUON 2007]. The routing relay could be implemented on the top of the SRAM without footprint penalty [Kwon 2017; Muñoz-Gamarra 2014]. The NEM switch hysteresis can be used to make FPGA programmable routing switches which do not require configuration of SRAM cells [CHEN 2012A]. The SRAM-based FPGA uses an SRAM to drive a NMOS pass transistor to implement programmable routing. Due to technology scaling after the signal pass through the NMOS pass transistor there is a threshold voltage drop, therefore there is a half latch-based buffers to restore the signal and speed up the rising edge. This voltage drop introduce performance penalty and the buffer introduce area penalty. [CHEN 2012A] shows that NEMS

relay-based programmable crossbars stacked on top of CMOS can be used to replace the SRAM and the NMOS routing switch Fig. 44.

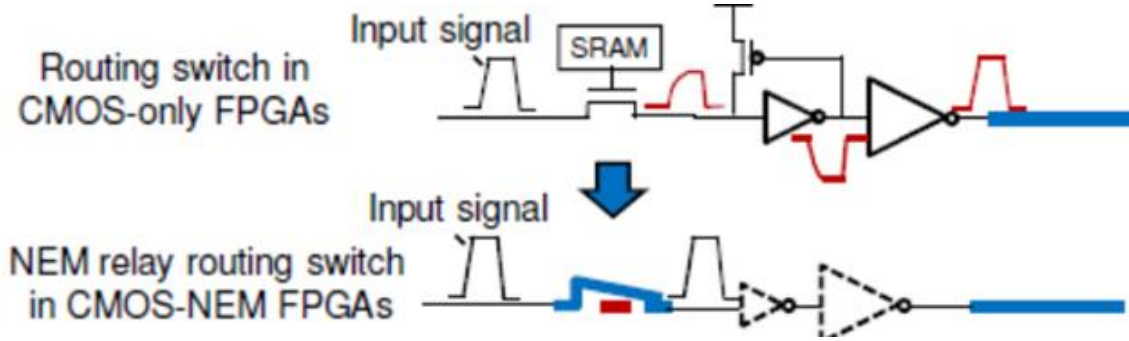


Fig. 44: Top) CMOS-only FPGA schematic showing the voltage drop with NMOS pass transistor as routing switch. Bottom) CMOS-NEM FPGAs schematic without voltage drop thanks to NEM relay. [Chen 2012a]

However, for FPGA containing millions of configurable routing switches is very difficult to configure correctly all the NEMS relays. In fact, NEMS relay voltage variability is large, causing a large variation between the relays pull in voltage V_{PI} and pull out voltage V_{RE} resulting in a very small tolerable noise margin (*i.e.* $Minimum \{V_{PI} - V_{RE}\} > V_{PI,max} - V_{PI,min}$) [CHEN 2012A].

Other example of NEMS relay used for FPGA routing can be found in [Dong 2011; Lee 2012; Zhou 2007].

2.3.1.2 LOGIC NEMS RELAY IN SRAM

NEM relay have been also proposed to help SRAM to be more power efficient. [BOTA 2017] analyzed the benefits of replace two or more MOSFET transistors of a conventional CMOS six-transistor SRAM cell with NEM relays. The paper shows that the substitution of the CMOS transistor by NEM relay reduce the leakage power by 1.2x if the NEM relays are at the place of the pull up transistors, by 2.6 id the relays are placed instead of the pull down transistor and of 4.4x if both pull up and pull down transistor are substituted. The intrinsic reading delay does not vary much between the 6T CMOS SRAM and the CMOS/NEMs cell. A failure to write the cells occurs when the pass transistor is not strong enough to overpower the pull-up relays. To be able to properly write the CMOS/NEMS cell having NEM relay at the place of the pull up transistor the relays ON resistance has to be larger than the pass transistor resistance. If the NEM relays are substituting only the pull-down transistor Fig. 45, the write speed is not limited by the relay, however the presence of the relay introduces a latency delay equal to the mechanical delay. NEMS relay could be an interesting option to help CMOS SRAM to be more energy efficient [CHONG 2009].

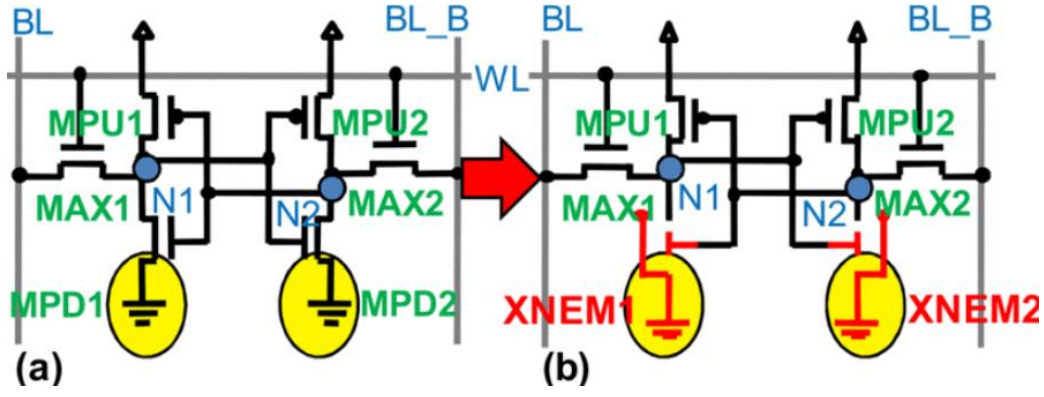


Fig. 45: Schematic of (a) a conventional CMOS 6T SRAM cell and (b) a NEM/CMOS hybrid SRAM cell with NEM relays replacing the pull-down NMOS. Figure from [Chong 2009]

2.3.2 NON-VOLATILE NEM RELAY

Internet of things (IoT) requires performing real time operation with high-energy efficiency. In-memory computing IMC architecture have been proposed to help IoT applications because being able to compute the information stored in site is much faster and less expensive than transfer the information from the memory to the CPU as is done in current Von Neumann architectures.

However, IMC to take the lead requires having a memory that have small footprint and that does not consume a lot of power. ReRAM-based IMC architecture have been proposed to address the challenge [Halawani 2019; Mittal 2019]. At our knowledge, until now the non-volatile memory footprint is limited by the access transistor size.

Non-Volatile NEM relay could be suited for IMC because of the zero leakage and the very high R_{OFF}/R_{ON} ratio and very low programming energy compared to the other NV devices see Fig. 46. NEMS relay can be used as nonvolatile memories elements, to implement circuit architecture for energy-efficiency computing.

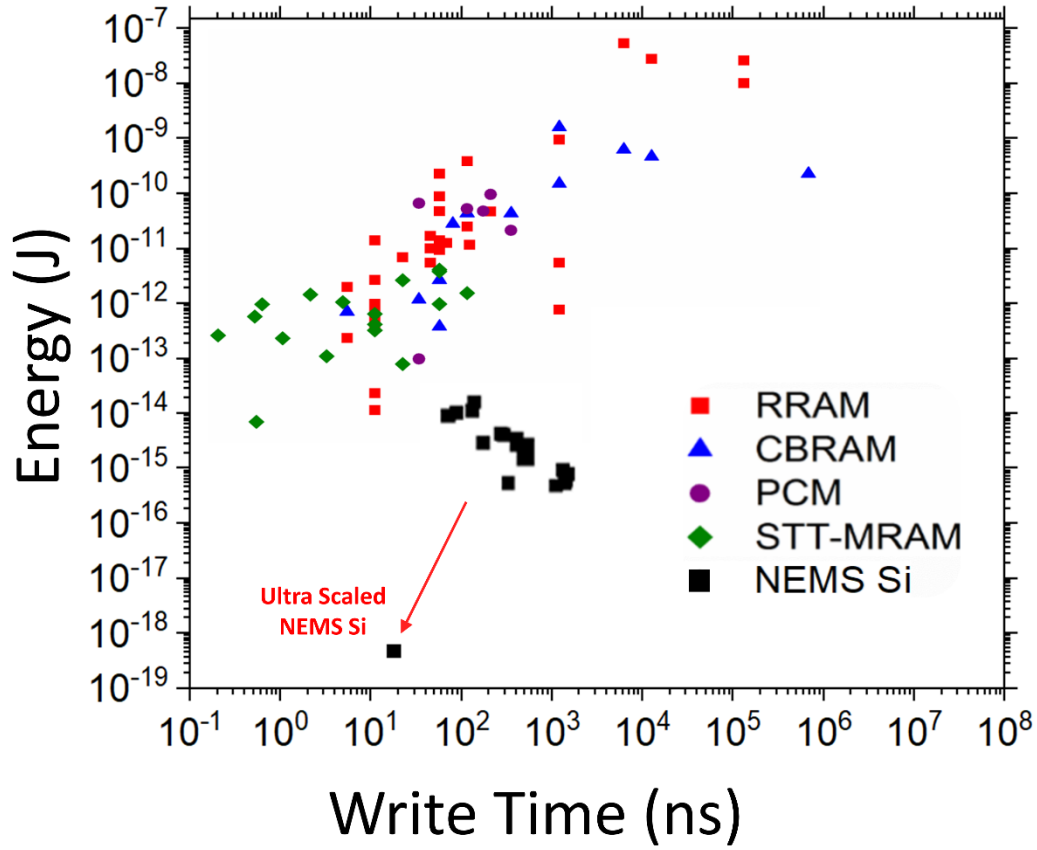


Fig. 46: Switching energy vs switching time for NV emerging device. RRAM, CBRAM, PCM, STT-MRAM data from: [Wong 2018] compared to NEMS Si data simulated made for the scope of this thesis. Additionally, a projected Ultra scaled NEMS relay is represented showing the value obtainable using more aggressive technology then the one used in this thesis

2.3.2.1 LOOK-UP TABLE

Lookup table are very useful instrument to realize fast computation. The input signal gives the address in which the data is stocked, and the output signal gives the data stocked on that address. The NEM relay LUT can be designed by $N+M$ columns, where N is the number of the inputs and M is the number of the outputs and with 2^N rows which corresponds to the possible input combination Fig. 47 Left. Each cell memory cell is made by the movable electrode, which is anchored at the bit line, and has two-actuation electrode on either side to program it. The contacting electrode one in either side of the movable electrode are implemented on the input/output line Fig. 47 Right.

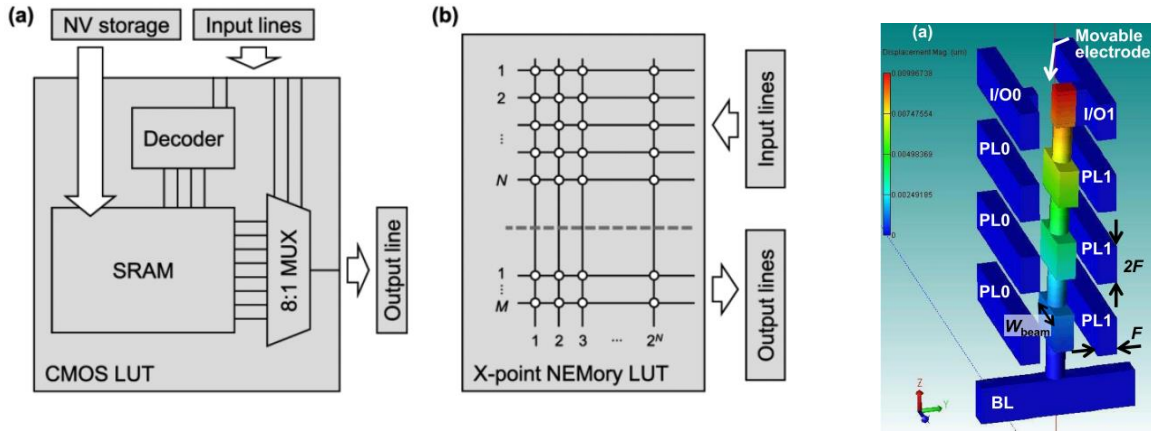


Fig. 47: Left) Circuit block-level comparison of CMOS Lookup Table and NV-NEM based. Right) Schematic of the NEM relay highlighting the contact electrode connected to the I/O Input/output line, the programming electrode connected to the programming line PL and the anchor connected to the bitline BL. From [Kato 2016a]

The advantage for the NEM relay used for LUT is that during the program operation, the movable electrodes is actuated in cold-switching condition, because the input and output node are floating resulting in a better endurance. The read operation requires enabling the read line, the input line is precharged at GND and the output line are precharged at V_{DD} . The input lines are driven to V_{DD} , besides the input line that corresponds to the wanted bit combination. Then, the output lines are enabled causing one of each pair to discharge toward grounds. The LUT architecture can be done using the metal interconnection levels of CMOS technology for instance.

[KATO 2016A] shows the energy, delay and area projection of scaled nonvolatile NEM Cell Lookup Table, in comparison with the conventional CMOS based or ReRAM based LUT. It results that NEM relay LUT have lower readout energy, delay and similar high density as ReRAM Fig. 48.

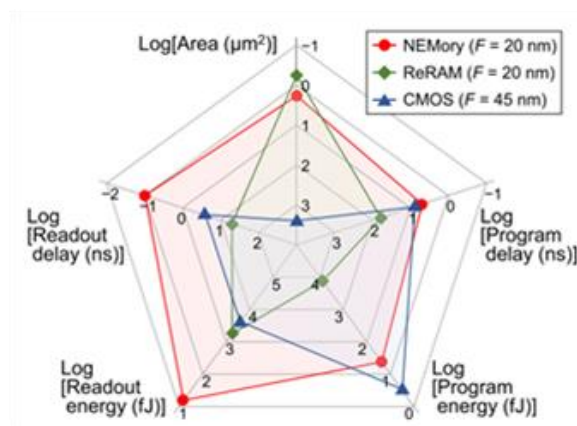


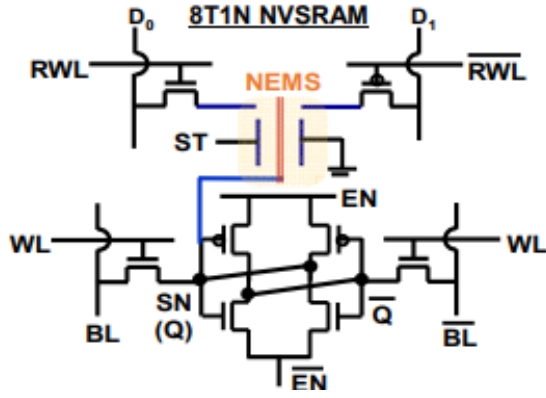
Fig. 48: Radar plot comparing the performance characteristics of a lookup table implemented with an array of NEMS on top of CMOS circuitry vs. Conventional only CMOS-based LUT (data from [Paul 2014]) and ReRAM-based LUT data from [Govoreanu 2011]. Figure from [Kato 2016a]

2.3.2.2 NV-SRAM

NV-SRAM has been proposed to maximize energy efficiency without degrading performance. Reducing energy consumption caused by the inactive circuit blocks could require using power-gating technique. However, in case of volatile memory, the need to store/restore state to/from memory blocks each time part of the circuit is powered off/on requires a lot of energy and time. 3-D monolithic integration of non-volatile memory (NVM) element within a single SRAM cell will help in achieving fast power-off/on speed with comparable energy [CHIU 2010B]. Various NVM devices such as MRAM, FeRAM, and ReRAM have been used to implement NV-SRAM cells.

However, NVSRAM requires either high voltage or significant current to program the nonvolatile element. It is often necessary that the NV device require an additional selector device to be isolated during SRAM operation. It can also be possible that the NV-store function needs to be separately implemented prior to power-off. Hence, NVM devices are often not sufficiently energy efficient to follow the content of an SRAM cell, therefore this NVSRAM solution has not been largely implemented.

NV NEMS switch offers a solution of the above issues. The NV cell can be built using a 6T SRAM cell with 1NV switch and two additional transistors to control and restore the data once the circuit is switched ON again. The NEMS can be driven by the same VDD value used to bias the SRAM cell. The NEMS actuation delay is slower than the SRAM actuation delay, this could help obtaining relaxed endurance requirement. Before powering OFF the circuit, a delay slightly larger than the mechanical delay (min of 1ns to $\sim 0.5\mu\text{s}$) should be waited. This will ensure that the data Q, stored on the storage node SN is saved; (*i.e.* the NEMS relay is in contact with the corresponding contact electrode (*i.e.* D0 if Q is low and D1 if Q is high)). The energy spent to store permanently the data is equal to the relay switching energy ($\sim 10^{-16}\text{J}$ - 10^{-18}J). To restore the data into the SRAM cell, the RWL line is enabled, D0 is driven to the ground, D1 is biased at VDD then the SRAM cell is powered ON and the stored data Q is automatically written back in the storage node.



Operation	EN	WL	BL	RWL	D _(0,1)	ST
WRITE	V _{DD}	V _{DD}	DATA	GND	Open	V _{DD}
READ	V _{DD}	V _{DD}	(pre-charged to V _{DD})	GND	Open	V _{DD}
HOLD	V _{DD}	GND	N/A	GND	Open	V _{DD}
STORE	GND	GND	N/A	GND	Open	GND
RESTORE	V _{DD}	GND	N/A	V _{DD}	{GND, V _{DD} }	GND

Fig. 49: Left) 8T1NEMS NVSRAM cell design with Right) bias voltage for Write/Read/Store/Restore operations from [Xu 2014a]

NV relay promise lower switching emerging and comparable switching time compared to the other nonvolatile elements see Fig. 46.

The NEM switch input capacitance ($\sim 0.01\text{fF}$) is much lower than that of a storage node ($> 0.1\text{fF}$) in a standard 6T SRAM cell, therefore the NVNEMS-SRAM cell should operate with speed and energy comparable to that of a conventional 6T SRAM cell [XU 2014A].

2.3.2.3 NV-TCAM

Ternary Content Addressable Memory (TCAM) is a type of memory used for high-speed searching applications. The input data is compared to the stored data and returns the address of matching data. It is possible to find large data set, and there is the possibility of using masks for one or more bits, resulting in large flexibility to the search. In contrary of classic memory system, TCAMs circuits allow to search the stored information by its content [LY 2018]. This characteristic can be exploit for Big Data application, Network Routing, Reconfigurable Computing Analytics [KARAM 2015]. Historically TCAM have been implemented using CMOS devices, but recently emerging nanoscale devices such as FeRAM, STTRAM, PCRAM, ReRAM and NV-NEM switches shows promising capability for TCAM. In fact, this NV devices can have lower search and update latency/power, lower leakage current compared to the CMOS technology and additionally non-volatility of stored information.

A typical TCAM search principle rely in three operations:

1. **Pre-charging** the match line (ML) at VDD_ML.
2. **Set** the search-line (SL) *i.e.* insert the data to be found in as input in the SL.
3. **Evaluate**, the ML is left floating, and during the evaluation operation ML discharge slowly through leakage current IML_m (match leakage

current) Fig. 50 a) or if at least one bit mismatches the input search data, then ML quickly discharge with an higher current IML_mis (match leakage current) through the mismatched bit Fig. 50 b).

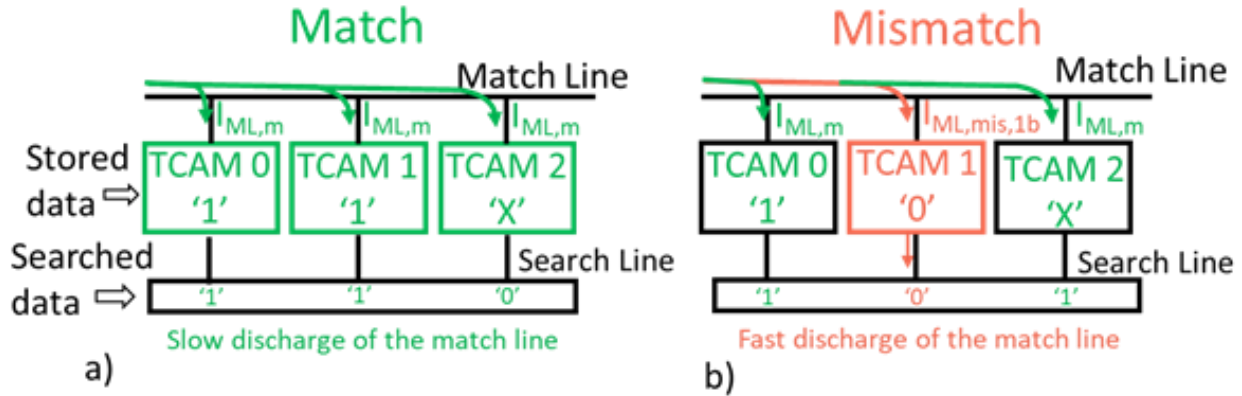


Fig. 50: Example of search operation. a) ML slowly discharge because the stored data matches with the search data. b) ML discharges faster because the stored data mismatches with the search data. Figure adapted from [Ly 2018]

For long TCAM words, it is very important that the resistance ratio between the OFF state and the ON state is large (in case of TCAM made by MOSFET $ROFF/RON \sim 10^5$). If the resistance ratio is not big enough, the sum of the leakage current IML_m in case of a match can be comparable to a mismatching current IML_mis . NEMS relay should have a large OFF/ON resistance ratio ($R_{ON} \sim kW$). Another characteristic of TCAM is that the write and read operation are well distinguished [ZAMBELLI 2015]. Most of the time TCAMs are accessed to perform search operations rather than writing, therefore the read/search operations are much less stressful for the devices [GROSSI 2018]. The reading/searching operation does not require a change of state ON/OFF, hence the NEMS relays should have a very high endurance. Applications such as hyper dimensional computing [Li 2017], requires 10^3 endurance cycles to obtain a high recognition accuracy while complex matching accelerators, an value of 10^4 endurance cycles can guarantee ten-year lifetime [ZHA 2018]. NEMS relay used in TCAMs combined with CMOS should be an interesting solution for TCAMs.

2.3.3 POWER GATING

Scaling has allowed increasing functionality per area, higher speed and lowering switching energy however it has increased the leakage. Additionally, the requirement to increases capability and performance demand to increases the numbers of transistors, which increases the leakage power. The mitigation of the leakage could be done during the stand-by periods, when no processing is occurring. Mobile phones and notebooks, wireless and sensor applications all of

them have idle time. That is why the industry has introduced the power gating technique.

Power Gating allows reducing the leakage path between the power supply V_{DD} and ground of the digital block during idle times. Power gating is achieved by adding between V_{DD} and ground a switch in series with the digital block, and turning off the switch during the idle time (see Fig. 51). Fine-grained functional unit power gating is when an independent switch is placed for each digital block. Coarse-grain power gating is the definition for the presence of a switch for a multiple number of digital blocks. The leakage is eliminated if the switch is ideal ($R_{OFF} = \text{infinite}$). However, the presence of a switch between the power supply and the digital adds an “ IR_{ON} ” voltage drop which increases the delay of the logic block. Hence, power gating requires to find a trade-off between the induced delay and the power leakage saving, which means that the switch resistance should be very high when the switch is OFF and very low when the switch is ON.

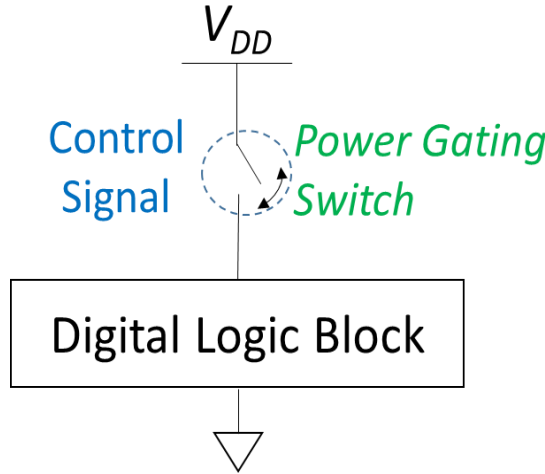


Fig. 51: Schematic representing the concept of power gating technique.

Many examples of power gating using NEMS relays can be found in the literature [FARIBORZI 2010; HENRY 2013; HENRY 2011; SAHA 2017B].

NEMS relay can offer zero leakage current, but the ON-state resistance depends on the relay design and structural/contact material. Additionally, NEMS relay have an intrinsic switching mechanical delay that can be lower than $\sim 10\text{ns}$.

[SANKAR 2017] describes all the important figure of merits to take into account while evaluating the total energy consumption using NEMS relays for power gating.

E_N : The total energy consumed by the circuit with NEMS power gating during time off T_{OFF} and time on T_{ON} (Eq. 14). P_{Active} : The average power consumed during the ON-STATE (Eq. 15). f_{LB} : The switching frequency of the Logic Block. f_{PG} : The

frequency of the power gating. I_{lea} : The leakage current during the OFF-STATE. I_{static} : The leakage current during the ON-STATE.

$$E_N = V_{PG}^2 f_{PG} C_{G,N} \left(1 + \frac{T_{ON}}{T_{OFF}}\right) + P_{Active} \left(\frac{T_{ON}}{T_{OFF}}\right) \quad Eq. 14$$

$$P_{active} = V_{DD} * I_{avarage} = V_{DD} [\alpha C_L f_{LB} + I_{static}] \quad Eq. 15$$

Where $C_{G,N}$ is the Gate capacitance of the NEMS power-gating switch and T_{ON} is the time the logic block is *ON* and T_{OFF} is the time which the logic block is *OFF*. The paper shows that the energy gain of using NEMS relay instead of FinFET technology (20nm/14nm/12nm) increases when the T_{ON} is much lower than T_{OFF} . Reaching an energy gain of 10^3 for $T_{ON}/T_{OFF}=10^{-5}$. Meaning that the use of NEMS relays for power gate is recommended for application that needs to stay OFF for the most part of the time. The paper also highlights that using NEMS relays for Power gating compared to FinFET gives a higher energy gain as the temperature increases.

Functional units in a SoC for mobile platform	Values used for			% of Net Energy saving using NEMS power gating
	α	T_{ON}/T_{OFF}	f_{Logic} (GHz)	
Application processor	0.1	10%	3.5	12.2%
Cache memory *	0.05 ⁺	10%	3.5	23.4%
Clock distribution *	1	10%	3.5	1.6%
DSP/GPU	0.1	10%	1.0	29.5%
Baseband processor	0.5	5%	1.0	16.3%

Temperature = 40 C. * Non-stacked gate. ⁺ Includes spatial activity also.

Fig. 52: Possible energy reduction using NEMS Power Gating in SoC made with FINFET (14nm). Figure from [SANKAR 2017]

2.3.4 REMAINING CHALLENGES

Microelectromechanical relays have been investigated for several decades, especially for RF signal applications. However, their use for logic or memory it is more or less recent (~10 years). Further work is needed to address three main challenges and for NEM relays to be real good candidate for ultra-low power in-memory computing.

2.3.4.1 CHALLENGES FOR REDUCING RELAY SIZE

Reducing the NEM relay size is quite challenging. Reducing its size, it is constraint by the requirement of low actuation voltages and the fact that the thickness of the

contact air gap has a minimum physical limitation. The minimum size of the actuation area is limited by the necessary amount of electrostatic force to overcome the spring restoring force. To reduce the relay footprint, a vertical relay design have been proposed [XU 2014A]. The vertical structure of the relay could be made by using multiple layer of metal (*e.g.* the metal and vias of BEOL CMOS technology). However, at our knowledge, the release steps for vertical relay structures are complex, needs further investigation and researcher are still working on it.

2.3.4.2 CHALLENGES FOR REDUCING OPERATING VOLTAGES

Obtaining a small footprint relay with low actuation voltages it is also limited by the contacting surfaces. Scaling the actuation voltages reduces quadratically the electrostatic force, which must be larger than the adhesion force. In relays, the energy efficiency limit is set by the adhesion force between the contacting surfaces [PESCHOT 2015]. The contact between nano-surfaces is asperities dominates, and the physical and chemical mechanisms between asperities dominated mechanical/electrical contact is yet to be completely understood. Researcher are working to understand and to meet the needs of nano-contact [DADGOUR 2011; HARRISON 2014; PROKOPOVICH 2011; SI 2015; XIANG 2012]. Low actuation voltages have been reached by applying body bias [CHEN 2012A; LEE 2013; OSOBA 2016; QIAN 2017; ZAGHLOUL 2014] and adding a surface coating that reduce the adhesion force, which allows to reduce the release voltage.

2.3.4.3 CHALLENGES FOR MONOLITHICALLY NEMS CMOS INTEGRATION

The monolithic co-integration between NEMS relay and CMOS circuitry is under development, papers shows results regarding the fabrication of relays using the back end of the line of CMOS technology [CHOI 2015; KWON 2017; MUÑOZ-GAMARRA 2016]. However, fabricating BEOL relay does not leave a large degree of freedom in the choice of structural and contacting materials. The use of the CMOS interconnection for relays has been promoted for vertically oriented design because of the possibility of a small footprint without voltage penalty.

3 MODELING OF NV NEM RELAYS

3.1 DESIGN AND MODELING OF NEM RELAYS

3.1.1 PARALLEL PLATE CAPACITOR APPROXIMATION

The simplest way to understand Micro and Nano electromechanical electrostatic relay actuation mechanism is to describe its behavior through the model of parallel plate capacitor approximation. The simplest relay design has two electrodes: a fixed electrode and a movable electrode. In order to prevent the total collapse of the movable electrode into the fixed electrode, dimples can be added to the movable plate. The two electrodes are equivalent to the plates of a capacitor separated by a gap, this gap vary together with the movable electrode displacement. The capacitor area (A) is proportional to the length multiplied by the width of the overlapping surfaces, the nominal gap thickness between the two plates is g_0 , and the nominal distance between the two dimples and the fixed electrode is g_d .

The static model of the electrostatic actuated parallel-plate capacitor is well described by a mass-spring system (Fig. 53).

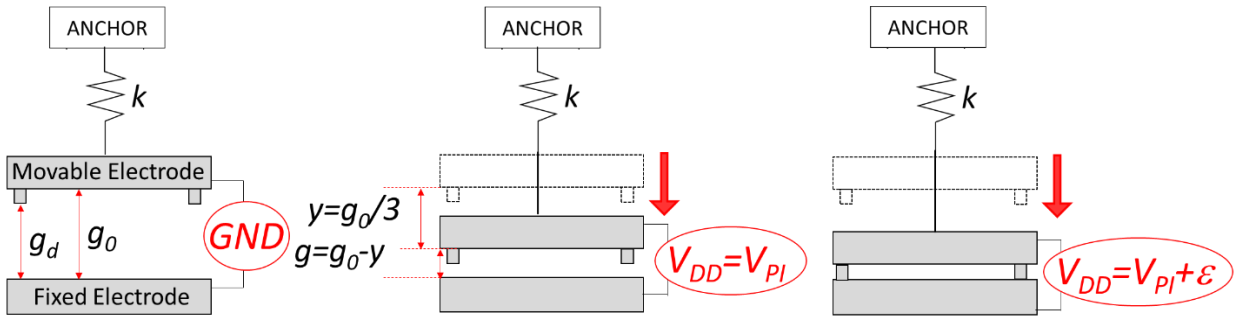


Fig. 53: Schematics diagram of the parallel plate capacitor NEM relay. Showing the important parameters, such as stiffness and actuation voltage and the critical displacement and remaining gap.

$y = g_0 - g$ indicates the displacement of the movable electrode with respect its original position, and k is the spring constant or stiffness, V_{DD} the applied voltage, V_{PI} the pull-in voltage.

When no bias is applied between the two plates ($V_{DD} = GND$), the switch is in the OFF state. Once a voltage V_{DD} is applied, an electrostatic force F_{ele} appears pulling the source towards the fixed electrode. The electrostatic force (Eq. 16) can be derived from the stored energy in the parallel-plate capacitor (between movable and fixed electrode). The spring restoring force is expressed in (Eq. 17)

$$F_{ele} = \frac{d}{dy} \left(\frac{1}{2} C_{plate} V_{DD}^2 \right) = \frac{\epsilon_0 A V_{DD}^2}{2} \left(\frac{d}{dy} \frac{1}{(g_0 - y)} \right) = \frac{\epsilon_0 A V_{DD}^2}{2(g_0 - y)^2} \quad Eq. 16$$

$$F_k = -ky \quad \text{Eq. 17}$$

Where ε_0 is the vacuum permittivity. When the structure is in equilibrium, the total force acting on the mechanical structure is zero (see Eq. 18). Therefore, by equating the electrostatic force with the spring restoring force the equilibrium equation is derived Eq. 19:

$$\begin{cases} F_{tot} = 0 \\ F_{tot} = F_{ele} + F_k \end{cases} \Rightarrow F_{ele} = -F_k \quad \text{Eq. 18}$$

$$ky = \frac{\varepsilon_0 A V_{DD}^2}{2(g_0 - y)^2} \quad \text{Eq. 19}$$

Fig. 54 shows both electrostatic force F_{ele} and spring restoring force F_k vs normalized source displacement characteristic. As highlighted in Eq. 16, the electrostatic force increases quadratically, while F_k increases linearly as the gap between the two plates decreases (see dotted purple line in Fig. 54).

Reading the graph from left ($y/g_0=0$) to right ($y/g_0=1$), at low V_{DD} (*i.e.* $V_{DD} < V_{PI}$), F_{ele} (dash-dot pink line) is larger than F_k ; however, F_k contends F_{ele} in some points (*e.g.* blue triangle). When the voltage V_{DD} increases and reaches a critical value ($V_{DD}=V_{PI}$), F_k contends F_{ele} (solid red line) only in a mono-stable critical point (corresponding to a critical displacement). A small increment on V_{DD} results in F_{ele} greater than F_k causing the gap to close abruptly.

This critical voltage value is known as Pull-in Voltage (V_{PI}). The pull-in phenomena for a parallel plate capacitor relay occurs when the displacement exceeds one third of the initial gap of the parallel plate. When V_{DD} is decreased the relay does not immediately open, but only when $V_{DD}=V_{RE}$ and F_{ele} cross again F_k (red circle).

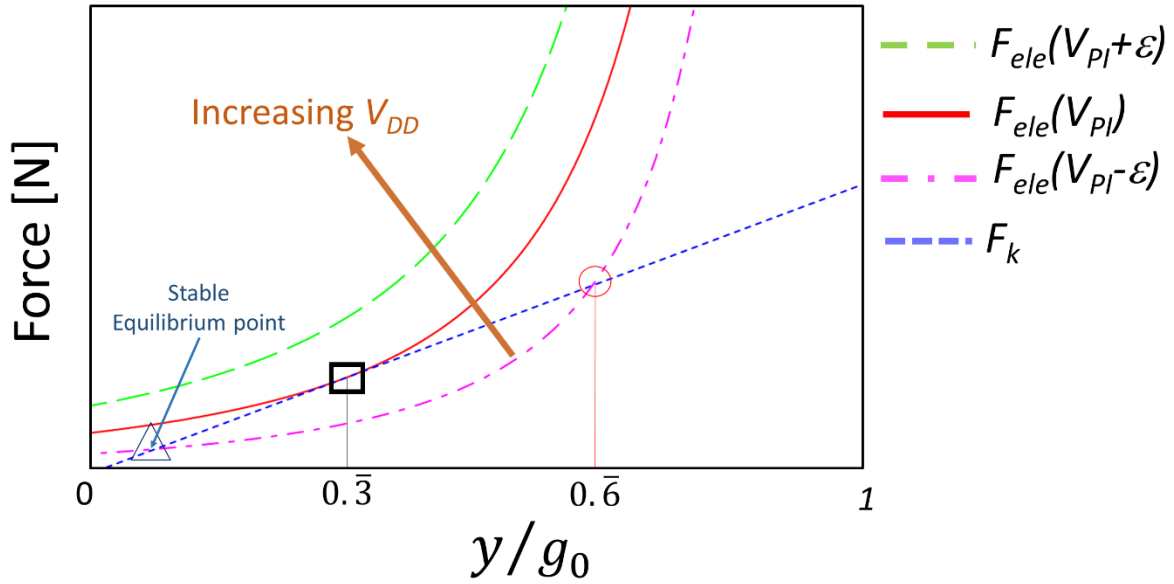


Fig. 54: Spring restoring force and electrostatic forces vs. normalized displacement y/g_0 characteristic for parallel-plate electrostatic actuator. F_{ele} is plotted for $V_{DD} < V_{PI}$, $V_{DD} = V_{PI}$ and $V_{DD} > V_{PI}$. The circle symbol indicates the point of release. The point marked by the square is called “Pull-In point”.

F_k is tangent to F_{ele} at $V_{DD} = V_{PI}$ and $y = y_{PI} = g_0/3$. The pull-in voltage for parallel-plate actuators is given by Eq. 20:

$$V_{PI} = \sqrt{\frac{8k g_0^3}{27 \epsilon_0 A}} \quad \text{Eq. 20}$$

To be more precise the parallel plate actuators having dimples is considered to be ON when the dimples are in contact with the fixed plate. Therefore, in presence of dimples the movable plate at contact is displaced from its original position by $y = g_d$.

The parallel plate capacitor is considered to operate in “pull-in mode” if $g_d \geq g_0/3$ and in “non-pull in mode” if $g_d < g_0/3$. When the relay operates in “pull-in mode” and the applied voltage between the two plates is $V_{DD} \geq V_{PI}$ the plate will suddenly move from $g_0/3$ to g_d turning ON the switch. The presence of this instability it is also detectable while turning OFF the switch. In fact, the switch does not simply turn OFF when the applied voltage between the two plates is smaller than the pull-in voltage (*i.e.* $V_{DD} < V_{PI}$). Instead, it turns OFF when V_{DD} reaches a certain voltage value called release voltage V_{RE} (red circle in Fig. 54). Additionally, upon contact the adhesion forces F_{adh} starts to appear and to counteract F_k . Therefore, F_{adh} must be included in Eq. 19 while computing the forces balance at the contact position, resulting in Eq. 21. At contact, F_k needs to overcome F_{adh} and F_{ele} that at this point ($y = g_d$), it is much larger than the one corresponding to the critical displacement value ($y = g_0/3$). The presences of F_{adh} further decreases the value of V_{RE} .

All the consideration above allows one for computing V_{RE} that is expressed in Eq. 22

$$kg_d = \frac{\epsilon_0 A V_{RE}^2}{2(g_0 - g_d)^2} + F_{adh} \quad \text{Eq. 21}$$

$$V_{RE} = \sqrt{\frac{2(g_0 - g_d)^2 (kg_d - F_{adh})}{\epsilon_0 A}} \quad \text{Eq. 22}$$

The hysteresis voltage margin V_{HY} is equal to the difference between the pull in voltage and the release voltage $V_{HY}=V_{PI}-V_{RE}$. This hysteresis margin is what ultimately limits the lowest possible operating voltage, especially for normally OFF logic relay. Non-Volatile applications require that the electrodes of the relay stay into contact when no voltage is applied. This implies that the spring restoring force of the relay has to be smaller than F_{adh} . Applying this condition, Eq. 22 results in the nonexistence of a voltage release V_{RE} .

The parallel plate capacitor model is an approximation suitable for switch designs leading to parallel movement and a uniform gap upon actuation. While it provides an easy qualitative understanding of the dimensional trends influencing V_{PI} , this model cannot be accurate for predicting the behavior of a cantilever-beam and clamped-clamped switch designs. In this thesis, two main switch designs have been analyzed and fabricated: Cantilever beam design and clamped-clamped beam design, both with in plane actuations. Therefore, a complete behavioral model of cantilever-beam nanoelectromechanical relay needs to be derived.

3.1.2 MODELING OF MECHANICAL BEAM

To understand the statics of the bending, it is possible to take as example a beam which has a free end (the tip) and a fixed end (the anchor) *i.e.* cantilever beam. In which a constant force P is applied at the tip in the $+y$ direction as shown in Fig. 55.

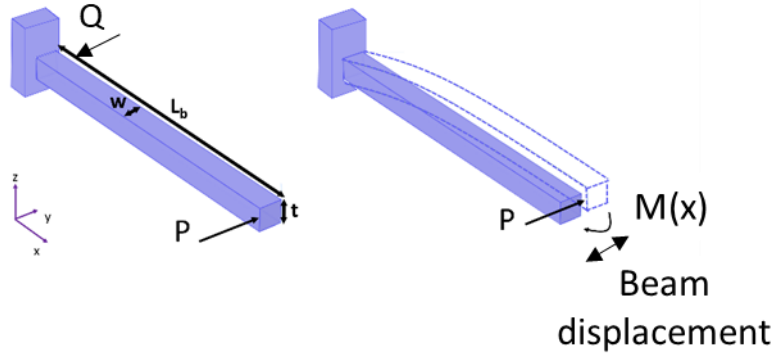


Fig. 55: Left) Elastic cantilever beam with Load P applied on the free end; Right) Beam deformed under compressive load P showing also bending moment $M(x)$ of the deformed beam caused by compressive load P .

If a constant load P is applied at the tip, a shear opposite force Q appears on the anchor, together with a moment $M(x)$ to maintain equilibrium. The static equilibrium can be express as Eq. 23 for the forces and Eq. 24 for the moments.

$$\sum P_y = 0 = Q + P \Rightarrow Q = -P \quad \text{Eq. 23}$$

$$\sum M_x = 0 = -M(x) + P\delta(x) \Rightarrow M(x) = P\delta(x) \quad \text{Eq. 24}$$

Where δ is a small displacement transverse to the y -axis. If the applied load is a distributed load rather than a point load, the shear force Q can be calculated as Eq. 25 by integrating the points load $p(x)$ applied to small sections of the beams.

$$Q = - \int_{x_0}^{x_1} p(x) dx \quad \text{Eq. 25}$$

Where x_0 is the position x in which the applied load $p(x)$ begins and x_1 where it ends, with dx the dummy length. Consequently, the bending moment is given by Eq. 26.

$$M = \int_{x_0}^{x_1} p(x) x dx \quad \text{Eq. 26}$$

3.1.2.1 STIFFNESS

The stiffness is defined by the quantity of load needed to induce a deformation ∂ in a material, and it is expressed in Eq. 27, which is also called Hooke's law or spring restoring force and it is valid for small displacement.

$$F_k = k\partial \Rightarrow k = \frac{F_k}{\partial} \quad \text{Eq. 27}$$

k is the proportionality constant between the force and the beam deformation, hence is measured in [N/m]. The stiffness depends on the geometry of the device. For simplicity, it is possible to assume that the load applied stretches all parts of the beam uniformly and this stretching can be calculated as Eq. 28:

$$\epsilon = \frac{\partial}{L_0} \quad \text{Eq. 28}$$

Where ϵ a dimensionless parameter that measures the stretching and is called tensile strain. L_0 is the original beam length. It can be interesting to adjust the stiffness linked to the material proprieties by normalizing the load to the cross-sectional area in which is applied. Considering a load per unit area and an elongation per unit length Eq. 27 can be written as Eq. 29 or as Eq. 30.

$$\frac{F_k}{A_0} = E \frac{\partial}{L_0} \quad \text{Eq. 29}$$

$$\sigma = E\epsilon \quad \text{Eq. 30}$$

E is the proportionality constant between the stress σ and the strain ϵ and it is called Young's modulus or modulus of elasticity. It has the same unity as stress [Pa].

Eq. 29 highlights how the stiffness of a solid bar is related to the Young's modulus and can be written explicitly in Eq. 31:

$$k = \frac{AE}{L} \quad \text{Eq. 31}$$

From Eq. 31 and Hooke's law the elongation can be written as Eq. 32, which highlights the dependency on the material proprieties and the applied load.

$$\partial = \frac{PL}{AE} \quad \text{Eq. 32}$$

The beam elongation is proportional to the applied tensile force and to the beam length, and it is inversely proportional to the cross-sectional area and the modulus of elasticity.

3.1.2.2 CANTILEVER AND CLAMPED-CLAMPED BEAM STIFNESS AND RESONANT FREQUENCY

It is also important to derive the beam deflection produced by the applied load to understand the pull-in equations for a cantilever beam relay. To derive these equations, the example of a cantilever beam on which a constant force P is applied at the tip towards $+y$ direction as been taken as shown in Fig. 55

To find the y displacement, it will be necessary to solve the deflection equation (Eq. 33) given by [TIMOSHENKO 1940]:

$$EI \frac{d^2 v(x)}{dx^2} = -M(x) \quad \text{Eq. 33}$$

Where $M(x)$ is the bending moment at position x and I is the moment of inertia, which for a rectangular cross section such as a beam, with centered rotational axes, it is given by Eq. 34:

$$I = \frac{w^3 t}{12} \quad \text{Eq. 34}$$

Considering that a constant force P is applied at the tip (the free end position), $x=L_b$, the bending moment is given by Eq. 35:

$$M(x) = -P(L_b - x) \quad \text{Eq. 35}$$

The boundary condition imposes that the displacement and the beam velocity at the position $x=0$ must be equal to zero:

$$\frac{dv(x)}{dx} \Big|_{x=0} = 0 \quad \text{Eq. 36}$$

By applying the boundary condition, while integrating (Eq. 33) we obtain the displacement for a cantilever beam:

$$y(x) = \frac{P}{6EI} (3L_b x^2 - x^3) \quad \text{Eq. 37}$$

Therefore, the maximum displacement at the tip (*i.e.* at $x=L_b$) is:

$$y(L_b) = \frac{PL_b^3}{3EI} \quad \text{Eq. 38}$$

From here, we can combine it to the Hooke's law to obtain the expression of the effective spring constant k :

$$k_{eff} = \frac{3EI}{L_b^3} = \frac{Et w^3}{4L_b^3} \quad \text{Eq. 39}$$

The cantilever beam fundamental frequency is given by Eq. 44:

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} = \frac{1}{2\pi} \sqrt{\frac{3EI}{mL_b^3}} = \frac{1}{2\pi} \sqrt{\frac{E}{4\rho} \frac{w}{L_b^2}} \quad \text{Eq. 40}$$

ρ being the beam density and m the beam mass.

For example, a Silicon cantilever beam has a value of $f_0 = 203\text{kHz}$, for $E=168\text{GPa}$, $w=50\text{nm}$, $\rho = 2330\text{kg/m}^3$, $L_b=12.9\mu\text{m}$.

For the Clamped-Clamped beam, having symmetrical actuation a constant load is applied to the beam center, and both ends are anchored. Therefore, the velocity and displacement of the beam are both zero for $x=0$ and $x=L_b$. Consequently, the bending moment and the $y(x)$ displacement can be written as Eq. 41 and Eq. 42, respectively:

$$\left. \begin{aligned} M(x) &= -\frac{PL_b}{8} + \frac{Px}{2} & \text{if } 0 < x < L_b/2 \\ M(x) &= \frac{PL_b}{8} - \frac{Px}{2} & \text{if } L_b/2 < x < L_b \end{aligned} \right\} \quad \text{Eq. 41}$$

$$y(x) = \frac{Px^2}{EI} \left(\frac{x}{12} - \frac{L_b}{16} \right) \quad \text{if } 0 < x < \frac{L_b}{2} \quad \text{Eq. 42}$$

At the maximum displacement (*i.e.* $x=L_b/2$), we can extract the effective spring constant which is given by Eq. 39:

$$k_{eff} = \frac{192EI}{L_b^3} = \frac{16Et w^3}{L_b^3} \quad \text{Eq. 43}$$

The Clamped-Clamped fundamental frequency is given by Eq. 44:

$$f_0 = 1.03 \sqrt{\frac{E}{\rho}} \frac{w}{L_b^2} \quad \text{Eq. 44}$$

For example, a Silicon clamped-clamped beam has a value of $f_0=2.81\text{MHz}$, for $E=168\text{GPa}$, $w=50\text{nm}$, $\rho =2330\text{Kg/m}^3$, $L_b=12.9\mu\text{m}$.

The fundamental frequency of the relay influences the mechanical delay.

3.1.3 MECHANICAL DELAY

As soon as a voltage bias is applied between the movable electrode and the fixed electrode, and produce an electrostatic force, the movable electrode starts moving and its movement is governed by Newton's second law of motion Eq. 45:

$$m\ddot{y} + \gamma\dot{y} + k_{eff}y = F_{ele} \quad \text{Eq. 45}$$

Where, m is the effective mass, y is the displacement, k is the spring constant, F_{ele} is the applied electrostatic force and γ is the damping coefficient that is defined in Eq. 46

$$\gamma = \frac{k_{eff}}{\omega_0 Q} = \frac{\sqrt{mk_{eff}}}{Q} \quad \text{Eq. 46}$$

where Q is the quality factor which indicates the total system energy and the average energy loss in one radian at resonant frequency [GAMARRA 2014]. For underdamped system the damping factor can be approximated by

$$\gamma \sim \frac{1}{2Q} \quad \text{Eq. 47}$$

A higher quality factor indicates a lower rate of energy loss; hence the resonator oscillations dies out more slowly. For a NEM relay, it may be preferable to avoid high- Q to minimize non-ideal switching effects such as contact bounce and long settling time. At the same time the relay does not have to actuate too slow down, therefore Q does not have to be too small. The optimal operation regime is $0.5 < Q < 2$ in order to have short settling time after pull-out, and optimal tip bouncing settling time after pull-in [BORREGUERO 2017]. For relays having micrometer-scale actuation gaps, Q is limited by squeeze-film damping. In contrast, for scaled logic relays with actuation gaps $\approx 10\text{nm}$, (*i.e.*: less than the mean free path of an air molecule at atmospheric pressure); squeeze-film damping is negligible. Therefore in nanorelays the quality factor is dominated by surface-related energy-loss mechanisms [KAM 2011].

The relay closing time can be calculated as Eq. 48 adapted from [RANA 2014]. Note that non-ideal effects such as contact bouncing and settling time are assumed negligible.

$$t_m = \frac{2\pi}{\omega_0} \sqrt{\frac{F_k}{F_e}} = 2\pi \sqrt{\frac{mg_d}{F_e}} \quad \text{Eq. 48}$$

ω_0 is the natural frequency or resonant frequency in radian and $\omega_0 = 2\pi f_0 = \sqrt{\frac{k}{m}}$ and m is the mass of the movable electrode. Eq. 48 shows that, in the presence of a constant electrostatic force, the dominant property, affecting the closing time is the beam mass and the gap. In order to reduce the delay, the beam mass should be minimized.

3.2 WORKING MARGIN OPTIMIZATION

In this thesis, the NEM relays are designed to be auxiliary element for CMOS circuitries, therefore V_{DD} applied to the switches has to be the same as the co-integrated CMOS. This means that the relay pull-in voltage should be low enough to guarantee that each device will consistently pull in for ($V_{DD} > V_{PI}$). At the same time, a too large overdrive may cause the movable electrode to collapse onto the actuating electrode in a “catastrophic pull-in or V_{CPI} ”, thus imposing $V_{DD} < V_{CPI}$. Therefore, this paragraph deals with the pull in voltage modeling of nanoelectromechanical relays. The model takes into account the relay actuation electrode size and position with respect the movable electrode. The goal of this study is to improve and to broad the actuation voltage margin called ($V_{CPI} - V_{PI}$) margin, and to center it on the supply voltage V_{DD} . This study has been published in [USAI 2017A]. The importance of designing relays with a broad operating ($V_{CPI} - V_{PI}$) margin is also important to improve reliability. In fact, among the NEM relay failure modes impacting reliability, mechanical fatigue is shown to affect relay movable structures [MUHLSTEIN 2001] potentially causing a drift in pull-in characteristics versus number of switching cycles. Additionally, a variety of environmental factors may alter the mechanical properties of the system [LEE 2016], which contribute to the dispersion in behavior between nominally identical devices.

The study presented below is done for a three-terminal relay design made in monocrystalline highly doped Si cantilever beam, laterally actuated by a Gate, with a contacting Drain electrode Fig. 56. The model and simulation are carried out for this type of design because we had the possibility to fabricate them after this study. The type of device structure material and actuation orientation is chosen because of

a collaboration between two of Leti's laboratories. In fact, we had the possibility to processes NEM relays and NEM resonators in the same dies and structural layer.

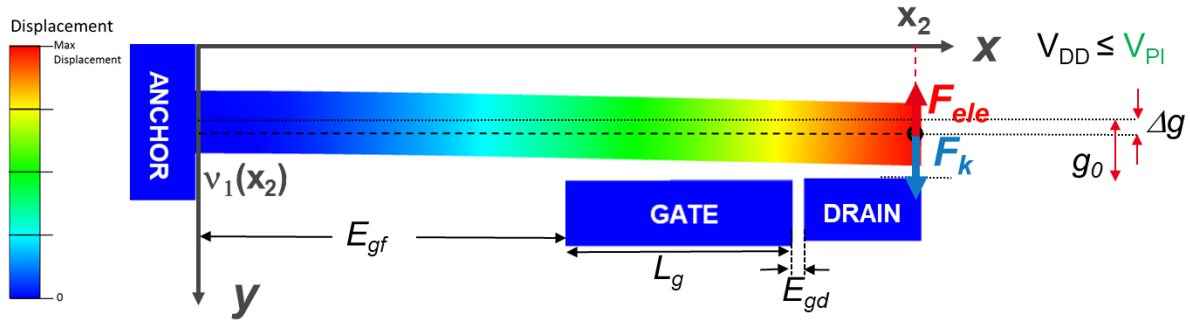


Fig. 56: Schematic top view of electromechanical cantilever beam relay, showing the main design variables, and introducing the forces, coordinates system and beam deflection function used in the next sections.

3.2.1 PULL IN VOLTAGE MODELING FOR “BEAM” NEM RELAY

When a bias V_{DD} is applied between Gate and Beam, an electrostatic force F_{ele} is created, proportional to the actuation area A and following the inverse-square of the separation distance noted g (Eq. 16):

$$F_{ele} = \frac{1}{2} \frac{\epsilon_0 A V^2}{g^2}$$

F_{ele} is competing with the spring restoring force F_k , which itself has a linear dependence on the beam displacement Eq. 17:

$$F_k = k(g_0 - g)$$

in which k is the spring constant and g_0 the separation distance at equilibrium, or nominal gap. Beyond a critical displacement of the beam, the electrostatic force always overwhelms the spring restoring force, causing the beam to “pull-in” and contact the Drain electrode, thus allowing current to flow (see example on parallel plate capacitor in 3.1.1).

In the parallel-plate capacitor approximation, this critical displacement can be shown to be $g_0/3$, but cantilever beam tips typically exceed this value before snapping to contact because in this case, the gap is not uniform. Depending on V_{DD} and on the electrode position, a second type of pull-in may occur, resulting in the collapse of the Beam onto the Gate electrode, and hence in functional failure. This is referred to as catastrophic pull-in voltage, V_{CPI} . Fig. 57 Left shows a SEM top view of a 5-terminals switch having the movable electrode (the beam) contacting the drain. The beam displacement is due to the potential difference V_{DD} between gate and beam, which is in this case equal or larger to the pull in voltage V_{PI} of this

structure. Fig. 57 right shows also a SEM top view of a 5-terminals switch, but in this case, the movable electrode (the beam) is contacting both drain and gate, hence the beam is catastrophically pulling-in. In this case, the applied potential difference V_{DD} between gate and beam is equal or larger than the catastrophic pull in voltage V_{CPI} of this structure.

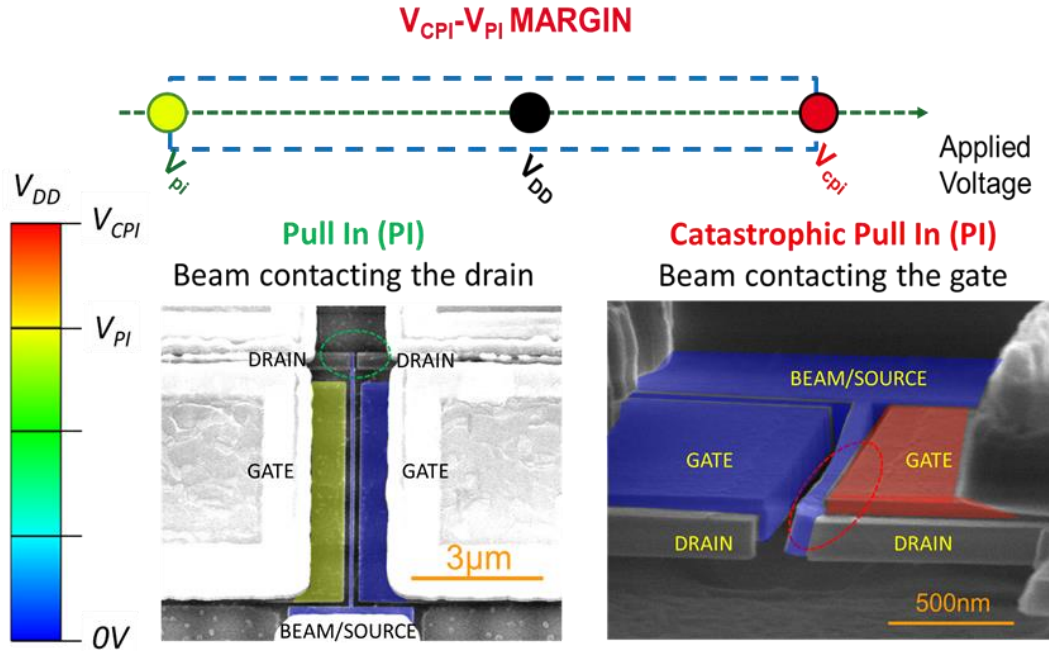


Fig. 57: Example of 5-Terminal NEM relay: left) SEM top view of electromechanical cantilever beam relay, showing beam displacement due to the application of the Pull in Voltage between gate and beam. The beam contacts the drain. Right) SEM view showing beam displacement due to the application of the Catastrophic Pull in Voltage between gate and beam (i.e. The beam contact both drain and gate).

The first step in maximizing the $(V_{CPI} - V_{PI})$ margin is to minimize V_{PI} as a function of the device design parameters: L_b , w , g but also E_{gf} and E_{gd} . [O'BRIEN 2001] proposed a cantilever beam electrostatic pull-in model, which presents the advantage of being simple to use, and free from technology-dependent fitting parameters. This model assumes that the electrostatic force F_{ele} (Eq. 51) that is created by the Gate is applied entirely and only to the beam tip. This results in a lateral beam displacement profile $v(x)$ which can be described as a quadratic function of the position along the x -axis:

$$v(x) = -\frac{M_1 x^2}{2EI} \quad \text{Eq. 49}$$

where M_1 is the moment applied to the beam tip due to F_{ele} , E is the Young's modulus and I is the inertia momentum. Taking into account this deflection profile, the actuation capacitance C_{beam} can be expressed as follows (Eq. 50):

$$C_{beam} = \int_{r_0=0}^{L=L_g=L_b} \frac{\epsilon_0 t}{g_0 - v(x)} dx \quad \text{Eq. 50}$$

since the actuation area in the case of a laterally actuated beam is defined by $L_g * t$. It is important to note that the boundaries of the integral consider that the Gate position and length along the x -axis match exactly that of the length of the movable beam. F_{ele} is then expressed as a function of the energy stored in the C_{beam} capacitor:

$$F_{ele} = \frac{1}{2} \frac{dC_{beam} V_z^2}{dg_{tip}} \quad \text{Eq. 51}$$

Computing the derivative of C_{beam} with respect the beam tip position g_{tip} and equating F_{ele} to F_k results in Eq. 52:

$$V_z = \sqrt{\frac{2k(g_0 - g_{tip})}{-\frac{dC_{beam}}{dg_{tip}}}} \quad \text{Eq. 52}$$

Plotting the solutions of Eq. 52 vs the beam tip displacement yields to a concave curve similar as the one shown in Eq. 21. The absolute maximum of the function indicates both the pull-in voltage and the critical displacement.

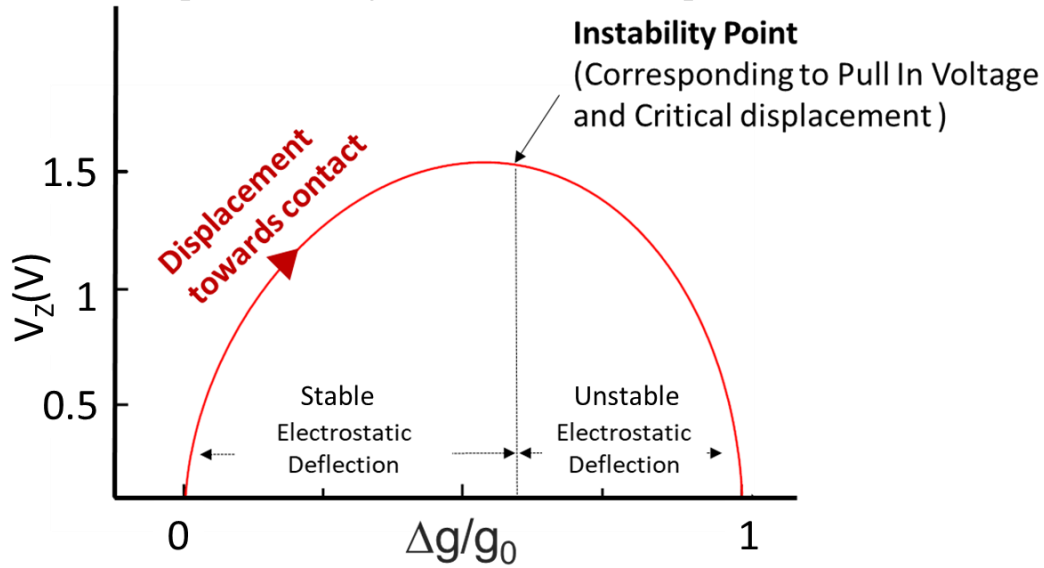


Fig. 58: Example of a plot showing applied actuating voltage versus beam tip displacement g_{tip} over the total gap g_0 for a cantilever beam. V_{PI} is extracted from the instability point at which the derivative changes sign.

Because of the simplification of Eq. 52 while expressing the beam/gate capacitance, a closed-form expression of the pull-in voltage can be obtained (Eq. 53):

$$V_{PI} = \frac{1}{3} \sqrt{\frac{2Ew^3 g_0^3}{\epsilon_0 L^4}} \quad \text{Eq. 53}$$

From this, it can be immediately seen that the minimum achievable actuation gap g_0 and beam width w should be targeted in order to minimize V_{PI} . Long Beams and long Gates are also desirable, though they tend to reduce the switch integration density (for in plane relays). This assumption cannot be used for real 3-Terminal NEM relays. In fact, it would be impossible to have an actuation Gate matching the entire length of the beam ($L = L_g = L_b$), since this would not allow the presence of a third (contact) terminal.

To the [O'BRIEN 2001] model, we added three geometric parameters while computing the pull in voltage [USAI 2017A]: the distance between the anchor and the actuation electrode E_{gf} , the actual actuation electrode length L_g (different from the beam length L_b) and the distance between contact and actuation electrode E_{gd} (Fig. 57).

Taking into account these three parameters in the pull in computation modifies the expression of the beam/gate capacitance. Eq. 50 needs to be rewritten by setting new integral boundary: $x = E_{gf} = \alpha$ and $x = L_g + E_{gf} = \beta$

$$C_{beam} = \int_{\alpha}^{\beta} \frac{\epsilon_0 t}{g_0 - \frac{k(g_0 - g_{tip})L_b x^2}{2EI}} dx \quad \text{Eq. 54}$$

Solving for Eq. 54 yields Eq. 55

$$C_{beam} = \frac{\epsilon_0 t L_b}{\sqrt{g_0 \Delta g}} \left(\tanh^{-1} \left(\frac{\beta \Delta g}{\sqrt{g_0 \Delta g}} \right) - \tanh^{-1} \left(\frac{\alpha \Delta g}{\sqrt{g_0 \Delta g}} \right) \right) \quad \text{Eq. 55}$$

The derivative of the capacitance with respect the beam tip position results in obtaining for V_Z the following expression Eq. 56:

$$V_Z = \sqrt{\frac{2k(g_0 - g_{tip})}{\frac{\epsilon_0 t L_b}{g_0^2} [D_1 - D_2]}} \quad \text{Eq. 56}$$

Where D_1 and D_2 expressed by Eq. 57 and Eq. 58:

$$D_1 = \frac{-\beta}{2L_b(1 - g_{tip}) \left(1 - \frac{\beta^2}{L_b^2} (1 - g_{tip}) \right)} + \frac{\tanh^{-1} \left(\frac{\beta}{L_b} \sqrt{1 - g_{tip}} \right)}{2\sqrt{(1 - g_{tip})^3}} \quad \text{Eq. 57}$$

$$D_2 = \frac{-\beta}{2L_b(1 - g_{tip}) \left(1 - \frac{\alpha^2}{L_b^2} (1 - g_{tip}) \right)} + \frac{\tanh^{-1} \left(\frac{\alpha}{L_b} \sqrt{1 - g_{tip}} \right)}{2\sqrt{(1 - g_{tip})^3}} \quad \left. \vphantom{\frac{-\beta}{2L_b(1 - g_{tip}) \left(1 - \frac{\alpha^2}{L_b^2} (1 - g_{tip}) \right)}} \right\} \text{Eq. 58}$$

Fig. 59 shows in blue solid lines the value of V_{PI} computed through Eq. 56 for various Gate electrode sizes (varying L_g) and positions with respect to the anchor, *i.e.* for various values of α and β and extracted using Mathcad software. As predicted, the V_{PI} deviates from the asymptotic lower limit of Eq. 53, and the deviation is particularly pronounced for short Gates located towards the anchor. Simulating the same geometries using the pull-in analysis feature of FEA software Coventor MEMS+ 6.0 with Timoshenko beam elements has validated our developed model. Similar trends have been reproduced (light blue dashed lines). The results can be qualitatively explained as follow: locating the Gate closer to the free end of the cantilever beam increases the value of C_{beam} , since this is where the beam deflection is the largest, the average electrodes separation the smallest, and thus the electrostatic force the strongest. Increasing the gate length further increases the effective electrostatic force. Therefore, in order to achieve low V_{PI} it is better to prefer long gates located towards the beam tip.

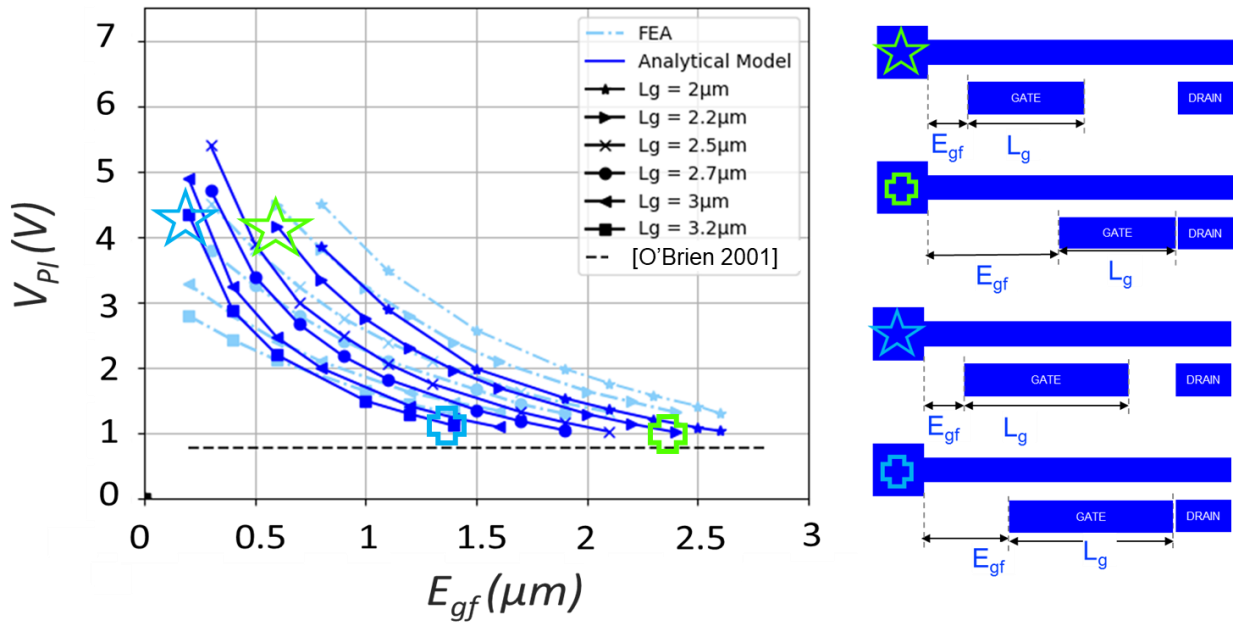


Fig. 59: V_{PI} as a function of E_{gf} for various L_g . L_b is fixed at $5\mu m$. Comparison between the simplified analytical model in [O'Brien 2001] (black dashed line) and the voltage trends (blue solid lines) obtained by extracted V_{PI} values using the improved analytical model. The light blue dashed/dotted lines were obtained on similar geometries, using Coventor MEMS+ FEA.

The quantitative mismatch between FEA and our analytical model can be explained for the most part by the underlying hypothesis of Eq. 49. Considering an electrostatic force concentrated at the beam tip gives indeed a second order polynomial variation for the beam deflection profile $v(x)$. In reality, the electrostatic force should be materialized as a non-uniform load distributed along a section of this beam, which would lead to a 4th order polynomial variation for $v(x)$. In the following section, we rely trends obtained by FEA for studying V_{CPI} .

3.2.2 CATASTROPHIC PULL IN VOLTAGE

The next step to optimize the operation ($V_{CPI}-V_{PI}$) margin around V_{DD} requires evaluating and maximizing V_{CPI} . Qualitatively, V_{CPI} has similar trends versus L_g and E_{gf} with respect V_{PI} but lesser sensitivity in terms of increasing voltage with decreasing E_{gf} .

This lesser sensitivity can be explained by the fact that the configuration “cantilever beam” of the movable electrode, becomes “simply-supported end (pinned) and clamped end” configuration once the movable electrode pulled onto the drain. Additionally, the gap between the beam and the gate becomes non-uniform. This new configuration places the point of maximal deflection more likely to be located towards the center of the beam, resulting in a C_{beam} less sensitive to a shift of the Gate electrode see Fig. 60. The consequence is that globally smaller ($V_{CPI}-V_{PI}$) margins are obtained when the Gate is located towards the beam anchor, as seen on Fig. 61.

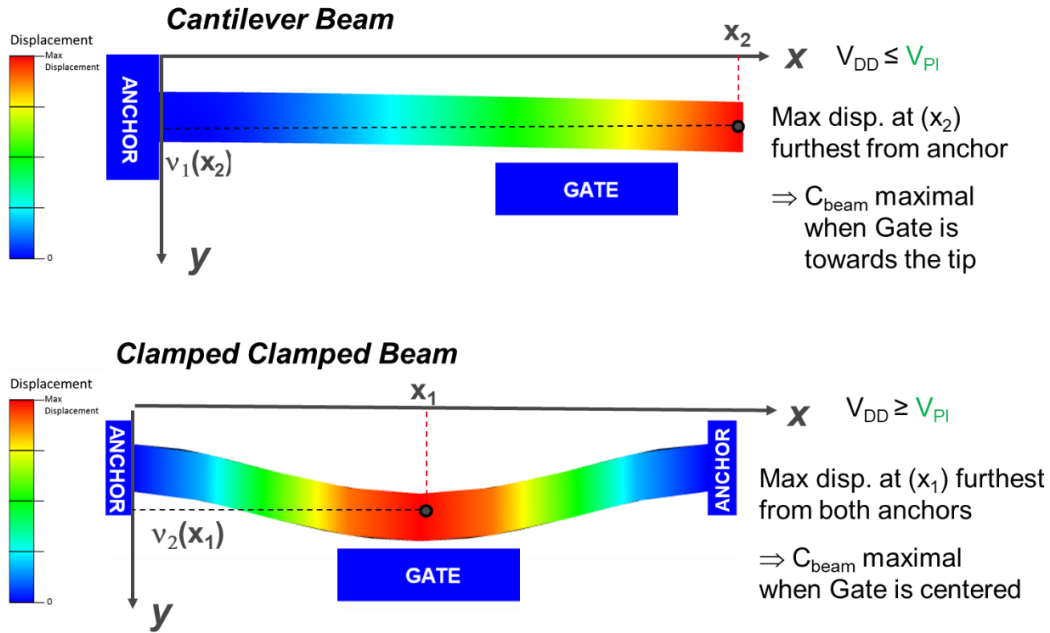


Fig. 60: Top) The figures shows the maximum displacement for Cantilever beam and Clamped-Clamped beam configuration.

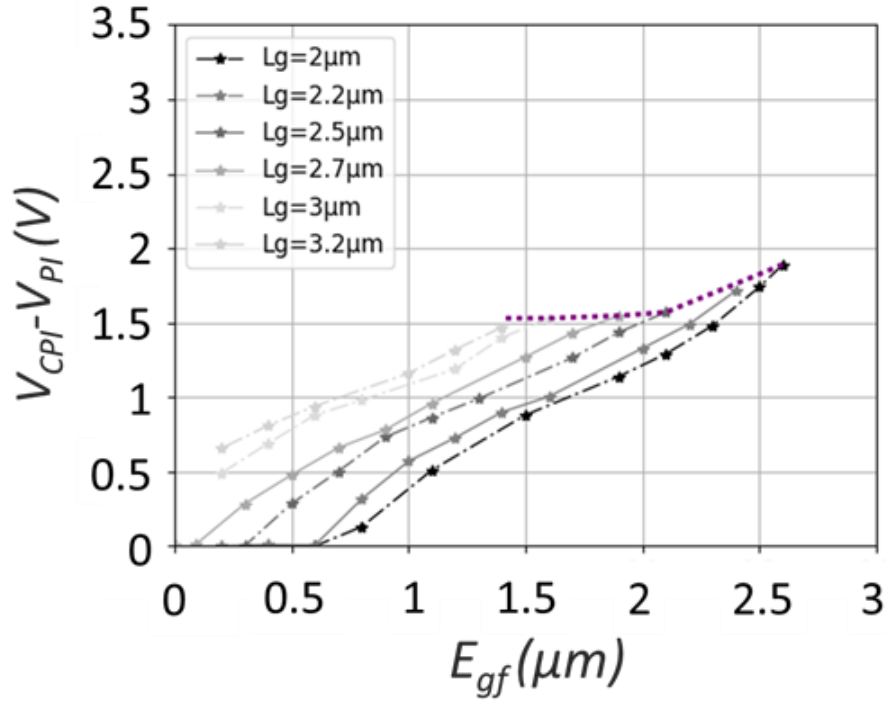


Fig. 61: The figure shows the margin $V_{CPI} - V_{PI}$ calculated as a function of the distance between the gate and the anchor E_{gf} for different gate length L_g . L_b is fixed at $5 \mu m$. The trends for the highest margin (dotted purple line) shows an increment on the margin for shorter L_g .

The maximal margin values are actually obtained for shorter Gates with large E_{gf} . This may seem counterintuitive at first, but it underlines the fact that the portion of the beam having the most impact on C_{beam} , shifts from tip towards its center between V_{PI} and V_{CPI} . Shortening L_g may enable to increase E_{gf} to a point at which the Gate is shifted past this central region, making catastrophic pull-in harder to occur.

In the previous section is shown that shortening L_g at large E_{gf} does barely affected V_{PI} (major contribution to C_{beam} towards the tip). This means that the larger margin is almost only attributed to a V_{CPI} increase, as evidenced by Fig. 62:

It seems that the downside is negligible with respect the benefits if the gate is designed relatively short and towards the tip. Moreover, on the contrary it seems that it can lead to a better operating margin with no penalty on minimum supply voltage, if it is sufficiently shifted towards the cantilever tip to lower V_{PI} , and if it is sufficiently off-center to increase V_{CPI} .

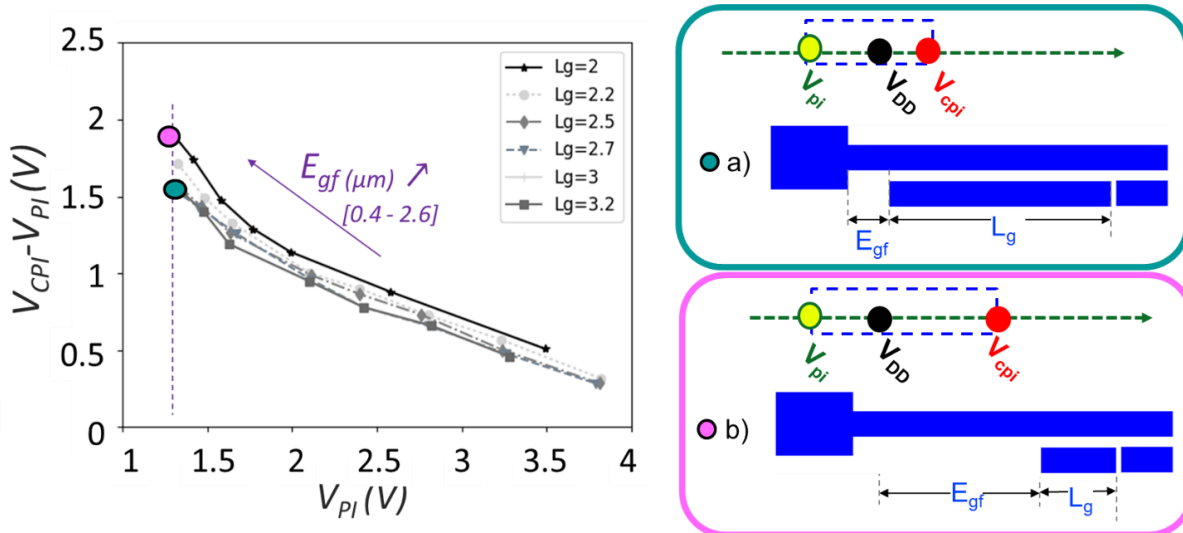


Fig. 62: Shows the working margin $V_{CPI} - V_{PI}$ vs the pull-in voltage. It is interesting to notice how the margin increases with the shorter gate length compared to an almost the same value of V_{PI}

3.2.2.1 WORKING MARGIN OPTIMIZATION SUMMARY

This Analytical and FEA modeling of the pull-in and catastrophic pull-in behavior of a three-terminal electrostatically actuated NEM relay allowed to obtain qualitative insight in optimizing the actuating electrode size and position with respect to the beam, in order to improve the robustness of operation. Reducing the Gate length L_g and shifting the gate towards the tip increases the $(V_{CPI} - V_{PI})$ margin, with little to no impact on the minimum V_{PI} . Leveraging this trend could help to counteract, through clever device design, the dispersion or drift of pull-in characteristics due to cycling, process variability, or environmental factors.

3.3 DESIGN OPTIMIZATION FOR NON-VOLATILE NEM RELAY

This paragraph deals with the design optimization of Non Volatile relay for memory applications, the study has been published in [USAI 2017B]. The suitable relay design is a SPDT design which can be programmed to one side and reprogrammed to the opposite side. The main challenge in designing functional non-volatile relay is the incertitude linked to the unknown and shifting value of the adhesion force. In order to design a functional Non-Volatile relay, it is necessary to optimize the forces operating margin ($F_{ele} + F_k > F_{adh} > F_k$). Additionally, in case of CMOS/NEMS co-integration the relay has to be programmed and re-programmed at supply voltage V_{DD} given by the CMOS circuitry. The design optimization can be adapted to all the type of structural material, but here the computations are made

for relays using single-crystal Silicon as structural and contact material. The electrostatically actuated Single-Pole Double-Throw (SPDT) designs are well suited to memory applications since the storage of a “low” or “high” logical state is performed in a complementary and symmetrical way. Fig. 63 Left a) shows an in-plane actuated 5-terminal (5T) NEM relay, with a movable beam electrode B (source), two actuation electrodes: left gate G_L and right gate G_R , and two corresponding contact electrodes: left drain D_L and right drain D_R . The gap g_g is defined as the separation between the Gates and the beam and the gap g_d as the separation between the Drains and the beam when the beam is in the neutral position, centered on the x-axis. Fig. 63 right) shows the complementary operation of the switch when the actuation electrodes G_R and G_L are respectively biased at GND and V_{DD} . The beam makes contact with D_R (respectively D_L) if it is electrically biased to a ‘ V_{DD} ’=‘High’=‘1’ (respectively to a ‘ GND ’=‘low’=‘0’’) state due to the attractive electrostatic force F_{ele} induced by the voltage difference with G_R (resp. G_L), and it switches to the other side when the bias applied to B is reversed (*i.e.* it is reprogrammable). When the supply voltages are removed from all the electrodes, the beam retains its position (*i.e.* it is non-volatile) due to adhesive force F_{adh} Fig. 63 (Right).

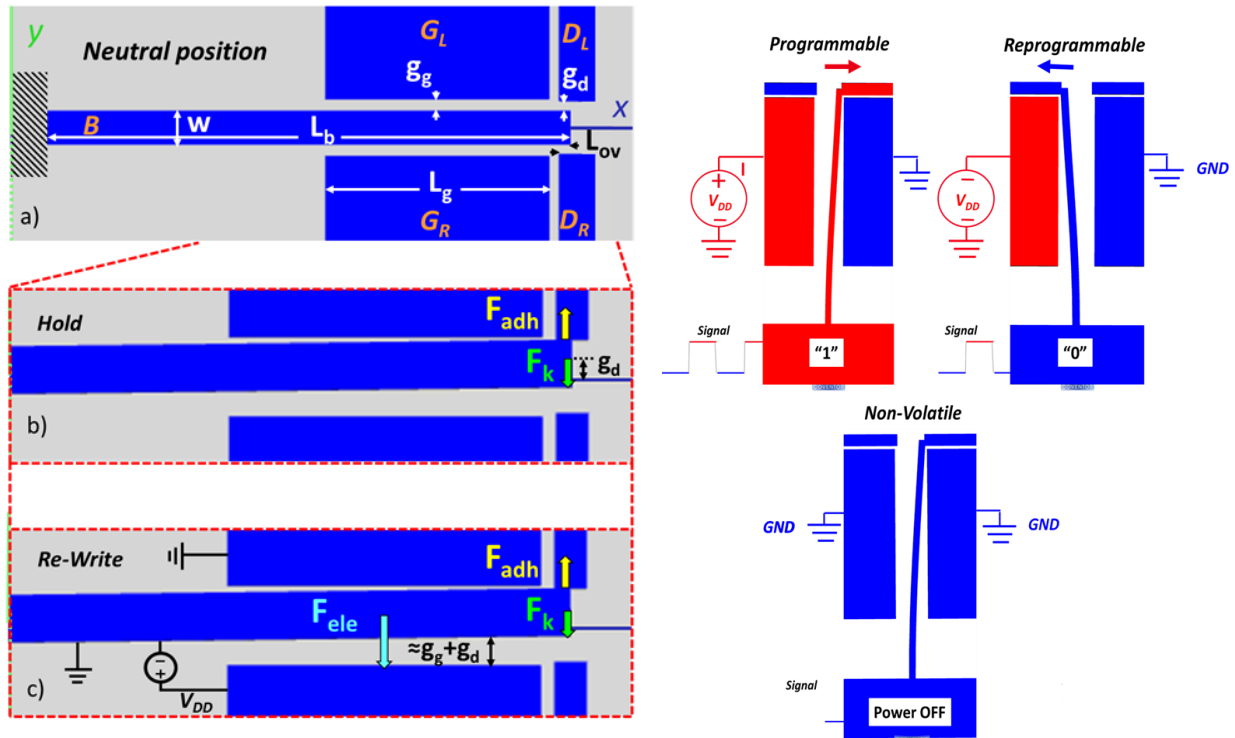


Fig. 63: Left) a) Design parameters for an electrostatic, in-plane-actuated 5T NEM switch. b) Hold state configuration and associated forces. c) Forces involved during a reprogramming operation. Right) Schematic description of the reprogrammable, non-volatile operation of the relay.

The applications requiring this functionality are described in Chapter 2 (*e.g.* Data Searching, Look-up-Table, Reconfigurable Logic, NV-SRAM and NV-TCAM [Kato 2016a; Kato 2016b; Liu 2017]). The functionality relies on careful balancing

of F_{ele} , F_{adh} and the spring restoring force F_k for a given V_{DD} that is fixed by the need for co-integration with CMOS circuitry. Fig. 63 b) shows the forces acting in the “Hold” state for an in-plane-actuated cantilever beam. In the hold state the power is switched off, this means that F_{adh} has to be larger than the opposing F_k . Eq. 59 expresses the adhesion force:

$$F_{adh} = t * L_{ov} * P_{adh} \quad Eq. 59$$

where t is the beam thickness, P_{adh} is the adhesion pressure and L_{ov} is the nominal contact length. In the multiple nano-asperities contact regime, P_{adh} depends not only on the contact material but also on the surface roughness [PROKOPOVICH 2011]. Typical values at the sub-micron scale are found in the range of 0.1-1 MPa [LI 2010A] for unoxidized single-crystal Silicon surfaces. From now on, in the following discussion all the variables used are layout-dependent variables, and the electrodes thickness t is considered fixed by the technology. This study has been carried out with the objective of fabricating the relays, hence all the parameters are optimized for LETI’s technology. Despite this, the study can be applied to any in-plane 5T NEM relay structure.

F_k is expressed as Eq. 17, where for an in-plane-actuated beam the stiffness can be approximated by Eq. 39: ($k=[(E*t*w^3)/(4*L_b^3)]$) (load concentrated at the tip) where E is the beam material Young’s modulus. Fig. 63 c) shows the forces acting during a “Re-write” operation: $F_{ele} + F_k$ must exceed F_{adh} to allow the beam to switch onto the other side. For simplicity F_{ele} is expressed using the parallel-plate capacitor approximation Eq. 16, and with a parallel actuation gap approximated as $g_g + g_d$.

In Fig. 64 the system forces are plotted vs. the beam length L_b in order to visualize the reprogram (blue hatching) and non-volatile (yellow hatching) domains. The crosshatched overlap between those two domains represents the working design window for which a non-volatile and reprogrammable relay is obtained (*i.e.* $F_k < F_{adh} < F_{ele} + F_k$).

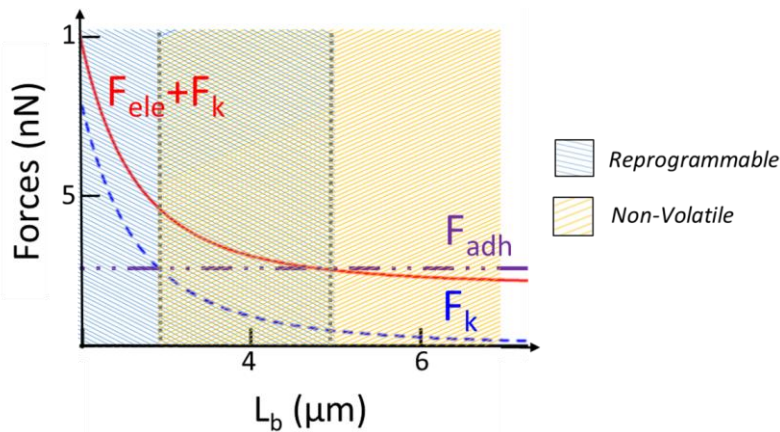


Fig. 64: Design window representation by the overlap between reprogrammable and nonvolatile domains. Parameters: $g_g = 80\text{nm}$, $w = 80\text{nm}$, $L_g = 8\mu\text{m}$, $g_d = 50\text{nm}$, $t = 160\text{nm}$, $V_{DD} = 1.8\text{V}$.

Robust switch designs should have a wide design window, ideally with $F_{ele}+F_k$ and F_k centered on the nominal F_{adh} value. In Fig. 65 we plotted the system forces and the p parameter vs various relay parameters.

The p parameter is the voltage overdrive factor $p=V_{DD}/V_{PI}$ and can be expressed by Eq. 60:

$$p \cong \frac{V_{DD}}{\sqrt{\frac{2}{27} \frac{Et w^3}{L_b^3} \frac{(g_g + g_d)^3}{\epsilon_0 L_g t}}} \quad \text{Eq. 60}$$

Eq. 16 and Eq. 17 clearly indicate that the window (*i.e.* F_{ele}), depends linearly on L_g whereas F_k does not and in Fig. 65 top left it is illustrated the window broadening associated with increasing L_g . The equations also show that F_k has a cubic dependence on the beam width w , which does not affect F_{ele} . Hence, w can be used to center the window on F_{adh} as shown on Fig. 65 (top middle). Another efficient way to increase the window is to reduce the actuation gap size g_g (Fig. 65 top right) down to the minimum distance g_d , since F_{ele} is proportional to $1/(g_g+g_d)^2$ (Eq. 17).

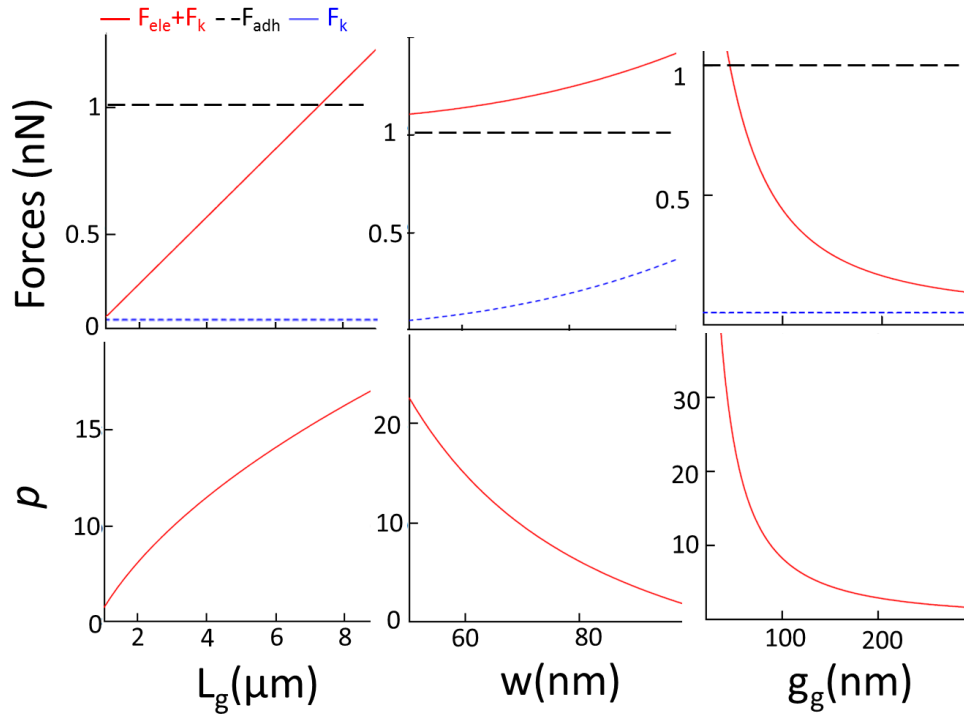


Fig. 65: System forces and voltage overdrive dependences on switch design parameters L_g , w , and g_g . Having $L_b = 9\mu\text{m}$, $g_d = 50\text{nm}$, $t = 160\text{nm}$, $V_{DD} = 1.8\text{V}$.

These considerations need to be accompanied by taking into account for parameter variations affecting the beam's “intrinsic” pull-in voltage V_{PI} which needs to be

kept below V_{DD} . That is why the parameter voltage overdrive p has been defined Eq. 60, and plotted against L_g , w , and g_g in Fig. 65 (bottom).

The p factor should be larger than unity to ensure the actuation of the beam at V_{DD} . However, the risk associated with an unduly large p factor is catastrophic pull-in (CPI), which has been defined on the paragraph above (V_{CPI} occurs when the movable beam collapses onto the actuation Gate). Therefore, decreasing w or g_g comes at the cost of increasing the risk of CPI. To complete the study of the NV NEM relay design optimization, the operating voltage margin $V_{CPI}-V_{PI}$ needs to be included. Fig. 66 shows the both adhesion (lines) and $V_{CPI}-V_{PI}$ (symbols) stability margins plotted against p , which is itself mapped to the corresponding g_g/g_d ratio with all parameters but g_g fixed. While the design margin increases as g_g decreases towards the minimum feature size, the CPI margin is reduced. Beyond a certain point ($p > 3.7$ in this case), the margin drops to zero, meaning that relay actuation without CPI becomes impossible.

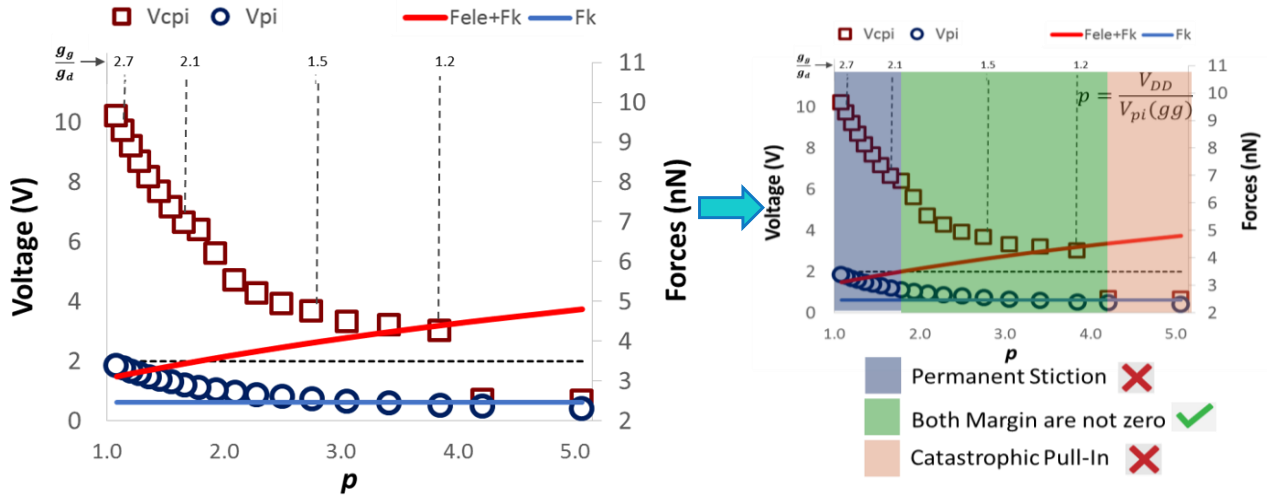


Fig. 66: Evolution of the adhesion (solid lines) and catastrophic pull-in (symbols) margins vs. the overdrive factor p corresponding to a variation of g_g . All other design parameters were fixed $L_b = 9\mu m$, $g_d = 50nm$, $t = 160nm$, $V_{DD} = 1.8V$. Gap ratio g_g/g_d values are indicated for reference. On the right a color code explanation is given to better understand the implication of considering both adhesion and CPI margin while design a 5T NV NEM Relay

Tracking p to ensure that it does not rise above a certain value (here 3.7) when adjusting w , L_g and g_g is necessary to avoid this issue. There is a trade-off between robustness against catastrophic pull-in, and robustness against F_{adh} variations.

Fig. 67 shows a simulated chronogram for a functional relay designs for the purpose of non-volatile and reprogrammable operation at fixed V_{DD} . This design uses the above described guidelines, *i.e.* increasing L_g to enlarge the design window, tuning w to center it on F_{adh} and checking p to keep a sufficient CPI margin. The simulation was carried out using Coventor MEMS+.

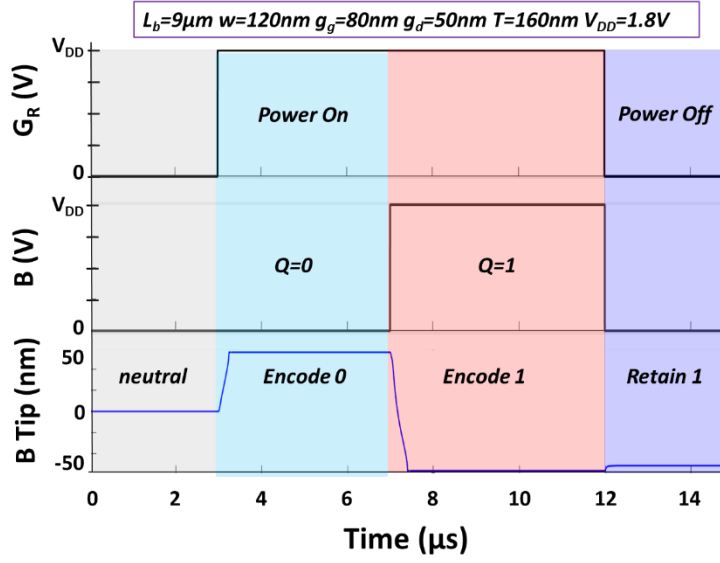


Fig. 67: Simulated transient analysis of 5T laterally actuated NEM relay. Used parameters:
 $L_b = 9\mu\text{m}$; $g_g = 80\text{nm}$; $g_d = 50\text{nm}$; $w = 0.12\mu\text{m}$; $T = 0.16\mu\text{m}$; $V_{DD} = 1.8\text{V}$; $L_g = 7.5\mu\text{m}$.

As a conclusion, the design of a non-volatile and reprogrammable electrostatic M/NEMS relay robust against process-induced variations and drift of F_{adh} and V_{PI} over the operating lifetime requires maximizing the design window defined by $F_{\text{ele}} + F_k > F_{\text{adh}} > F_k$. Broadening the window and centering it on F_{adh} can be carried out independently. However, the parameters enabling a broader this window needs to be the same that keep the overdrive factor p sufficiently low, as it is correlated to the risk of catastrophic pull-in. A trade-off between adhesion and $V_{CPI} - V_{PI}$ margin needs to be considered.

3.4 CONTACT INSTABILITY DUE TO SCALABILITY

In this paragraph we deal with the effects of volatile relay scaling on the contact resistance. Let us take as an example highly p-doped silicon three-terminal nanorelays having an actuation electrode (gate) of length L_g , one contact electrode (drains) of length L_{ov} and a movable cantilever beam electrode of length L_b , width w_b and gap g_g such as the one in Fig. 68. The relay is of the out-of-plane actuation type. The doping level is assumed to be $N_A=6\times 10^{19}\text{cm}^{-3}$, which corresponds to a resistivity $\rho=5\times 10^{-4}\Omega\cdot\text{m}$. The nominal contact area A_n is given by the contact length L_{ov} and the electrode beam width w_b because in this example relay is vertically actuated. (L_{ov} multiplied by thickness t would correspond the contact area, if the relay was laterally actuated). In μ -dimensional contacts (*i.e.* L_{ov} and/or w_b have $\sim \mu\text{m}$ length), the effective contact area A_{eff} is defined by multiple asperities. Therefore, A_{eff} is smaller than A_n and is given by the number and area (assumed to be circular) of the contacting asperities.

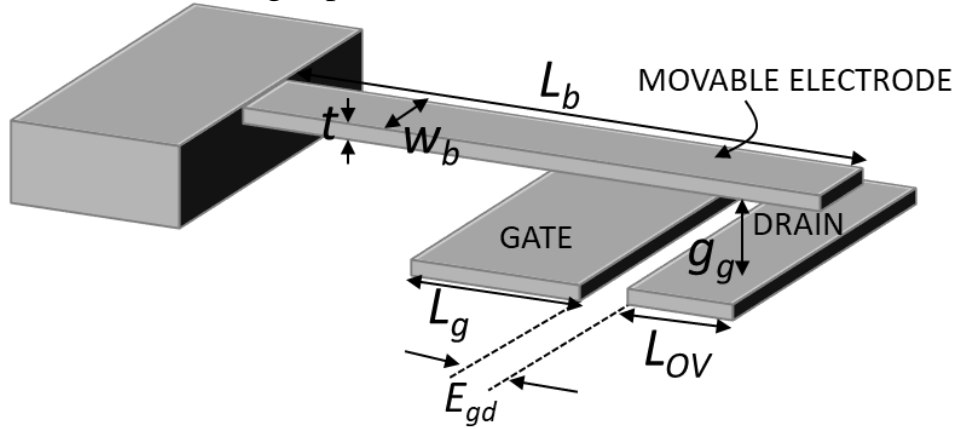


Fig. 68: Sketch of a vertically actuated MEM switch highlighting designed variables.

In [PAWASHE 2013], the average asperity radius is assumed to be $a=5\text{nm}$ and the average separation between two consecutive contacting asperities $d=300\text{nm}$, consistent with a $\sim 10^{-5}$ A_{eff}/A_n ratio suggested by AFM adhesion force measurements [LEE 2010B]. The parameters used in this analysis are reported in the table below.

t	L_b	L_g	w_b	L_{ov}	E_{gd}	V_{PI}
5nm	400nm	280nm	5nm	10nm	10nm	$<0.2\text{V}$

This type of vertically actuated structure is chosen because allows a more aggressive scalability compared to the laterally actuated relay. Indeed, the gap is here defined by the thickness of the sacrificial layer rather than lithography.

We considered L_{ov} being smaller than $2a+d$, in order to remain in the single asperity regime Fig. 69.

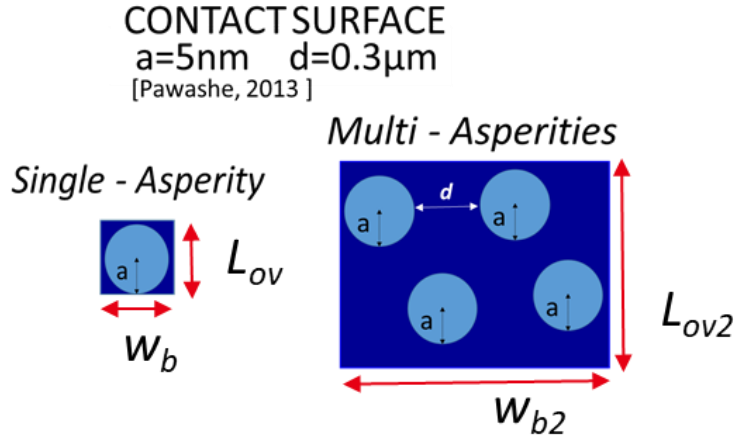


Fig. 69: Illustration of multiple vs. single contacting asperity regimes, with radius and inter-contacting asperity distance values chosen similar to [Pawashe 2013].

This minimal operating voltage is estimated using the analytical model in [USAI 2017A], which takes into account gate size and position with respect to the beam (i.e. L_g and E_{gd}). To register a switching event and characterize the transfer properties of the device, a DC bias is applied to the contact electrode (drain). From the perspective of cascadability is natural to consider that the same amount applied in the gate V_{PI} should also be applied to the drain. After the pull-in event, this potential difference existing between drain and beam is distributed along the movable beam and the contact, in a way that depends on the ratio between the beam resistance R_b and contact R_c (see figure Fig. 70). The drain electrode resistance and the interconnection resistance are neglected for the sake of clarity. Although they do have a layout-dependent impact, the focus is narrowed to the intrinsic device geometry.

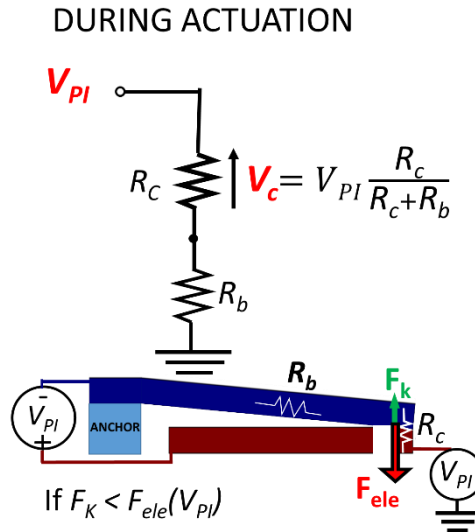


Fig. 70: Schematic of the programmed relay in the test setup showing the system resistances.

The dependency of R_b on the design parameters is shown in Fig. 71 left. The contact area is enlarged in the figure and the contact resistance is highlighted. The contact resistance of a single asperity is estimated by the Sharvin resistance $R_c = R_{\text{Sharvin}}$ [CHEVALLIER 2010]. The Sharvin resistance can be used when the contact radius a , is large compared to the Fermi wavelength $\lambda_F = \rho\lambda$, where ρ is resistivity of the contacting electrode, and λ is the mean electron free path. Therefore, the Sharvin resistance is dominant for small contact sizes ($<10\text{nm}$).

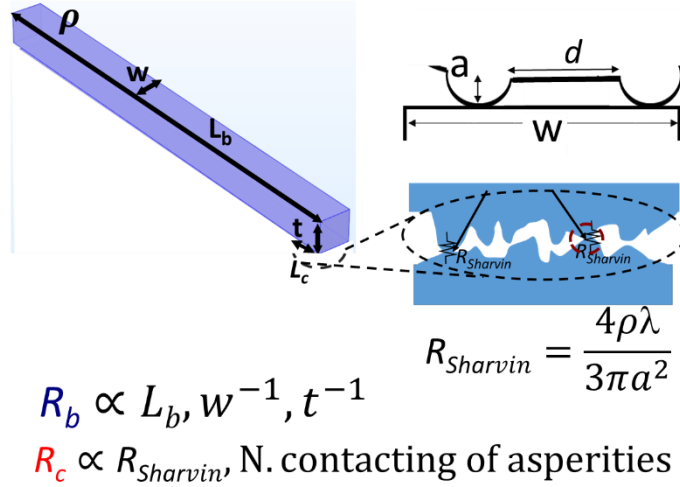


Fig. 71: Difference between the nominal contact and the actual contact surface showing the resistance dependency on the material and geometry parameters.

The contact resistance of multiple asperity is proportional to the Sharvin resistance and the inverse of the number of asperities. Plotting R_b and R_c vs the beam width allows to observe how R_c and R_b scale with the design parameters in the asperities-dominated regime Fig. 72. R_b increases continuously as w decreases, whereas R_c saturates in some intervals due to the discrete nature of contacting asperities. In the single asperity regime (i.e. $2a \leq w \leq 2a+d$), R_c can become significantly smaller than R_b , which scales with w^{-1} .

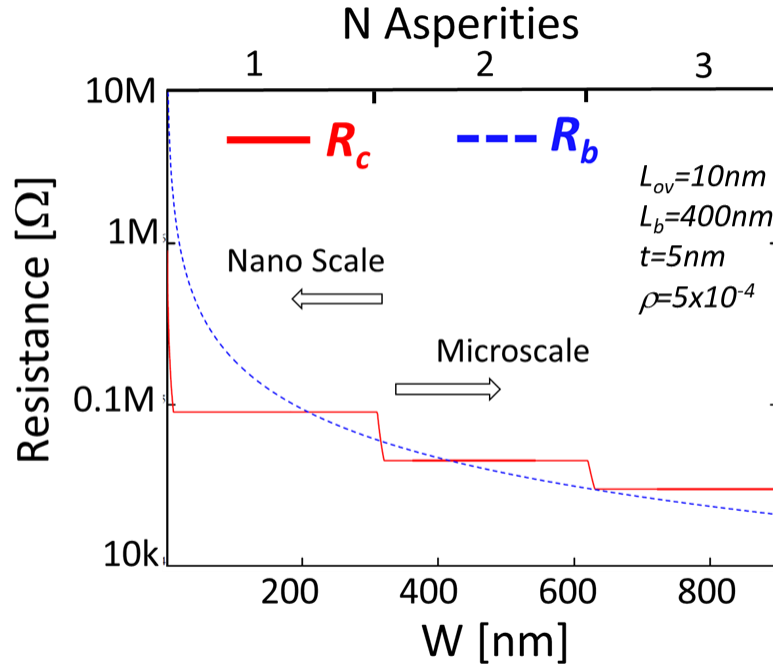


Fig. 72: R_c and R_b evolution with beam width w . R_c shows plateaus due to the discrete nature of contacting asperities.

Fig. 73 shows the voltage distribution along the beam in both case of $R_c \gg R_b$ (Left) and $R_c < R_b$ (right). In the first case, the electrostatic force F_{ele_1} between the actuating electrodes can be written as Eq. 61, F_{ele_1} is obtained from considering the voltage difference between the beam and gate electrodes as uniformly equal to V_{PI} (estimated by the analytical model given in [USAI 2017A]) and integrating the non-uniform gap $g_g(x)$ along the gate length (L_g) direction.

In the $R_c \gg R_b$ case, the voltage difference between the beam and gate is considered to be V_{PI} all along the total beam length. In fact, as soon as the contact between beam and drain is made, V_{PI} drops entirely across the contact, therefore the electrostatic force at pull-in increases by a larger amount of the spring restoring force.

$$F_{ele_1} = \frac{\epsilon_0 w V_{PI}^2}{2} \int_{E_{gf}}^{L_g - \delta} \frac{1}{g_g^2(x)} dx \quad R_c \gg R_b \quad Eq. 61$$

In the second case (right) where, $R_c < R_b$, as soon as the contact is made, a significant fraction of V_{PI} redistributes linearly within the movable electrode. Therefore, F_{ele_2} is obtained by integrating a varying voltage difference ($V_{PI} - V_b(x)$) in the previous integral along a non-uniform gap (see Eq. 62).

$$F_{ele_2} = \frac{\epsilon_0 w}{2} \int_{E_{gf}}^{L_g - \delta} \frac{(V_{PI} - V_b(x))^2}{g_g^2} dx \quad R_c < R_b \quad Eq. 62$$

In this study the adhesion force ($F_{adh} < 0.3 \text{ pN}$ for two contacting Silicon surfaces, $F_{adh} < 0.065 \text{ pN}$ for Si-SiO₂ contacting surfaces) are estimated as negligible considering, the surface energy and the Hamaker constant H for Si, A_{eff} the effective contact area, g_d the separation and D_0 is the effective cutoff distance in the approximation given by [PAWASHE 2013].

$$F_{adh} = \frac{HA_{eff}}{6\pi(g_d - D_0)^3} \quad \text{Eq. 63}$$

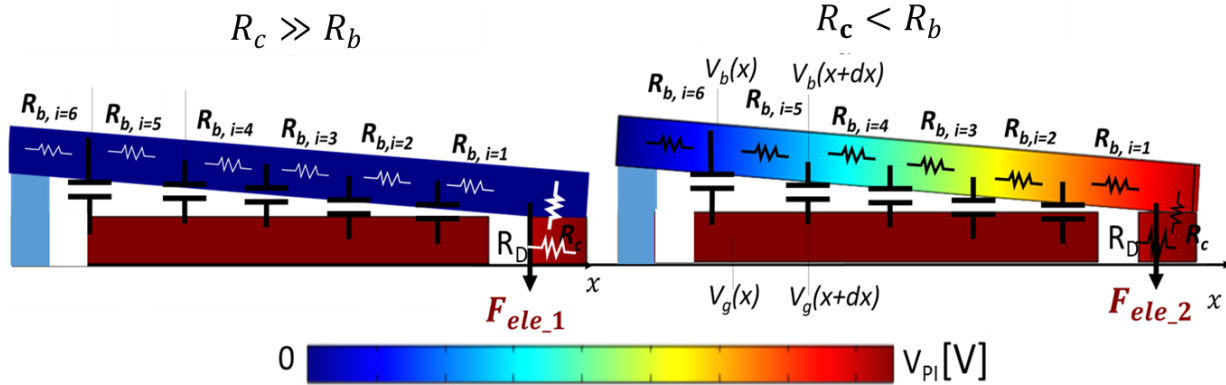


Fig. 73: Schematic representation of the electrostatic potential distribution after contact. When $R_c \gg R_b$ (left), the voltage difference between drain and beam is mostly dropped at the contact. If $R_c < R_b$ (right), V_{PI} drops mainly within the movable electrode along which the potential varies linearly.

This potential redistribution along the beam is of particular importance in the single asperity regime. Fig. 74 left, shows the system forces F_{ele_2} and F_k plotted vs L_g and E_{gf} ($E_{gf} = L_g - E_{gd}$). For some parameters, F_{ele_2} can become smaller than F_k , which would cause the beam to move in and out of contact, thus introducing measurement artifacts. This singularity depends on the gate size and its position with respect to the beam, both of which change with the design parameter $E_{gf} = L_g - E_{gd}$. Short gates located towards the tip of the cantilever, identified in [USAI 2017A] as more robust against catastrophic pull-in, may here be problematic in some cases.

Counter-intuitively, increasing L_g by extending it towards the anchor will simultaneously decrease V_{PI} and increase the electrostatic force F_{ele_2} by rendering the effect of voltage redistribution less substantial. F_{ele_1} is shown for reference, indicating that including the effect of potential redistribution as $R_c < R_b$ may cause a severe decrease of the electrostatic force. For small values of L_g , F_k is larger than F_{ele_2} , meaning that there, the contact is unstable. However, lengthening L_g by decreasing E_{gf} enables to recover enough F_{ele_2} to overcome F_k , and making the contact stable. V_{PI} decreases since it roughly depends on $L_g^{-1/2}$ through the actuation area term. In the F_{ele_1} calculation, this decrease is compensated at large L_g by the integral boundaries. The $(F_{ele_1} - F_{ele_2})$ difference shrinks with L_g since the redistribution has less impact when the gate extends towards the anchor, thus the trend in Fig. 74 right.

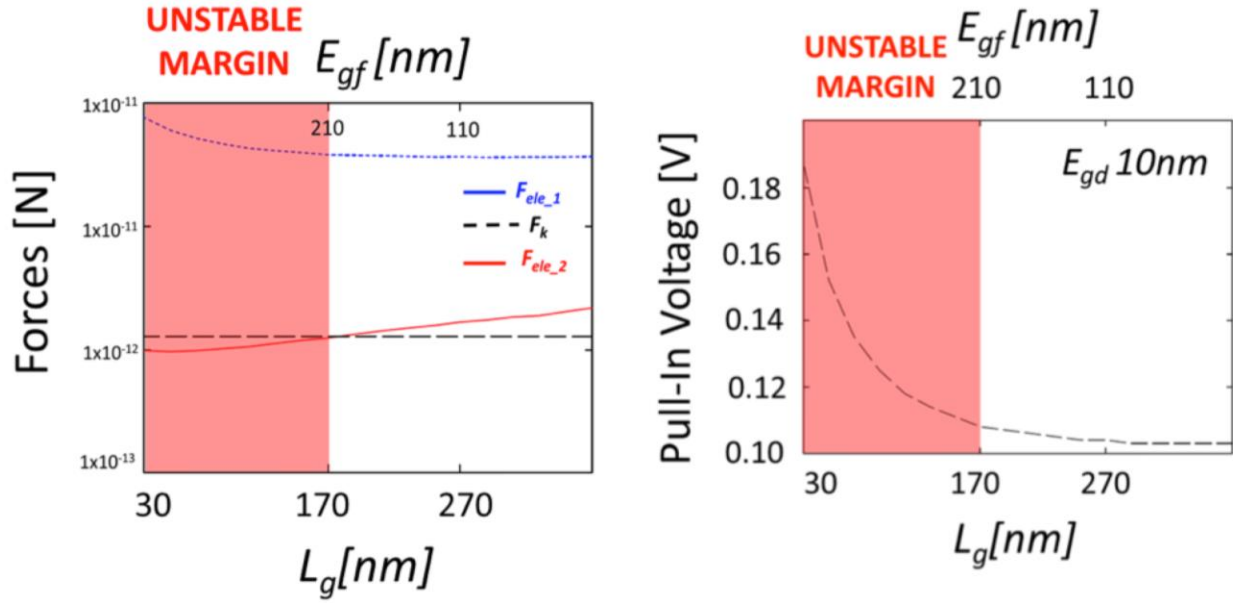


Fig. 74: Left) Electrostatic and spring forces versus L_g (and corresponding variation of the E_{gf} distance). Right) Pull-in voltage as a function of L_g (and corresponding variation of the E_{gf} distance).

In summary, in the single asperities dominated contact regime, if the adhesion force is negligible compared to the spring restoring force, and the drain is biased at V_{PI} the NEM relay scaling could come at the expense of contact stability. A saturation of R_c at $R_{sharvin}$ could cause R_c to become significantly smaller than R_b over a significant range of design parameters. Voltage redistribution upon contact could then cause the electrostatic force to decrease below the spring restoring force and cause an instability. Besides extrinsic series resistance management, it is possible to recover contact stability by adjusting the length and position of the gate along the beam.

3.5 CASE STUDY: NV SRAM

3.5.1 6-T SRAM

An SRAM is a volatile memory cell made using transistors (Fig. 75). Usually it has two n-FET access transistors (also called pass gates). It also contains two cross-coupled inverters, each comprising a pull-up (PU) p-FET transistor and a pull down (PD) n-FET transistor. The bit is stored on the storage node by the means of the two cross-coupled inverters. The bit stored can be either '0' (*i.e.* GND) or '1' (*i.e.* V_{DD}). The memory is accessed through the access transistors that are connected to the word line (WL), and to the bit line (BL) or the bit line bar (BLB). The different operations are storage, reading and writing operations.

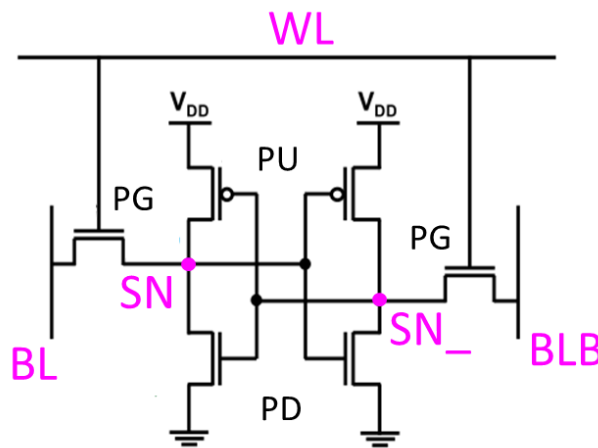


Fig. 75: Schematic of a standard 6T SRAM cell

3.5.1.1 STORING-WRITING-READING OPERATIONS

During the storing operation (Fig. 76 left) the PG transistors are biased on the OFF state; therefore, the WL is grounded. The data is held in the latch and kept in a stable state owing to the two cross-coupled inverters.

For a reading operation (Fig. 76 middle), the PG transistors are activated in the ON state (*i.e.* WL at V_{DD}). Additionally, both bit lines are precharged at V_{DD} . On the side of the bit cell where the bit '1' is stored, no current is able to flow through the Pass-Gate, in fact source and drain have the same potential ('1'). On the side of the bit cell where the bit '0' is stored, a reading current flows through both PG and PD transistor because they are in series. The '0' must be maintained; therefore, the reading operation does not have to alter the stored bit of the SRAM. Therefore, the PD must be "stronger" than the PG, this is usually obtained by designing a larger channel width for the PD.

The writing operation (Fig. 76 right) requires the WL to be at '1' and the bit line biased according to the bit to be written. There is a small subtlety on this system, since the reading condition requires to have $PD > PG$, the stored '0' cannot be changed to '1'. For example, if BL is at '1' and the storage node had a '0' stored, nothing can be done in this side, and the writing operation relies on the state variation of the other side, where the saved '1' can be changed to '0'. Being able to write a '1' requires having PG stronger than the PU. Therefore, writing operation imposes $PG > PU$.

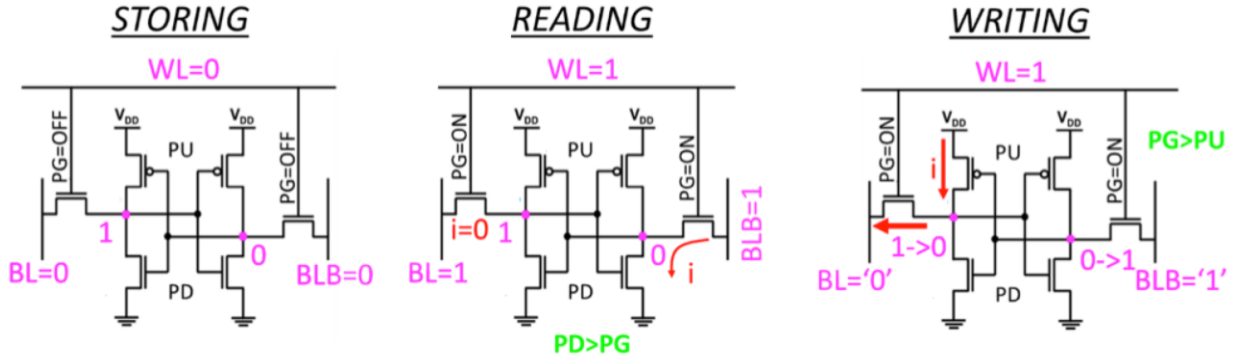


Fig. 76: Schematic of the storing, reading and writing operation of a 6T-SRAM

3.5.2 1-RELAY (R)-8T NON-VOLATILE (NV) SRAM

Fig. 77 shows a NV SRAM made by the standard 6T SRAM cell plus two additional transistors and a Non-Volatile NEM relay [Kato 2016a; Liu 2017; Xu 2014]. The source of the NEM relay is connected at the storage node of the SRAM left side. The left actuation electrode is connected at V_{DD} and the right actuation electrode is connected at GND . The left contact electrode is connected to an n-FET, which is activated if the RWL is activated (*i.e.* at V_{DD}).

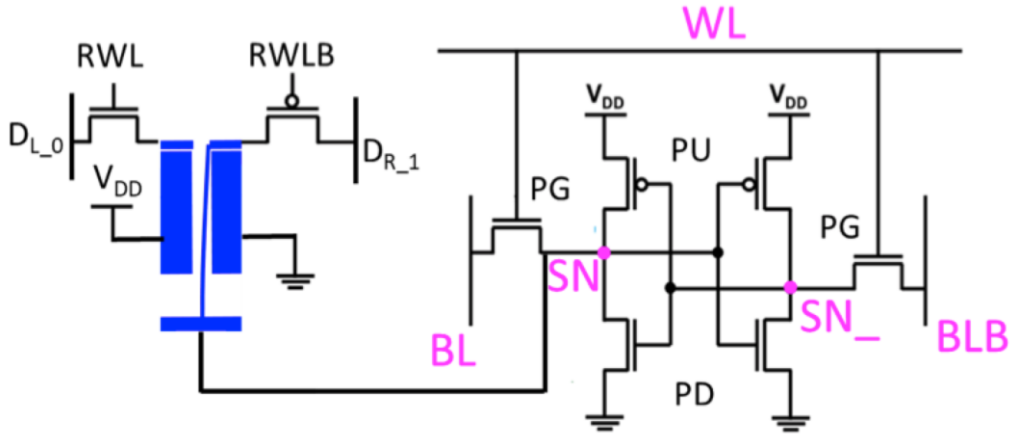


Fig. 77: Schematic of a 1R-8T NV SRAM adapted from [Xu 2014a]

The right electrode is connected to a p-FET transistor, which is activated if the RWLB is at (GND). In order to analyze the circuit operation, a Coventor MEMS+ SPICE macro-model of a NEM relay device is used. The operation of the NV-SRAM cell during normal SRAM operation proceeds in the same manner as an ordinary SRAM cell with one exception. The relay movable electrode follows the state of the SN: If the bit stored on the left side of the SRAM is ‘0’, the relay will move towards the left contact electrode and if the bit stored on the SN of the SRAM is ‘1’ the beam will pull in contact with the right electrode. The charge on the SN stays memorized because there is not current flowing through the relay (note: the RWL/RWLB pass transistor are OFF).

Once the SRAM power is OFF the data of the storage node stays saved by the means of the nano relay physical position. In fact, the relay, since it is nonvolatile, maintain its position also when the power is OFF.

To be sure that the relay had the time to contact the corresponding electrode (*i.e.* DL if bit stored ‘0’ and DR if bit stored ‘1’) it is important to let pass a time t_{OFF} between the last writing operation and powering OFF the cell Fig. 78.

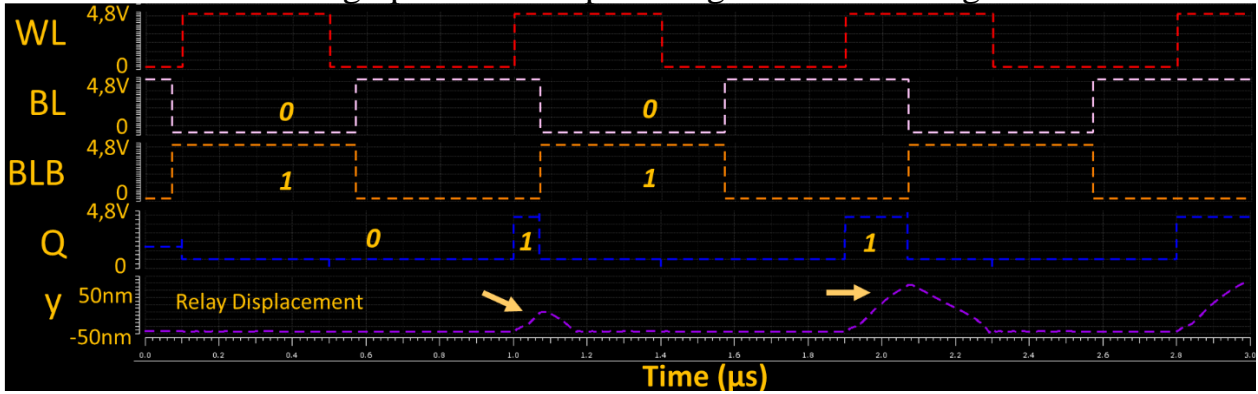


Fig. 78: Transient simulation of the SRAM writing operation, the relay does not have enough time to follow the change of the storage node. Relay parameters: $L_b=9\mu m$ $L_g=2.3\mu m$, $w=80nm$, $L_{ov}=70nm$ $E_g=6.58nm$. SRAM technology 130nm ST (working at 4.8V). Simulation done using MEMS+ plugin for cadence

This time corresponds to the RC delay plus mechanical delay t_m . In this case, the R depends on the pMOS/nMOS transistor ON resistance, on the resistance of the path towards the beam and on the beam resistance. The capacitance is the relay capacitance Eq. 55, which is usually very small (on the order of an attofarad) and for simplicity can be estimated by Eq. 64, which is the capacitance for a parallel-plate capacitor.

$$C_r = \frac{\epsilon_0 A_{ac}}{g_g} = \frac{\epsilon_0 t * L_g}{g_g} \quad Eq. 64$$

Where ε_0 is the vacuum permittivity, A_{ac} is the actuation area, g_g is the gap between the actuation electrode and the movable electrode. The mechanical delay can be estimated using Eq. 48.

Taking as example a Silicon relay, being the parameters listed in the table below, the capacitance value is $C_r = 48\text{aF}$. Considering a resistance of $R_C = 10\text{M}\Omega$, the RC delay becomes equal to $\sim 0.48\text{ns}$. The mechanical delay is $t_m \sim 312\text{ns}$. Resulting in a mechanical delay predominant with respect the RC delay. Therefore, before powering OFF the cell one needs to wait for a time $t_{OFF} \sim 312\text{ns}$ in order to be sure that data of the SRAM is mechanically saved on the beam. In Fig. 79 a simulation made by using Cadence with Coventor MEMS+ plug-in shows the time before the relay is in contact with the opposite drain (*i.e.* beam tip position from -50nm to 50nm).

L_b	w	t	Density	g_d	V_{DD}	g_g	L_g	E	L_g	m
$12\mu\text{m}$	50nm	50nm	2330Kg/m^3	50nm	$1,8\text{V}$	55nm	$1,14\mu\text{m}$	$168,9\text{GPa}$	$6\mu\text{m}$	$6.99 \cdot 10^{-17}\text{kg}$

The fact that the relay has a mechanical delay allows to preserve the contact surface of the relay and to preserve its endurance. In fact, the relay is not able to follow every single SRAM state change, but only those that are spaced by a time than t_m , resulting on an overall endurance improvement.

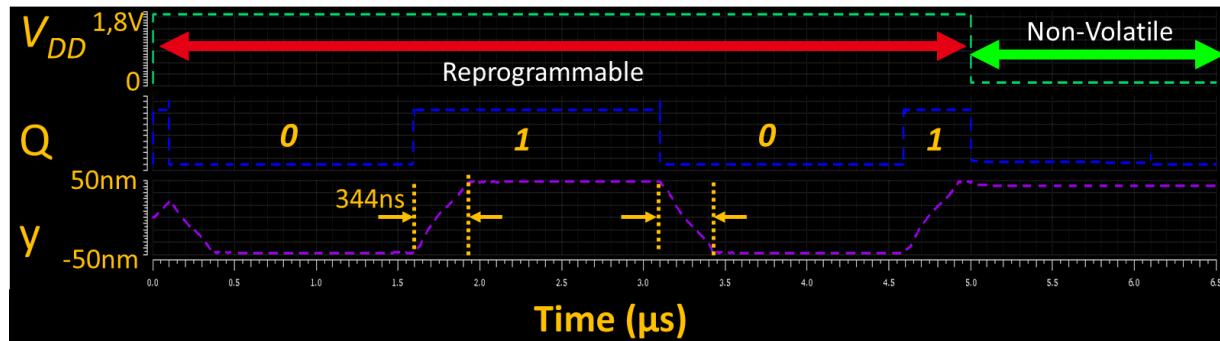


Fig. 79: Transient simulation of the SRAM writing operation, showing that the relay needs a certain amount of time to follow the SRAM state, and showing the relay non-volatility (*i.e.* when the power of the NV-SRAM is OFF the relay keeps its positions). Simulation done using MEMS+ plugin for Cadence.

3.5.2.1 RE-WRITING OPERATION

The term ‘re-writing’ is here intended to describe the operation of injecting on the storage node of the SRAM, the last saved bit after a power shutdown. This operation requires some shrewdness. The WL line has to be at ‘0’ and the SRAM has to stay disable for a time t_{re} (V_{DD} not applied). The RWL has to be at ‘1’, the RWL_b at ‘0’, and both actuation electrodes have to be grounded. The source of the n-FET (D_L) as to be biased at ‘0’ and the source of the p-FET (D_R) has to be at ‘1’. After a time t_{re} , the SRAM can be enabled (V_{DD}) and the data is restored. Fig. 80 shows an example of writing ‘1’ powering OFF the cell and re-writing data ‘1’

operation. The resistive path through the relay plus the re-writing assistant transistor RC delay gives the waiting time t_{re} . Waiting t_{re} before powering back on the 6T-SRAM pre-charges the node Q_- allowing to inject the data even if the ON resistance of the relay is very large (within a certain limits). Compared to writing the SRAM through the access transistors, the delay associated to this re-writing automate sequence is made by the t_{re} and the RC delay made by the resistive path through the relay and the re-writing assistance transistor. This RC delay has a higher value compared to the RC delay of the only access transistor connected to the storage node and it is highly influence by the relay ON resistance.

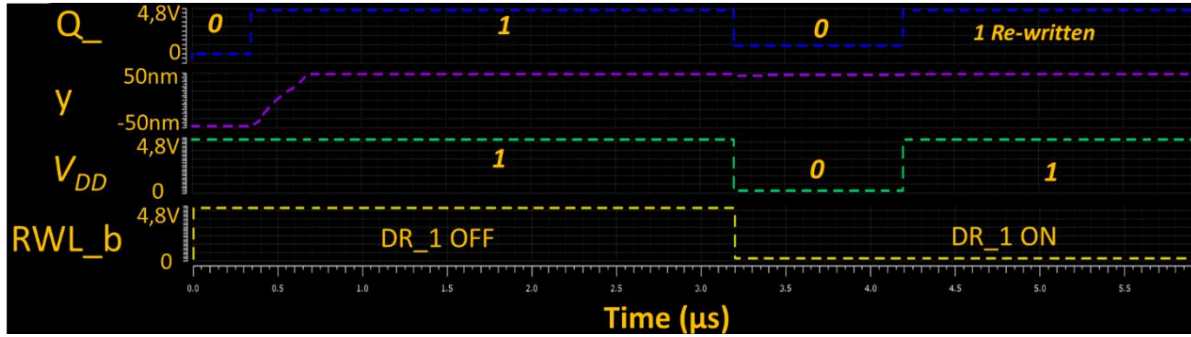


Fig. 80: Transient simulation of the re-writing operation. The “1” needs to be re-injected inside the storage node Q of the SRAM. The “1” is written once the V_{DD} is back ON.

Once the SRAM supply voltage is enabled, there is a charge and discharge current flows with respect the node Q and Q_- as shown in Fig. 81.

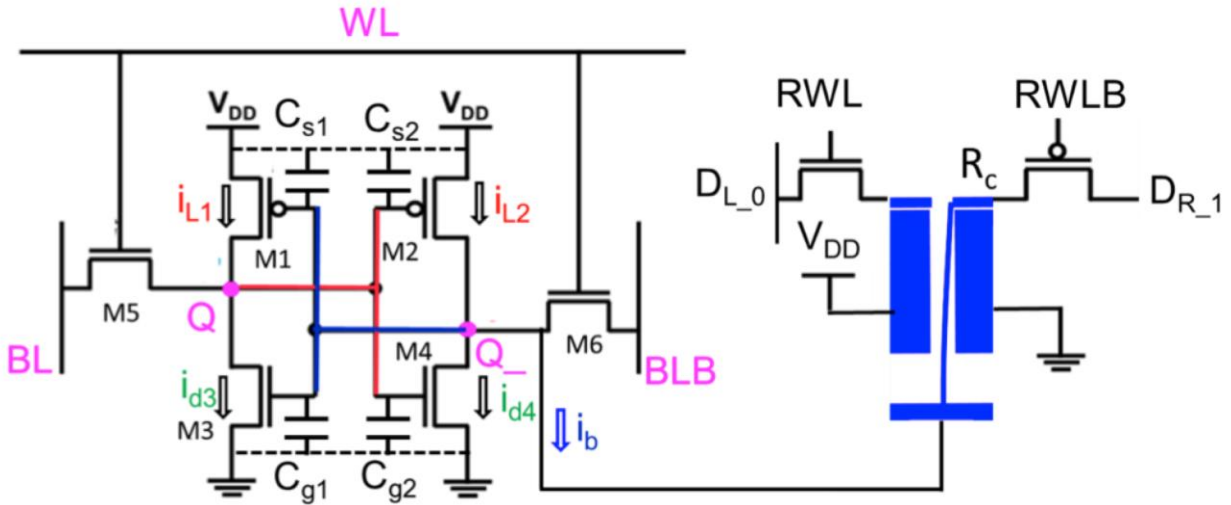


Fig. 81: Schematic illustrations of re-writing operation showing the charge and discharge current flows with respect the storage nodes.

In this figure C_{s1} , and C_{g1} represent the parasitic capacitances at nodes Q_- , C_{s2} , and C_{g2} represent the parasitic capacitances at nodes Q and R_c represent the relay contact resistance. The re-writing ‘1’ operation is explained below. Let us assume that RWL_B is low, DR_1 is high and then V_{DD} of the SRAM enabled. Eq. 65

shows the voltage at the storage node Q and Eq. 66 shows the voltage at the node Q_- :

$$V_Q = \frac{C_{s1}}{C_{sg1}} V_{DD} + \int \frac{1}{C_{sg1}} (I_{L1} - I_{D3}) dt \quad \text{Eq. 65}$$

$$V_{Q_-} = \frac{C_{s2}}{C_{sg2}} V_{DD} + \int \frac{1}{C_{sg2}} (I_{L2} - I_{D4} - I_b) dt \quad \text{Eq. 66}$$

Where $C_{sg1} = C_{s1} + C_{g1}$ and I_{L1} is the charging current and I_{D1} is the discharging current with respect the node Q through the p- and n-channel MOSFETs of the Left inverter. $C_{sg2} = C_{s2} + C_{g2}$ and I_{L2} is the charging current and I_{D3} is the discharging current with respect the node Q_- through the p- and n-channel MOSFETs of the right inverter. I_b is the current leaking through the relay, and the channel of the RWL-B p-mos. I_b is highly influenced by the contact resistance of the relay (higher the contact resistance of the relay lower I_b). We can assume that $C_{s1} \approx C_{s2}$ and $C_{g1} \approx C_{g2}$ because of the SRAM cell symmetry and because of the similar bias condition of both inverters. Therefore, V_Q and V_{Q_-} increases gradually, however V_{Q_-} is higher than V_Q because of the charges coming from the write assist transistor (DR_1) and presence of the NEM relay. The presence of the relay creates a difference on the discharge rate of the node Q_- and Q. Once $(V_{DD} - V_{Q_-})$ and $(V_{DD} - V_Q)$ are larger than the threshold voltage of transistor M_1 and M_2 (i.e. $V_{DD} - V_{Q_-} > V_{TH1}$ and $V_{DD} - V_Q > V_{TH2}$), nodes Q and Q_- are charged by I_{L1} and I_{L2} .

It means that M1 and M2 are working on the saturation region and the current flowing through them depends on the gate bias voltages V_Q and V_{Q_-} . Since, the potential V_{Q_-} is higher than V_Q the relation $V_{DD} - V_Q > V_{DD} - V_{Q_-}$ is maintained resulting in $I_{L1} > I_{L2}$. This last relationship allows enhancing the voltage difference between V_{Q_-} and V_Q because it accelerates the charging of node Q_- . In the end, the amount of voltage in the node Q_- and the one present on the node Q allows to turn ON the M3 transistor and turning OFF the M1. At the same time, the node Q is discharged through M4 and V_Q is reduced, leading to a further increase in the charging of the node V_{Q_-} by M1 and the impossibility to switch ON M3. At this moment, there is a complementary push-pull operation of the two inverters (M1&M3 and M2&M4) of the SRAM. V_{Q_-} and V_Q are in a bi-stable condition, and V_{Q_-} approaches $\sim V_{DD}$ together with the increment of V_{DD} . At this point, '1' at Q_- node is re-written. However, V_{Q_-} it might not be totally at the high level because of some current leakage towards the switch and the re-write assistance transistor. There is a limitation linked on the maximum contact resistance that the relay can have. Fig. 82, shows the re-write operation through the write assistant p-MOS transistor for a vary number of NEM relay ON resistances R_R . The SRAM cell does not have power (V_{DD} not biased) and storage node Q_- has a certain value, which depends on the previous data stored and on the time of power off (therefore on the

time that the node had to discharge). As the contact resistance of the relay increases the time of restoring the data on the Q_- node increases.

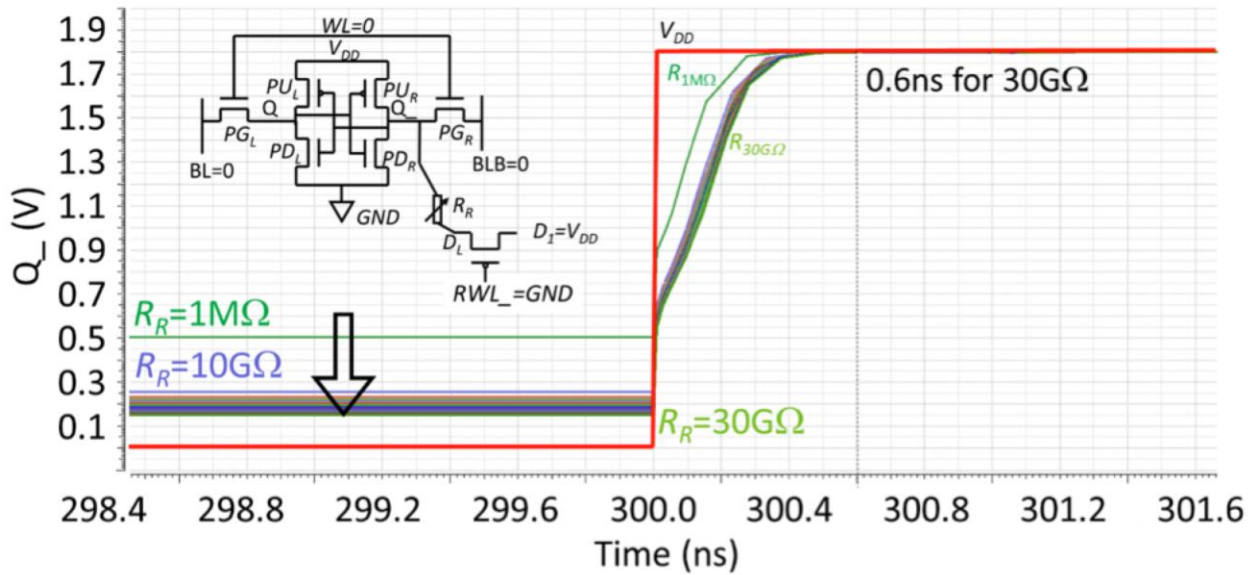


Fig. 82: Transient analysis of the rewrite assistant operation for an interval of ON resistance going from $1M\Omega$ to $30G\Omega$.

If the beam ON resistance is too high (in our case larger than $30G\Omega$) the data might not be restored, the level of the storage node Q and Q_- takes random values. In general, the ON resistance should be smaller than twice the pull up resistance at the off state, in this way the charges go through I_b faster than they discharge through I_{D3} .

However, the advantage on having high relay ON resistance, it is that there is less leakage towards the relay hence, therefore less power consumption of this restoring operation.

For example, for a 1R-8T NV SRAM made by 130nm technology, the delay of a re-writing operation is 0.45ns in case of an ON resistance of $1G\Omega$.

3.6 CONCLUSION

In the beginning of this chapter, an introduction of NEM relay physical behavior is explained through the basic equations of the parallel plate capacitor approximation. The modeling of cantilever and fixed-fixed beam design has been explained, introducing the beam stiffness and the relay mechanical delay. All of above helps on the understanding the cantilever relay pull in modeling. In order to increase the relay robustness and to increase the relay endurance, an optimized pull-in margin (*i.e.* VCPI-VPI) can be found. The detailed analytical study of paragraph 3.2 shows how important it is, while designing the relay, to pay attention to the actuation electrode length and position with respect the beam electrode. It

shows that for some cantilever beam lengths, it could be more convenient, for improving the *VCPI-VPI* margin, to have shorter gate collocate towards the tip.

In this thesis, we evaluate the use of 5-Terminal SPDT NEM relay as auxiliary nonvolatile element for CMOS circuitry. Therefore, a careful balance of the forces equations governing the system in order to obtain programmable/reprogrammable non-volatile relays is designed in 3.3. Unfortunately, there is a lot of uncertainty on the value of the adhesion force and additionally, this force varies during cycling. This creates a need for a robust relay design with a working margin that is as large as possible. A new design methodology is explained in these sections, which allow one to find and to validate various functional NV NEM relay designs through FEM simulations.

Paragraph 3.4 is a small digression to understand how size scaling the relay could influence the contact stability on 3-Terminal volatile NEM relay.

A case study of non-volatile relay for SRAM has been completed, exploring the influence of the relay on the writing operation, and analyzing the re-writing operation through the relay and the write assistance transistor.

***4 FOCUS ON 3D MONOLITHIC
NEMS/CMOS COINTEGRATION
PROCESS FLOW***

4.1 NEMS/CMOS MONOLITHIC CO-INTEGRATION

The monolithic co-integration of NEMS devices with CMOS circuits opens a vast number of possibilities in science and engineering. NEM relays combined with IC can enhance CMOS performance and capability, *e.g.* giving nonvolatile properties to SRAM [KATO 2016B; XU 2014B]. NEM sensors can be used to measure virus and molecules weight or can be used for DNA detection [BERTRAND 2011]. Usually the NEM sensors need CMOS circuitry to read the collected data, or to amplifying the electrical signal converted from a physical response (*e.g.* [SAGE 2018, P.]).

MEMS and CMOS were usually fabricated in two different substrate and bonded together off-chip, allowing an independent optimization of both MEMS and CMOS technology. However, the interconnection parasitic coming from the wires used for pads/to pads bonding can be prohibitive now that the electromechanical systems have reached the nano-size. Additionally, the packaging cost of not-monolithic hybrid approach is higher than the one of the monolithic approach.

In summary, the monolithic co-integration between NEMS and CMOS allows to improve performance, signal-to-noise ratio, power consumption, and global system compactness [OLLIER 2008]

Fig. 83 shows the schematics of an IC, pointing out the various CMOS parts and highlighting the area in which the NEMS can be monolithically fabricated within the CMOS substrate. The standard CMOS stack is divided into three main parts: The front-end-of-line (FEOL), the middle-end-of-line (MEOL) and the Back-end-of-line (BEOL). The FEOL is the first fabricated portion of IC that includes individual devices transistor, capacitor, resistor etc. It contains all the CMOS fabrication process needed to form fully isolated CMOS elements. It is the first part to be processed and it includes high temperature steps such as dopant activation ($\sim 1050^{\circ}\text{C}$) and annealing to improve transistors performance. The middle-end-of-line MEOL includes silicidation and contacts. The BEOL is the third portion of IC fabrication where individual devices get interconnected with wiring within the wafer. Common metal used for the interconnections are copper (Cu) and aluminum (Al). BEOL includes vias, insulating layers (dielectrics), metal levels and pads.

Knowing the schematic of an IC allows dividing the NEMS monolithic fabrication approaches into three main categories: Pre-CMOS, inter-CMOS and post-CMOS. The post-CMOS category can be sub-divided into above IC and BEOL.

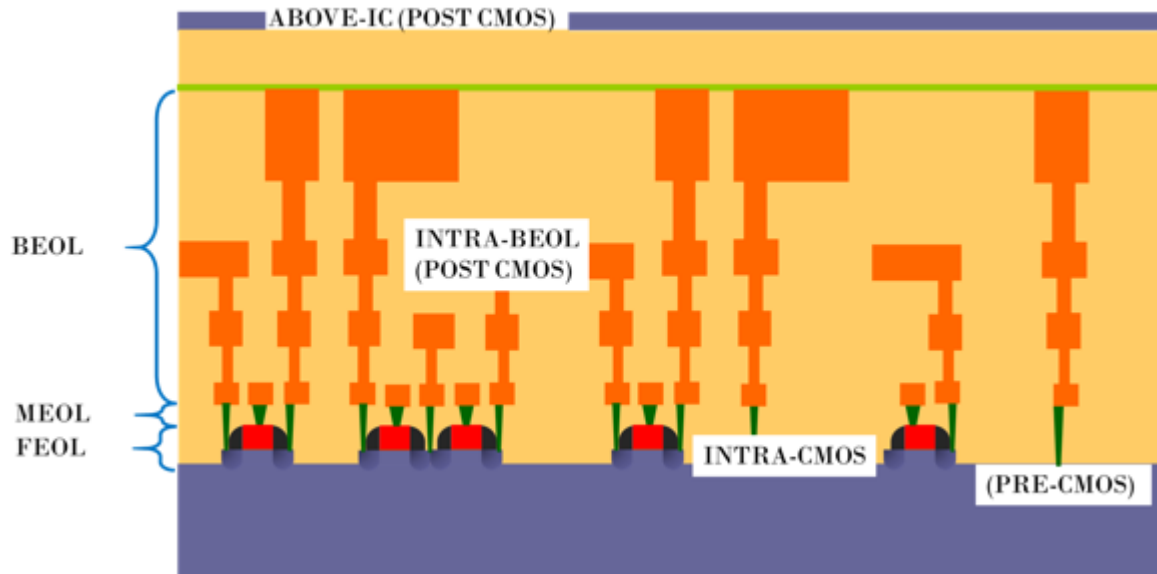


Fig. 83: Schematic cross section of a CMOS stack to identify the area in which NEMS can be fabricated

Pre-CMOS: All the steps required to fabricate the NEMS are performed prior to CMOS device fabrication. The NEMS is usually fabricated in a recessed trench of the silicon wafer (see Fig. 84). After the NEMS patterning and encapsulation, there is a planarization step that allows the CMOS fabrication. In the Pre-CMOS NEMS fabrication, there are no thermal budget constraints. The NEMS release is done as final step after the interconnections fabrication. One drawback of the pre-CMOS fabrication is that typically pre-processed wafers are not allowed in dedicated CMOS fabrication fabs. Another drawback is that the large circuit area, in fact the NEMS are fabricated next to the CMOS device. Example of this fabrication method can be found in [SMITH 1995].

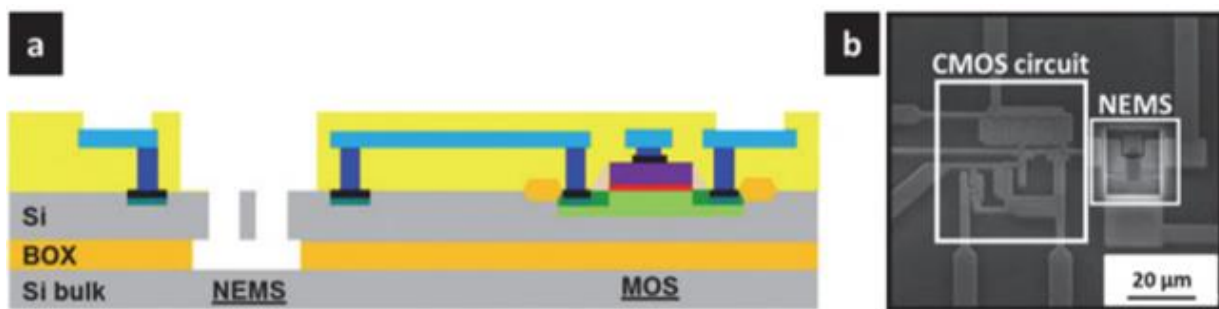


Fig. 84: Schematic representation of the fabrication process of the sensor element and the SEM TOP view by [Durauffourg 2015]

Intra-CMOS: This method, to our knowledge has been used mostly for MEMS structures (e.g. pressures sensors of tire monitoring system). The MEMS and IC processing are fabricating by interleaving CMOS device fabrication process with poly-Si deposition and micromachining steps to form MEMS structures. The MEMS release is done after the IC fabrication. Monolithic MEMS inter-CMOS

integration using interleaved MEMS and IC processing offers the possibility of integrating high-performance MEMS materials in the CMOS circuit substrate. Usually the MEMS and CMOS structures are placed side-by-side on the substrate, limiting the overall integration density.

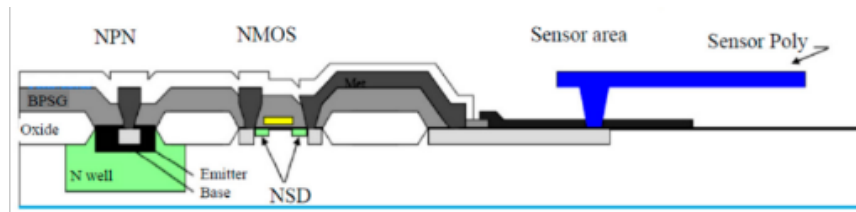


Fig. 85: Schematic representation of the fabrication process of the sensor element [Kuehnel 1994] taken by [Qu 2016]

- **Post-CMOS:** The NEMS are fabricated after the CMOS devices have been fabricated. This technique will be explained in detail below and in the annex.

The NEMS can be fabricated within the metal layers of the BEOL interconnections (here called BEOL NEMS), or above the encapsulation layer of the CMOS IC. The post-CMOS approach allows to considerably decrease the fabrication cost. In fact, besides the lower cost due to monolithic approach, there should not be any need of modifying the standard baseline CMOS process. Another advantage is that the NEMS are defined above the CMOS devices, allowing area saving compared to the pre-CMOS methods.

The definition of the BEOL NEMS is made on the CMOS foundry. The processing steps to be added after the NEMS and IC fabrication are 1) the hard mask deposition and patterning in order to define a release window for the NEMS, and 2) the release itself. The main drawbacks are the limitations on terms of material choice and device design (*e.g.* minimum gap, minimum width defined by the technology design rules). Post-CMOS BEOL approach can be found in [KWON 2017; MUÑOZ-GAMARRA 2014; RIVEROLA 2018]. During this thesis, we have worked at the release process of BEOL NEM relays fabricated on the CMOS BEOL in collaborations with UC Berkeley. The process is described in the annex 6.2. The Cu NEMS, it is a lateral actuated relay with a vertical structure made by more than two metal layers. The vertical definition should help on decreasing the actuation voltage. Defining a relay in more than one metal layers increases complexity of the release process.

NEMS above IC manufacturing requires additional processing steps with respect the BEOL NEMS, such as 1) oxide deposition 2) interconnections patterning between the NEMS and the CMOS and the NEMs itself 3) metal deposition in our case 4) wafer bonding/grinding. Fabricating a NEMS above CMOS requires respecting the thermal budget given by the CMOS technology. For instance, for not advanced CMOS technology node the thermal budget limits is 400°C for 2h because of the Copper and Aluminum interconnections (Cu and Al are likely to

diffuse above 400°C). This constrains limits the use of front-end materials such as mono-crystalline Silicon (mono c-Si) or the use of high temperature fabrication process such as dopant activations through annealing.

In the following sections, we described the approach we used to fabricate highly doped mono-crystalline Si NV NEM relay monolithically integrated above CMOS.

In the annex 6.1, we describe the manufacturing process flow we used to fabricate, doped poly-Si NEMS above CMOS obtained with a-Si deposition and laser annealing for both dopant activation and a-Si recrystallization. This approach was investigated because of the lower cost compared to the mono c-Si relays above CMOS. Both of the approaches besides the NEMS/CMOS co-integration, allow to hybrid co-integrate NEMS relay and NEMS sensor (*e.g.* resonator). In fact, the relays and the sensors can be both defined in the same layers.

In the annex 6.2, we describe the manufacturing process flow we used to fabricate, Cu BEOL NEMS. This approach was investigated because of the growing interest and promising success towards this technology.

4.2 c-Si NEMS ABOVE CMOS: FLOW OVERVIEW

This section describes the integration process of mono-crystalline Silicon (c-Si) NEMS relays above CMOS, developed at Leti and improved during the work of this thesis. The 3D process is monolithic, sequential and involves wafers bonding. The relay patterning is done after bonding a Silicon-On-Insulator SOI wafer above the “CMOS” wafer. Therefore, the NEMS patterning and the interconnection fabrication process require withstanding the thermal budget of the CMOS wafer. The maximum thermal budget allowed for metal interconnections in our technology is 400°C for 2 hours. Overpassing this budget will cause a degradation of the circuits due to an increase of the interconnection resistance variability and of the access resistance, and cause a shift of the threshold voltage [FENOUILLET-BERANGER 2014; TAKEUCHI 2005]. Additionally, the presence of heavy metals, such as Co, Cu and Ta, typical used in the middle and back end of line of the CMOS wafer, can be problematic for certain tools, especially during the Si NEMS patterning in Si or SiO₂ dedicated etching tools. The presence of those contaminants could degrade the electrical behavior of the processed wafers and it will leave the tools contaminated.

The wafers used within the scope of this thesis are part of a Leti multi-project. This means that the integration process needs to be versatile for various applications (*e.g.* Back-End memories or NEMS mass sensors). For that reason, the CMOS technology, the materials, the thickness of the Si mono-crystalline layer and its doping level, are fixed by conditions resulting from a trade-off between the targeted applications of this multi-project.

In the following sub-sections, the process flow is explained in detail. It is separated into six major steps listed as follows: preparation of the SOI substrate 4.2.1; preparation of the CMOS substrates 4.2.2; Bonding, Grinding and Trimming 4.2.3; NEMS patterning 4.2.4; interconnections and pads definitions 4.2.5; and finally NEMS Release 4.2.6.

4.2.1 PREPARATION OF THE SOI SUBSTRATE

This sub-section describes the steps to prepare the SOI substrate for the integration. The goal is to modify the original SOI wafer in order to obtain a substrate suitable for both SiO₂-SiO₂ bonding and NEMS device definition. Therefore, the Silicon top film has to be relatively thick to allow for the formation of a top layer of SiO₂ and at the same time to leave the right amount of mono c-Si after oxidation. The table below describes the layers after being processed and their scope:

LAYER	SCOPE
<i>TOP Layer</i> Thermal Oxide (SiO ₂)	<ul style="list-style-type: none">• Direct SiO₂-SiO₂ Bonding• Sacrificial layer for NEMS release
<i>MIDDLE layer</i> Mono-crystalline Silicon (mono c-Si)	<ul style="list-style-type: none">• NEMS patterning

The choice of having a SiO₂-SiO₂ direct bonding allows for burying the bonding interface below the sacrificial layer, helping to avoid risk of de-bonding during the switch release. The oxide can be deposited or formed by thermal oxidation; the following process uses thermal oxidation. This choice is motivated by the fact that thermal oxide gives higher etching controllability and less residues during HF vapor release (see section 4.2.6) compared to other SiO₂ types.

The thermal oxide layer is used for the wafer bonding but also as sacrificial layer for the NEMS release. Hence, it has to be sufficiently thick to allow the release without causing any impact on the bonding interface. The c-Si layer is used for the relay electrodes definitions. Since the wafers are part of multi-projects, the c-Si layer thickness depends by the specification given by the mass sensors present in the same mask. For in plane laterally actuated relays, such as ours, the actuation voltage does not depend on the thickness (see Chapter 3)

Fig. 86 shows the integration process flow of the SOI wafers step by step. All the fabrication steps with temperature exceeding 400°C (2h) must be performed before the bonding. The process flow starts with the thermal oxidation of the c-Si. The thermal oxidation reaches very high temperature (~1050°C for ~1h). The total thickness of c-Si is 400nm, meaning that to obtain 160nm during the thermal oxidation it is necessary to consume 240nm of Si and to target 520nm of SiO₂, which is enough to guarantee the release of the relays designed on the mask set.

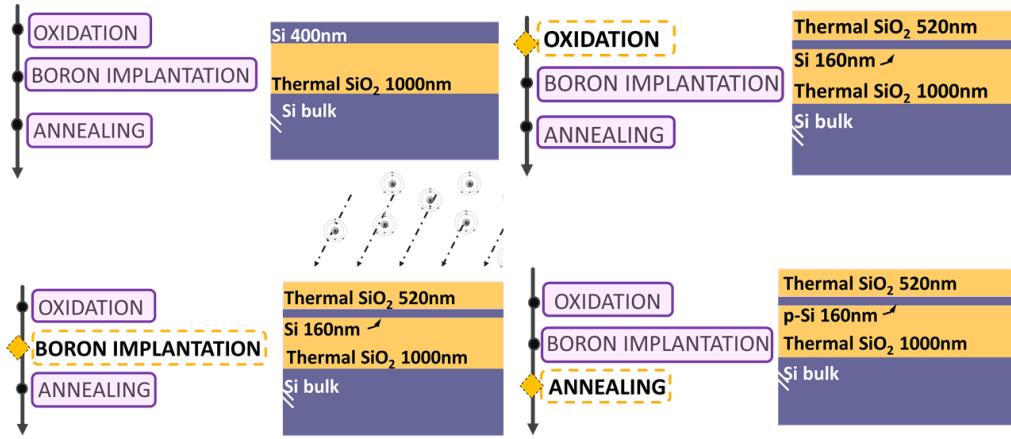


Fig. 86: Cross-sectional schematic of an SOI wafer and the fabrication steps before molecular bonding: Thermal oxidation at 1050°C to obtain the Si thickness required (here 160nm), Boron implantation to dope the Si at 5.5×10^{19} at/cm³ and thermal annealing to activate the dopants at 1050°C.

The c-Si layer has to be doped for the NEMS electrodes to be conductive. In our case, Boron (B) is used as doping species as required by piezoresistive mass sensors present on the same mask. Fig. 87 shows the B penetration distance used for a thickness of 520nm of SiO₂, and 160nm of c-Si. The implantation dose chosen 10^{15} at/cm² allows having a dopant concentration of 5.5×10^{19} cm⁻³.

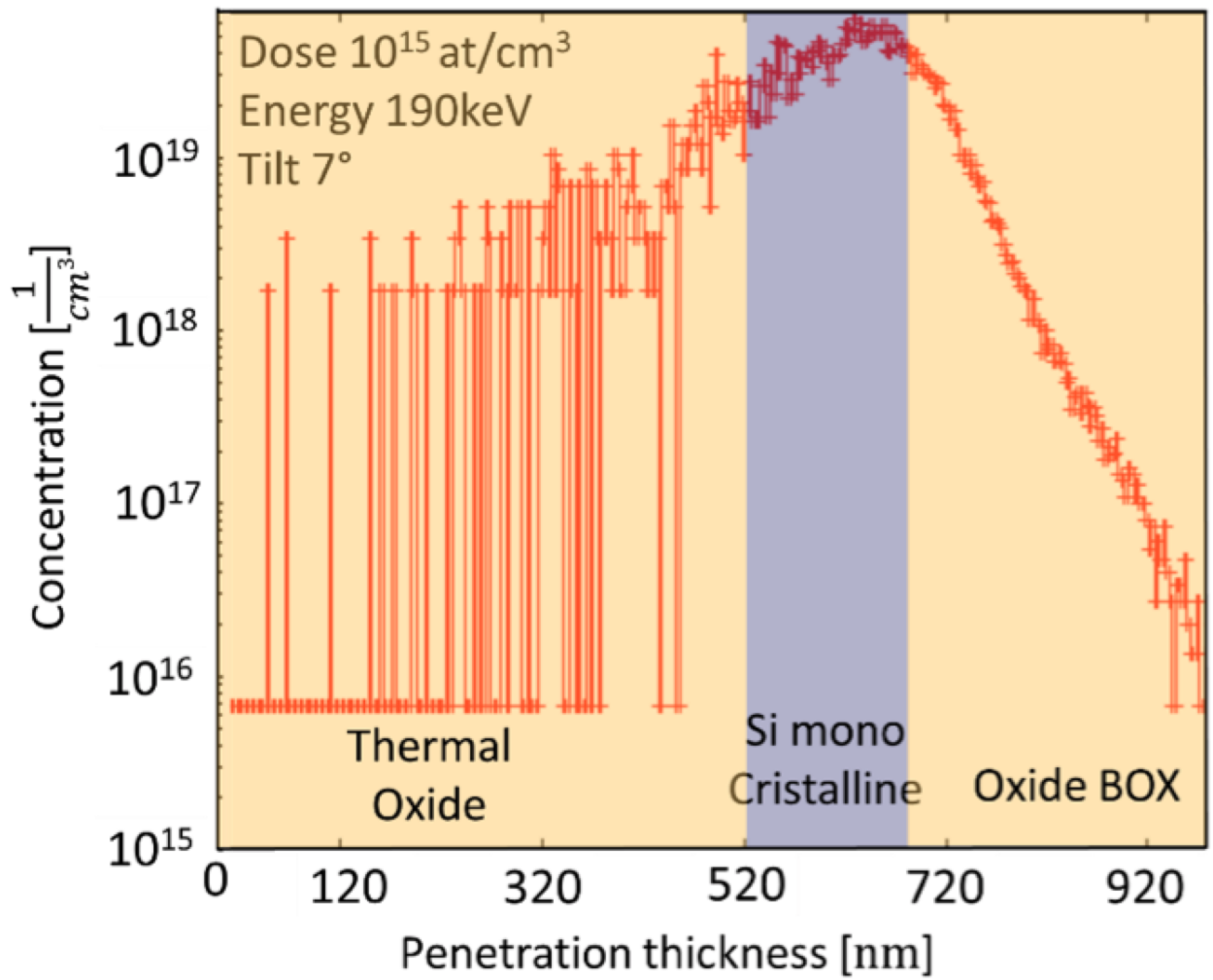


Fig. 87: Doping concentration depth profile for a dose of 10^{15} at/ cm^2 with energy of 190keV and tilt of 7° .

The B is activated by an annealing at $1050^\circ C$ for 5min under N_2/O_2 environment (1% of O_2 diluted in N_2). The small percentage of oxygen allows avoiding surface degradation (pitting effect) due to the presence of volatile SiO_2 in the annealing chamber. After the B activation, the dopant concentration is up to 5×10^{19} making the c-Si layer quasi-metallic (resistivity value of $\sim 0.002 \Omega \cdot cm$).

The final steps before the bonding consist into an annealing under N_2 ambient at $400^\circ C$ for 1h20min to dehydrate the surfaces before the bonding. This dehydration steps are needed to eliminate the water adsorbed by the oxide coming from the clean room humidity. The presence of an excessive amount of water on the bonding interface can causes significant damage during the bonding annealing step [MASZARA 1988]. In fact, the presence of large amount of water molecules trapped at bonding interfaces can cause low temperature oxidation ($\sim 200^\circ C$). This phenomenon is a source of hydrogen gas and annealing voids. Therefore, a large number of voids occur at elevated annealing temperatures [WANG 2017].

4.2.2 CMOS SUBSTRATE PREPARATION

In the frame of this thesis, we post-processed 200mm wafers with STMicroelectronics 130nm technology node and four interconnection levels.

The CMOS wafer needs to be prepared to form direct SiO₂ bonding with the SOI wafer, the bonding requires a very flat surface without any contaminant particles; see Table 6 for the bonding specifications. The wafer dies present some surface topology especially above the measurement boxes (100μm x 100μm) or to the large pads (90μm x 50μm) defined in the last metal level. The topology is often caused by dishing and erosion due to the damascene fabrication process of the Cu levels [GUPTA 2009].

The surface also needs to be cleaned from possible contaminant particles. In fact, a particle present on the bonding interface creates a void one order of magnitude larger than its diameter [PHILIPPE 2016]). It is also necessary to control the wafer bow, after a certain value the bonding activation energy is higher than the one available for a low temperature bonding, resulting in a bonding failure.

Surface Roughness	AFM area: 1μm x 1μm => RMS < 0.5nm
Bow	<150μm
Particles	Diameter = 90nm Max. Number of particles = 100

Table 6: Bonding specification

Fig. 88 shows the integration process flow to prepare the CMOS substrates for bonding. To smooth out the top surface 500nm of TEOS PECVD oxide is deposited and planarized by a CMP, leaving 300nm of oxide thickness. To obtain a very flat surface it may be necessary to perform more than one cycle of deposition/planarization. In our case, the wafers obtained an RMS surface roughness of 0.3nm after the second cycle of deposition and CMP.

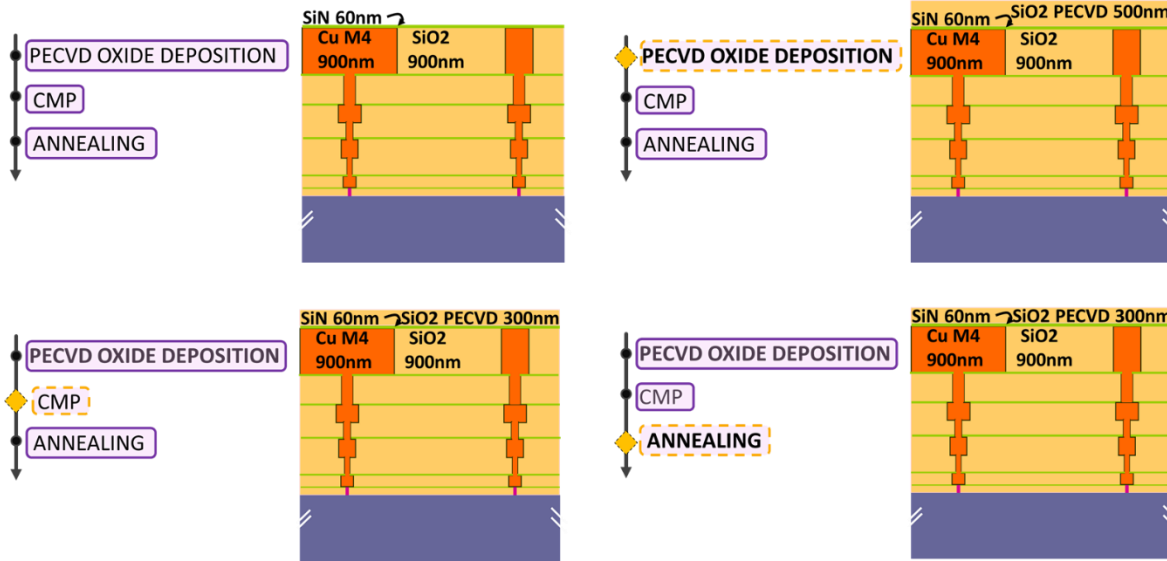


Fig. 88: Cross-sectional schematic of the CMOS wafer and the process steps before molecular bonding: PECVD TEOS oxide deposition (at 400°C) as bonding material, CMP to maximize the planarization of the top surface and thermal annealing for bonding oxide dehydration.

The final steps are the same as for the SOI wafer. The oxide needs to be annealed to eliminate the excess humidity as done with the SOI wafer 4.2.1.

4.2.3 BONDING, GRINDING AND TRIMMING

In the previous sections we explained how the wafers are prepared for the bonding steps. The SOI wafer has seen the only high temperature steps, and both wafers top surfaces are now ready for bonding. Fig. 89 shows the integration flow including bonding, grinding and trimming.

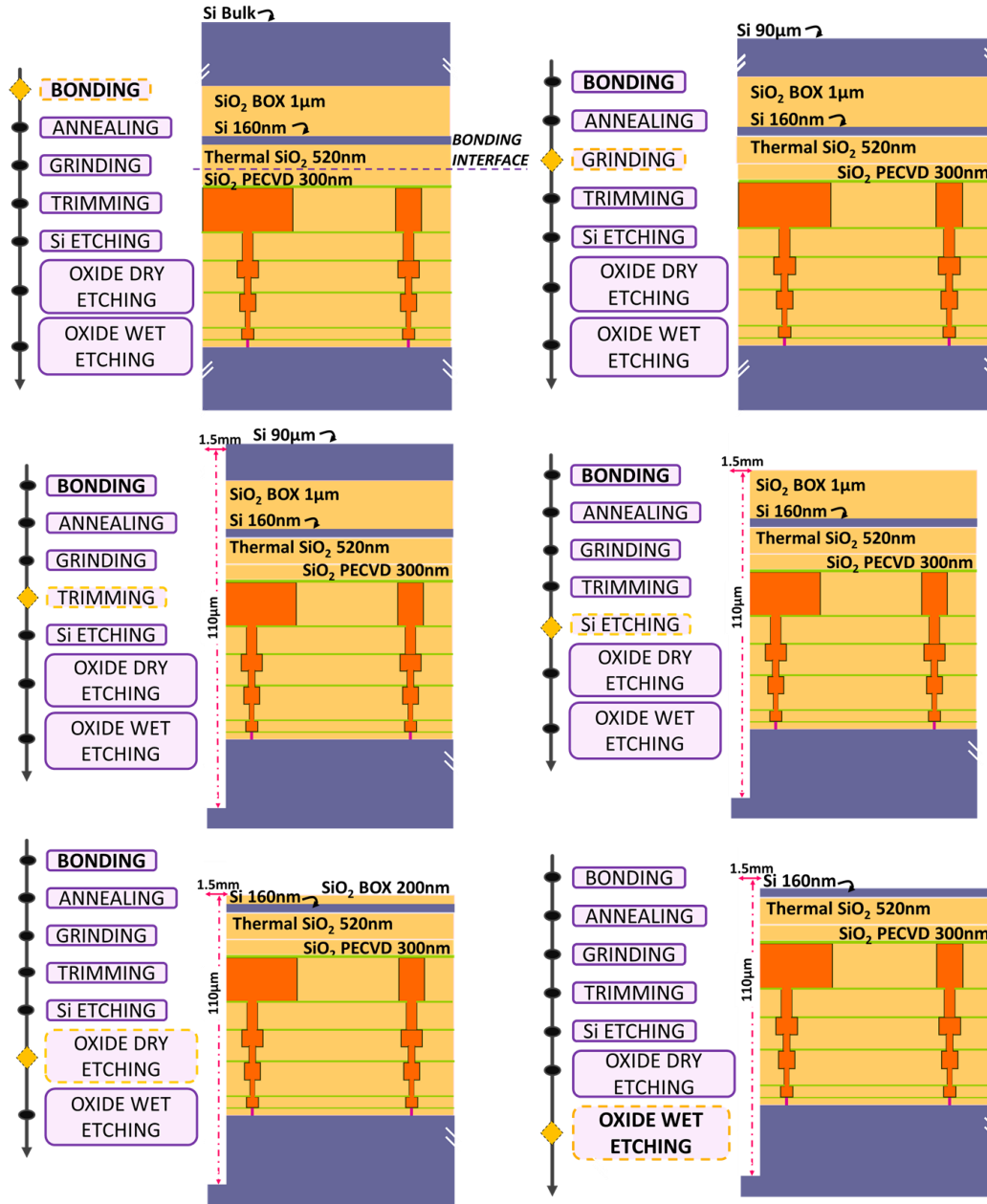


Fig. 89: Cross sectional schematic view of the final assembly process of the SOI and CMOS wafers. Molecular bonding and thermal annealing at 400°C, Si bulk grinding (rough polishing of 723 µm), wafer trimming to remove the damaged wafer side, Si and SiO₂ anisotropic etching (SiO₂ left 200nm) and the final oxide etch is done through etching wet process to avoid Si surface damage.

After planarization made by a slight CMP the wafers surface are activated in a nitrogen plasma. All the requirements for low temperature bonding are fulfilled. Finally, the two wafers are placed in contact under ambient conditions. There are several methods, *e.g.* infrared camera or acoustic waves, to check whether the bonding has succeeded [CUNNINGHAM 2011; GOSELE 1998; LIEBENS 2019]. Here the bonding quality is checked by a scanning acoustic microscopy (SAM). This tool

sends an acoustic signal and it measures the time it takes to reflect back to the detector. The resolution depends on the wafer thickness and on the signal frequency; its maximum resolution value is $0.5\mu\text{m}$. Fig. 90 and Fig. 91 show two images obtained by this method: both are CMOS wafers bonded to an SOI wafer. White zones mean that acoustic reflection occurred indicating unbonded areas, whereas black zones mean that no acoustic reflection occurred indicating bonded area. Dies contours are well visible in both figures, however Fig. 90 shows a large void on the bonding surface, making impossible to continue the process. The more metal layers are present on the CMOS wafer, the more challenging it is to obtain a successful bonding. To consolidate the bonding, annealing is performed under N_2 ambient at 400°C for 90min.

After bonding, the c-Si layer is buried under the Si thin film, Buried Oxide and Si handle. A first grinding step is performed, coarsely grinding a large part of the Si handle. A finer grinding step is used in the second step to polish the wafer and to accurately grind the wafer to the required thickness. The finer grinding step is itself followed by a dry polishing. A short TMAH Si-etching step is included to eliminate some grinding residues that may otherwise result in micromasking of the underlying oxide. The BOX is etched in two steps; the first 800nm are etched by RIE and the remaining 200nm using a wet process.

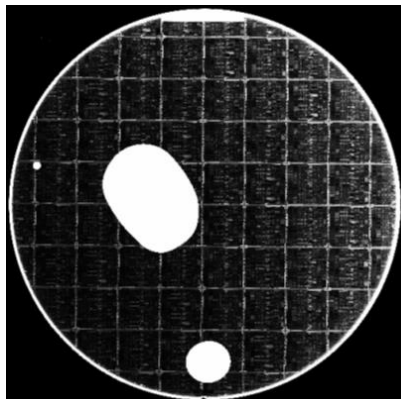


Fig. 90: CMOS-SOI wafers showing bonding defects: large unbonded surfaces are clearly visible through the SAM.

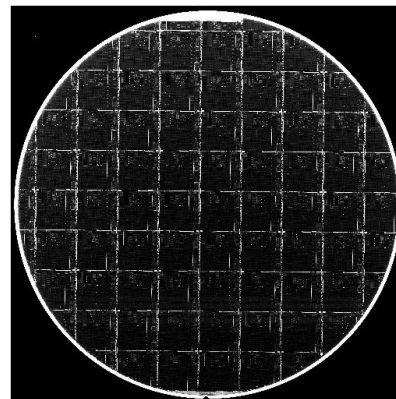


Fig. 91: CMOS-SOI wafers showing successful bonding.

As the top wafer is thinned down a sort of “knife edge” shape can appear on the top wafer edge. Additionally, after grinding and oxide etching the wafer edge becomes fragile, increasing the risk of edge delamination. During cleaning and etching steps, the presence of edge delamination could cause copper exposure and diffusion. Therefore, to eliminate the knife-edge shape and the edge delamination a trimming step is done. However, the trimming step leaves some exposed copper on the sidewalls. In this batch, we performed a final trimming of $5\mu\text{m}$ thickness and 8mm width.

4.2.4 NEMS PATTERNING

At this point of the flow, the top layer is the c-Si film in which the NEMS will be patterned. An alignment strategy must be implemented to connect the NEM relays to the underlying CMOS circuits (Fig. 92). Scribe line primary marks (SPM) are defined on M4 for this purpose. A first cycle of lithography and etching is done under coarse alignment on the wafer notch to clear the marks and facilitate later use.

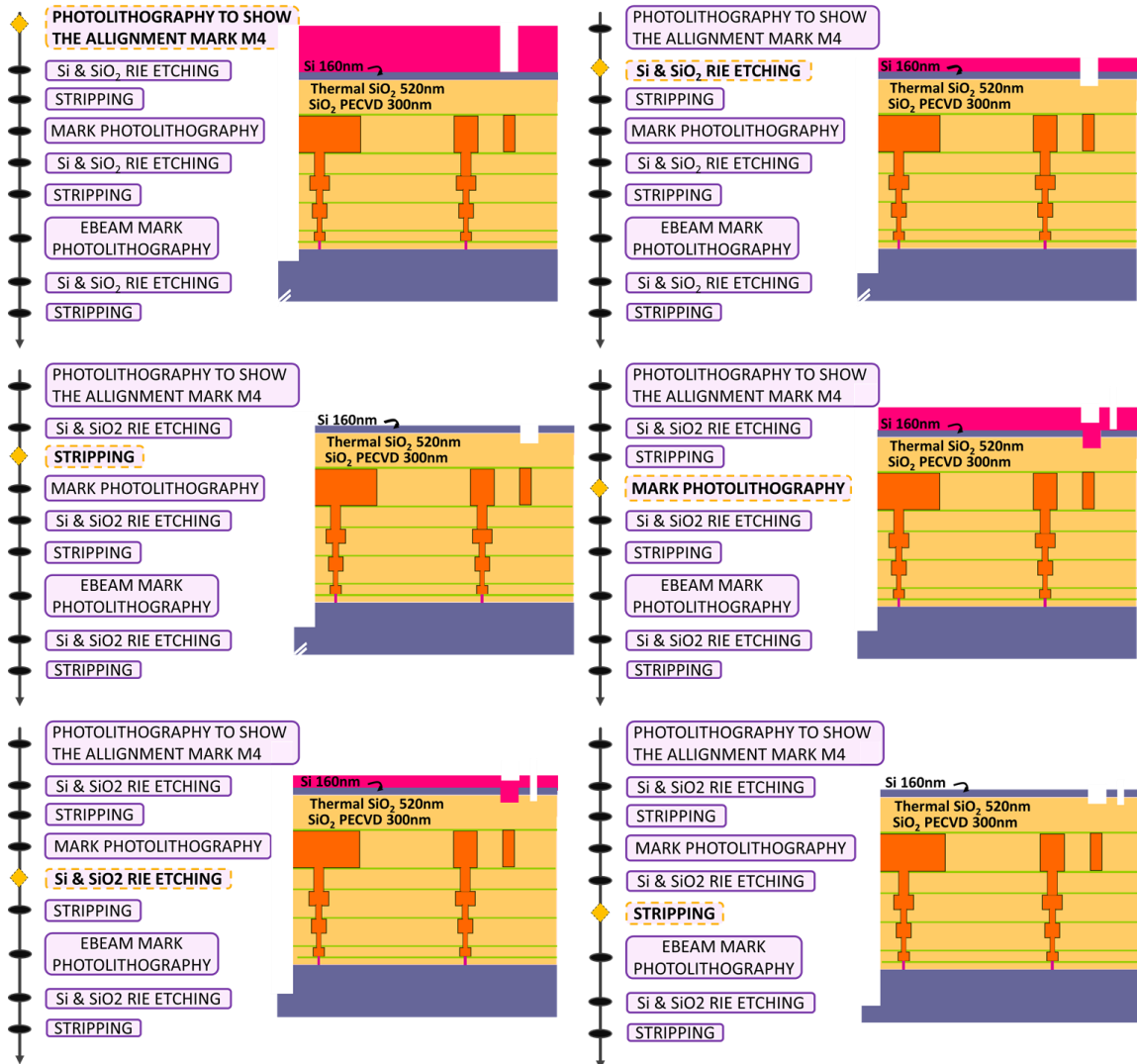


Fig. 92: Cross-sectional schematic view of the lithography alignment marks technology process step. The NEMS patterning will be made by a hybrid DUV/EBEAM lithography, one aligned on the M4 of the CMOS underlying and the other aligned on the EBEAM tool related special marks which themselves are aligned on M4.

The NEMS will be defined using hybrid DUV/EBEAM lithography technique. The DUV lithography is aligned on an intermediate level of marks called mark 99 see Fig. 92. These marks are defined by lithography steps, which are aligned on M4. After the alignment mark definition, the wafer sometimes requires some post-

processing steps such as oxide deposition or CMP, resulting on changing the marks appearance or compromising their visibility. Thus, it is often necessary to uncover them using an additional lithography step. The EBEAM lithography tool needs special designed EBEAM marks. The EBEAM marks are lithography defined and aligned on the marks 99. In this way the EBEAM defined level will be aligned on the DUV level and on the M4 level. The EBEAM marks are located on the dies periphery zone and have a size of 10 μ m x 10 μ m and are 800nm deep. The EBEAM tool uses an electronic scanning to read the alignment mark, meaning that a certain contrast and topology is needed to be able to read the signal amidst the electronic noise [WEIDENMUELLER 2008].

Since the wafers have the sides trimmed and the Cu on the sidewall is already exposed, the stripping, especially the wet step, needs to be performed with a recipe that is less aggressive and Cu compatible in order to avoid risk of Cu contamination.

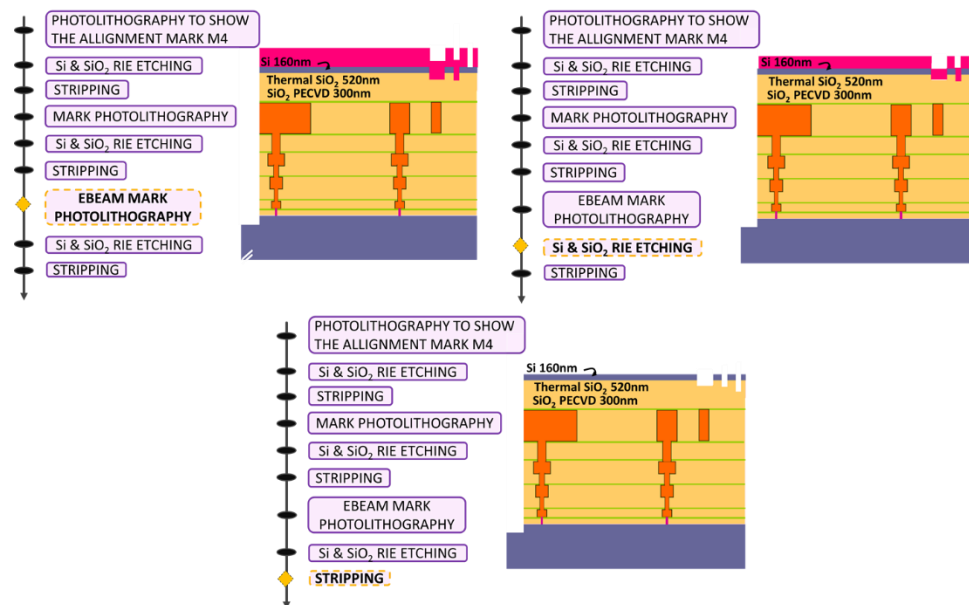


Fig. 93: Cross-sectional schematic view of the EBEAM lithography alignment marks technology process step.

Once all the alignment marks are defined, the patterning of the NEMS can be done, see Fig. 94 for the corresponding process steps. As said before, this layer is patterned using hybrid lithography steps combining DUV lithography with minimum width of ~350nm and EBEAM lithography having minimum width of 50nm and minimum space CD of 50nm. The alignment between the NEMS relay and the M4 is critical for this integration steps. Moreover, the alignment between the DUV and the EBEAM needs to respect the design specifications. In our mask, it was decided to define by EBEAM the NEMS electrodes (actuation electrode, contact electrodes and movable electrodes) and to leave to DUV lithography the c-Si lines and Pads. The hybrid patterning is double exposed and single etched. The

use of the negative resistance allows for saving EBEAM writing time. The resist thickness is chosen such that it can withstand the etching 160nm of c-Si.

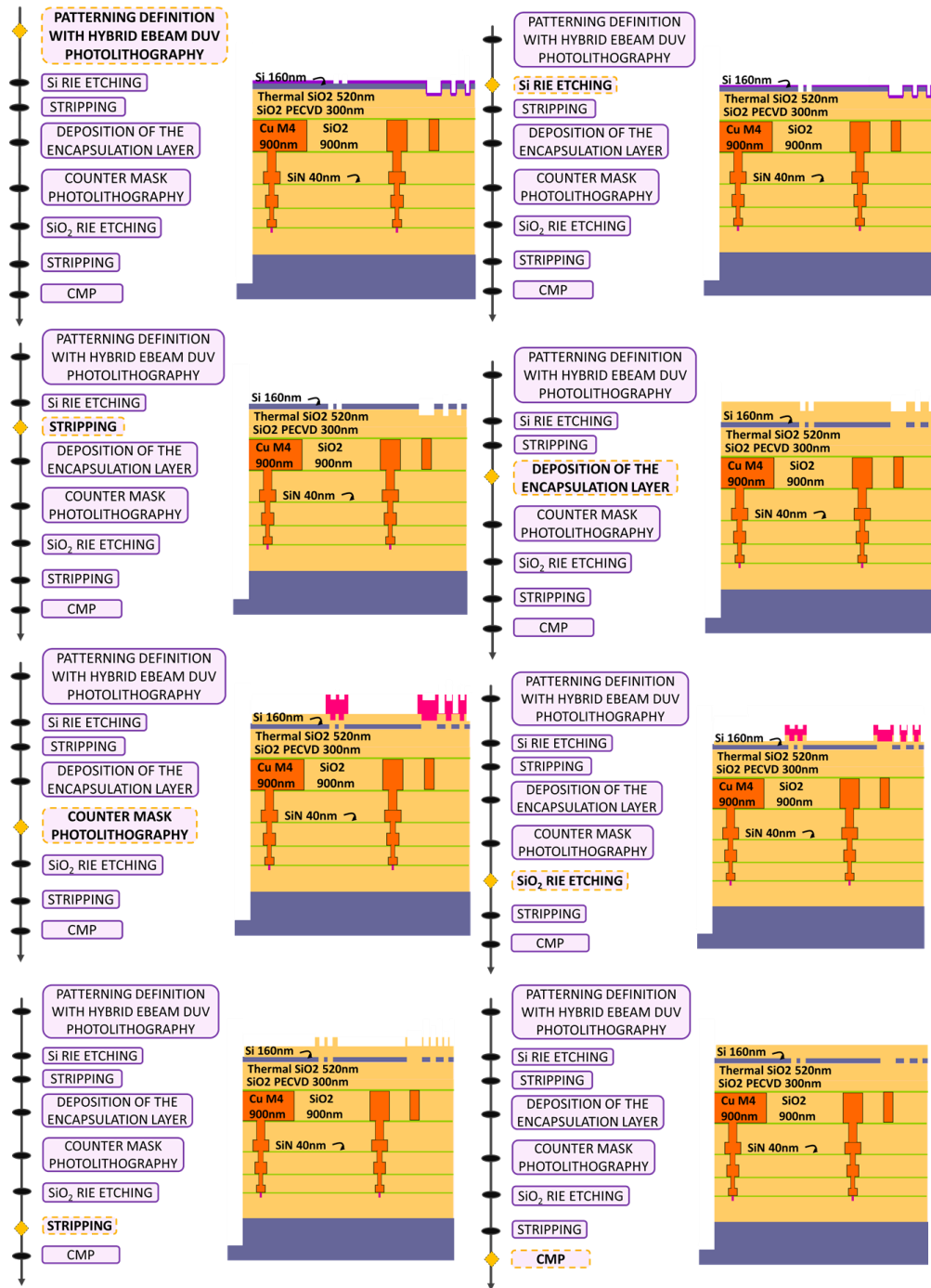


Fig. 94: Cross-sectional schematic view of the NEMS hybrid DUV/EBEAM patterning aligned on the CMOS circuitry underlying. The NEMS is then encapsulated by an oxide layer.

The DUV and the EBEAM are overlapped by ~100nm for misalignment tolerance. Lengths of movable structures defined in this mask range from 5μm to 20μm and widths go from 50nm to 400nm. Spaces between the movable structures and the contact electrodes range between 50nm to 80nm. The fabricated structures are

shown in SEM top view in Fig. 95, Fig. 96 and a SEM cross-sectional view in Fig. 97.

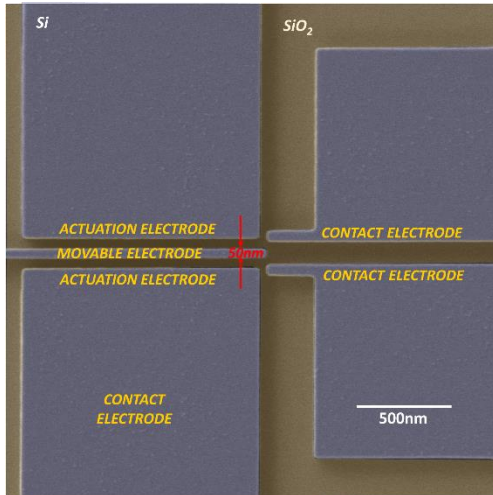


Fig. 95: Top SEM view of Cantilever type relay

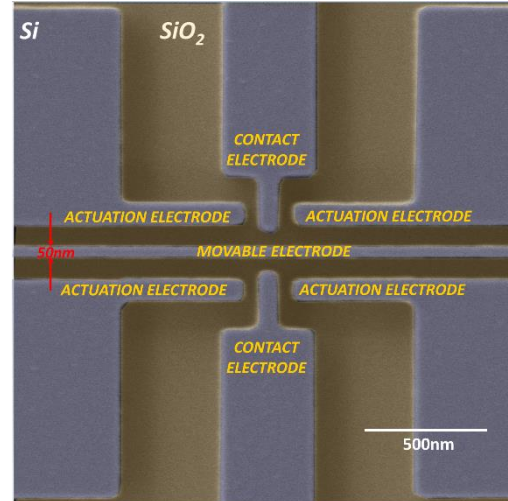


Fig. 96: Top SEM view of Clamped-Clamped type relay

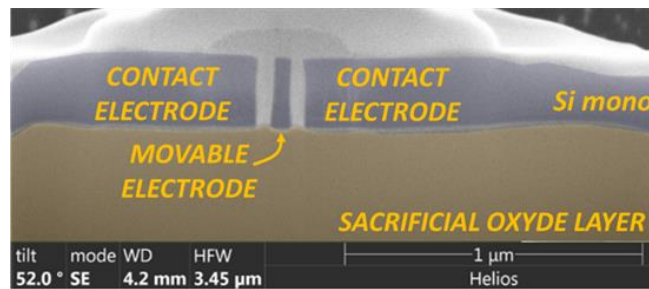


Fig. 97: Cross sectional tilted (52°) SEM view of the movable electrode between two contacts in a Clamped-Clamped Si relay.

After NEMS patterning, 400nm of PECVD silane-based Oxide (SiO_2) are deposited as a passivation layer. The deposition thickness is given by the already developed technology by [PHILIPPE 2014]. The role of this layer is to protect the NEMS and to allow for the fabrication of the metal interconnections between the NEMS and the CMOS and to the pads. The choice of this oxide is justified by two reasons: first, it is necessary to deposit a low temperature oxide, second, even if silane based SiO_2 produces many residues during the oxide HF etching, they can be easily removed. The oxide used does not completely fill the small gaps between the contact electrodes and the movable one. A more conformal oxide that can be deposited at low temperature is e.g. Plasma Enhanced Atomic Layer Deposition (PEALD) SiO_2 . However, it needs to be deposited in several passes to reach the required 400nm thickness. Depositing the oxide in several steps can take long time. Using a thin layer of SiO_2 ALD to fill the voids, and another oxide to complete the total thickness would cause interface degradations during future annealing. After the oxide deposition the top surface presents a surface topology, which will be

problematic during interconnection fabrication. The layer is flattened by a chemical mechanical polishing step, CMP. Before the CMP step counter mask lithography and etching steps are performed. The counter mask allows for etching the topology created by the NEMS patterning, thereby facilitating the CMP.

4.2.5 INTERCONNECTIONS AND PADS DEFINITION

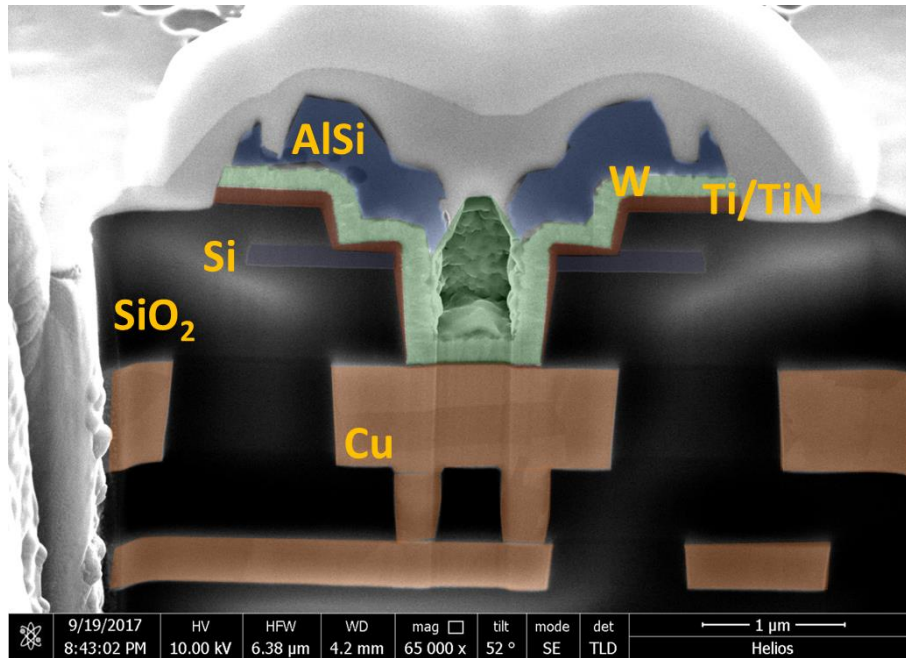


Fig. 98: Cross sectional tilted SEM view of metals interconnection between NEMS defined in the Si layer and M4 of the CMOS wafer.

After the encapsulation layer is planarized the VIAs, routings and pads are patterned. Fig. 99 shows the process flow to open the area in which the metal is deposited to make the routing, the pads and the interconnection between the c-Si NEMS and the M4.

The first DUV Lithography, anisotropic etch and stripping iteration is done to open the encapsulation oxide ($\sim 400\text{nm}$) in the areas that will be occupied by the pads, the routings and the VIAs. A second DUV lithography is performed to open the VIAs. The lithography is aligned on the SPM created during the c-Si etching, which are aligned to M4. The VIAs needs to be aligned to the NEMS but also to the corresponding CMOS metal via. The resist needs to withstand the etching of a total stack of $\sim 2\mu\text{m}$ thickness. The etching is divided into 3 steps: the first is a dry etching of the c-Si, the second etch it is also dry and it removes all the oxide stack (thermal and TEOS) and the last etch removes the SiN passivation layer. Once the SiN is etched, the copper is exposed to air and it tends to oxidize very quickly. Therefore, the metal barrier deposition needs to be deposited after a reactive pre-clean of the copper and within a few hours from the SiN etch.

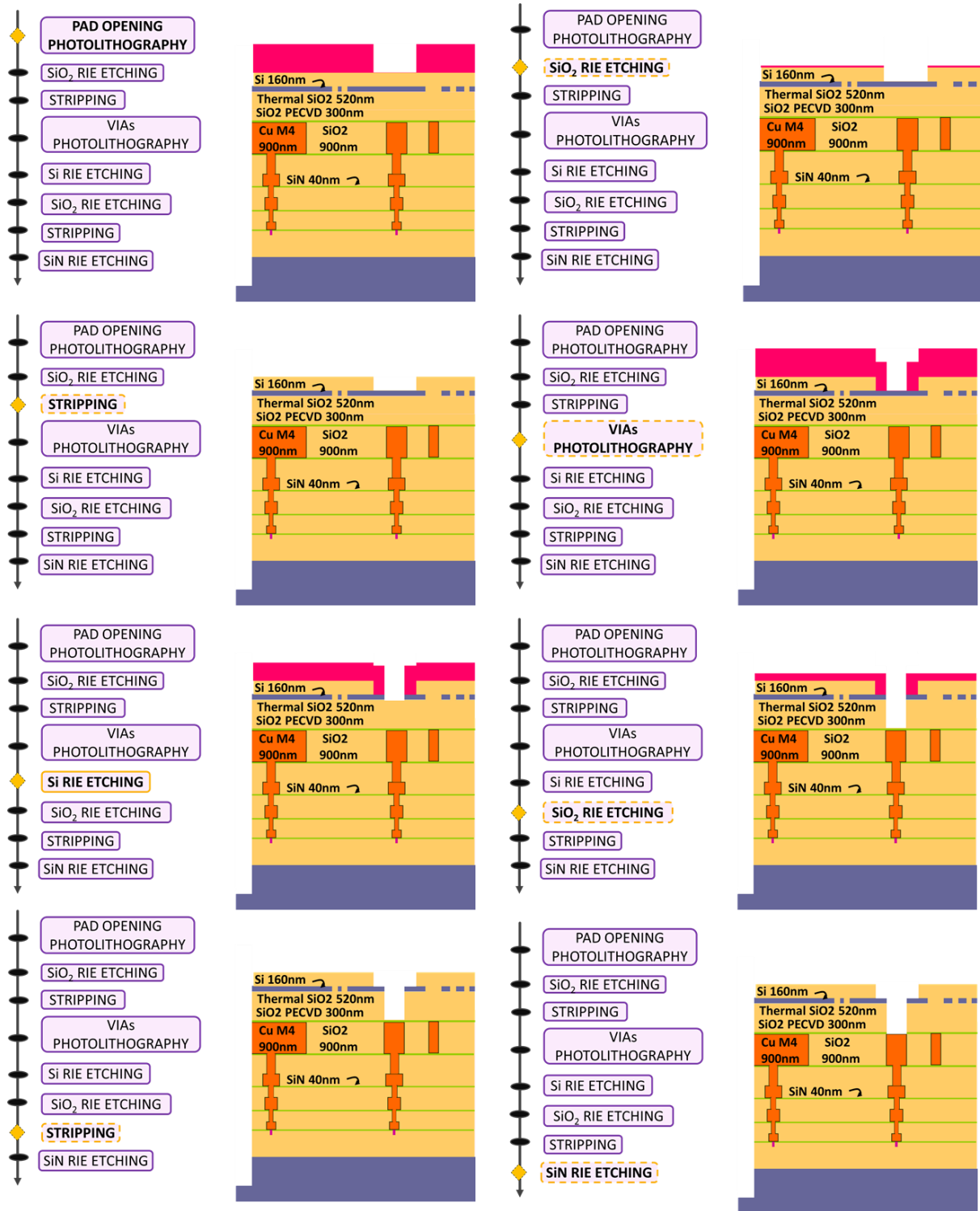


Fig. 99: Cross-sectional schematic view of SOI-CMOS interconnection etching. The etching is done in multiple steps: First the encapsulation oxide etch, then the mono c-Si etch, then the bonding oxide and eventually the metal encapsulation SiN etch.

The metal used to contact the BEOL metal and the NEMS part has to withstand several characteristics. First it has to be deposited at low temperature (below

450°C), second it has to be inert to HF vapor during the release step. Moreover, it has to be deposited conformal to fill the VIAs, which are deep 1040nm and side of 2000nm. Here, as shown in Fig. 99 the metal barrier deposited to avoid copper diffusion is PVD Ti/TiN (110nm), the via are filled by 200nm of CVD W and the last layer, useful for the contact, especially to electrically test the circuit is a CVD deposition of 600nm of AlSi.

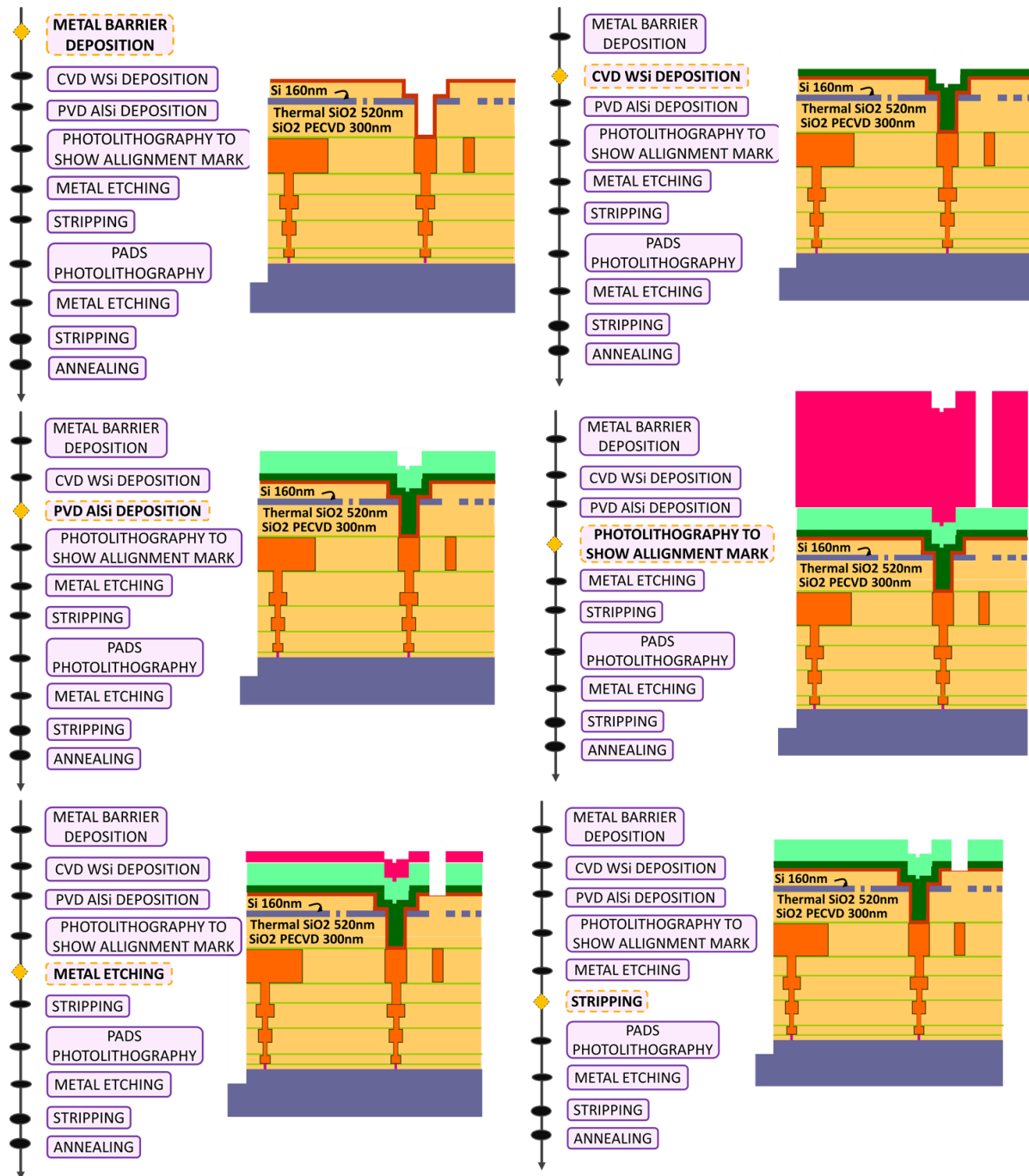


Fig. 100: Cross-sectional schematic view of the NEMS/CMOS interconnection metal deposition flow

After the Metal deposition, a metal stack of 910nm covers the scribe line primary (SPM) marks present on the c-Si layer and on the encapsulation layer. These marks allow for Metal pads alignment with the c-Si level. A cycle of lithography etching is done under coarse alignment on the wafer notch to clear the SPM marks. At this point of the flow, it is possible to pattern the interconnections and the pads as shown in Fig. 101. The lithography is aligned to the SPM marks made below the encapsulation layer. The etching is done in the same tool but with different chemistry to etch first the AlSi, then the W and at the end the TiTiN. It is important to produce enough polymers to protect the side of the AlSi during the W and TiTiN etching in order to keep a straight profile.

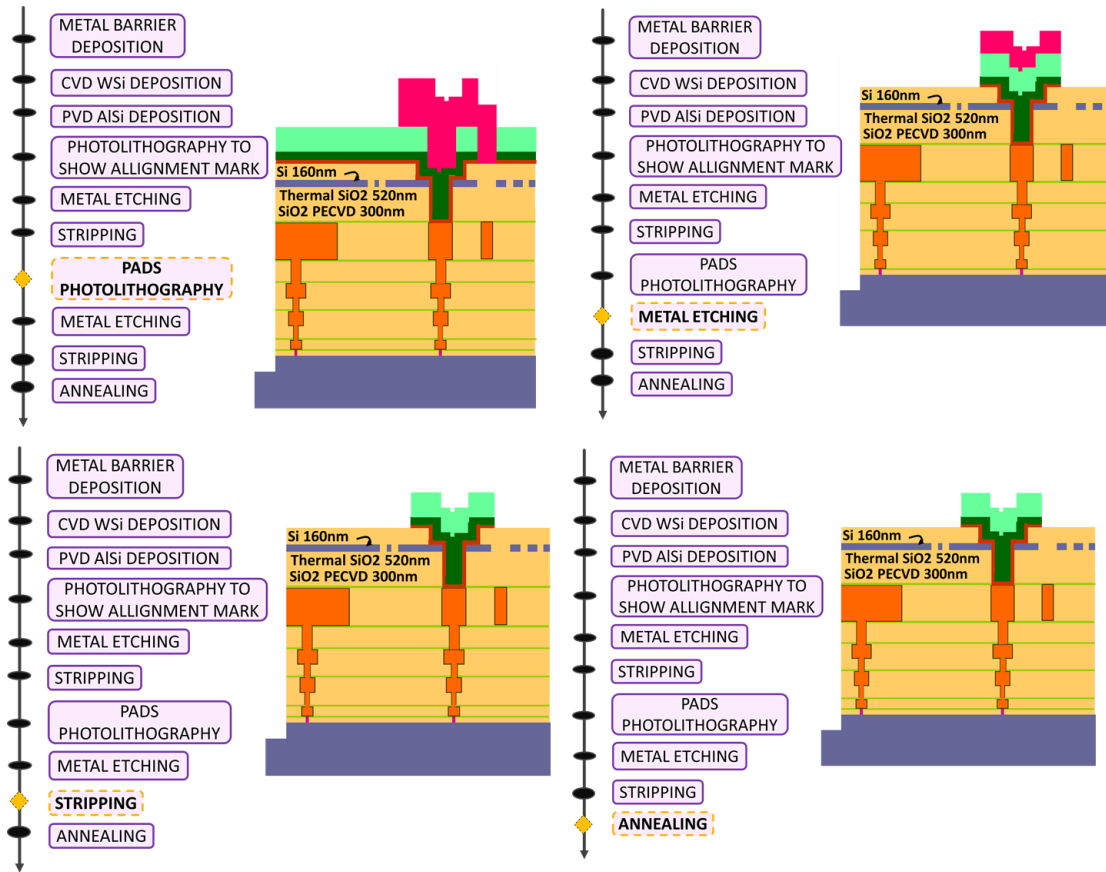


Fig. 101: Cross-sectional schematic view of the pads and interconnection patterning

The fabricated flow shown in Fig. 102, the NEMS are fabricated above the CMOS, with a compatible CMOS process and they are connected to the circuitry underlying in order to enhance the CMOS functionality.

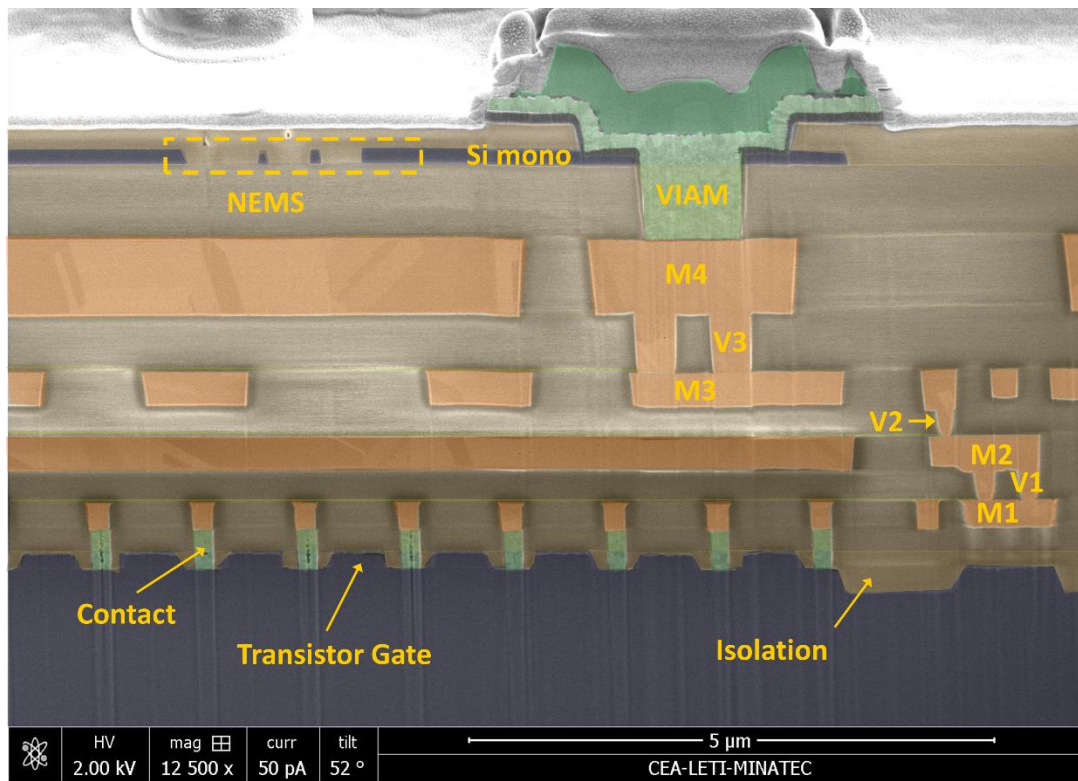


Fig. 102: Cross sectional tilted SEM view of the entire stack: AlSi, W, TiTiN metals interconnections connecting the Si NEMS to the Cu M4 of the CMOS underlying circuitry.

4.2.6 NEMS RELEASE

Finally, the NEMS movable part release is the last step of the fabrication process. It consists of etching the encapsulation layer that covers the top of the NEMS and the oxide surrounding the movable electrode, the actuation electrode and the contact electrode without damaging the rest of the relay. For Si-based structures, the common release etchant is HF because it has a high selectivity to SiO_2 with respect Si.

The reaction between HF and SiO_2 is well described in [KANG 2001]. In summary, H_2O or ethanol are catalyst and products of the reaction between HF^+ and SiO_2 . HF captures a proton from the catalyst and reacts with SiO_2 etching it and producing more water (or ethanol). However, NEMS structures are more subjects with respect MEMS structures to be deflected under capillary forces and encountering stiction.

Liquid HF etching is more likely to generate stiction issues because the liquid could be trapped under the structure and, during liquid draining, capillary forces could stick the structure on the substrate or in the electrode. The capillary forces can be stronger than the spring restoring force of the NEMS movable beam. The HF etching using its vapor phase can partially solve the problem. The key point on developing a stiction-free reactions is the good management of the desorption of

the byproducts (namely water and silicon tetrafluoride SiF_4) that can be achieved by controlling (and lowering) the partial pressure of the etchant and the chamber pressure see [POLLET 2013] for reference.

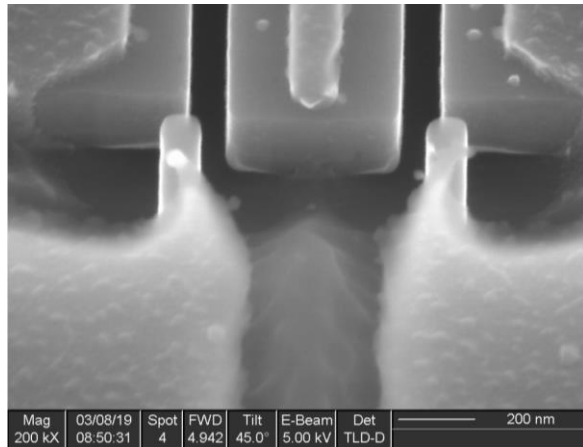


Fig. 103: SEM Cross View of 5 terminals cantilever like relay after the pre-release step.

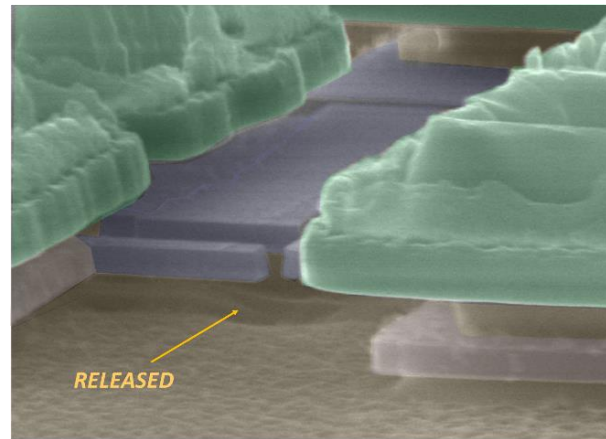


Fig. 104: SEM Top tilted view of 5 terminals cantilever like relay showing the oxide undercut after the release step.

Different oxides have a different amount of residues after the HF vapor etch. In our case, the encapsulation is done with silane based SiO_2 that after being expose to HF vapor leaves some residues see Fig. 105. Nevertheless rinsing the structure and drying it at 120°C for one hour allows to get rid of those residues Fig. 106. However, the rinsing step can cause stiction of the movable electrode onto fixed electrodes or onto the substrate if the movable structure is already released. That is why the release should be divided in more than one step.

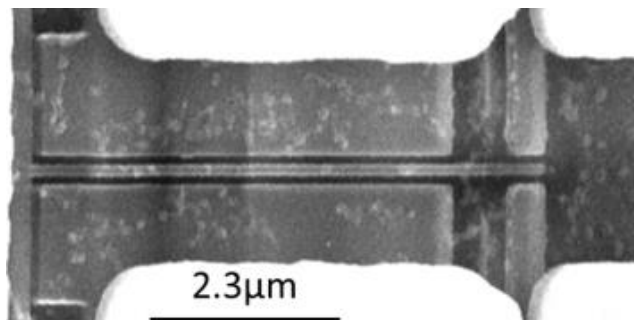


Fig. 105: SEM Top view of released relay showing the residues formed during VHF etching

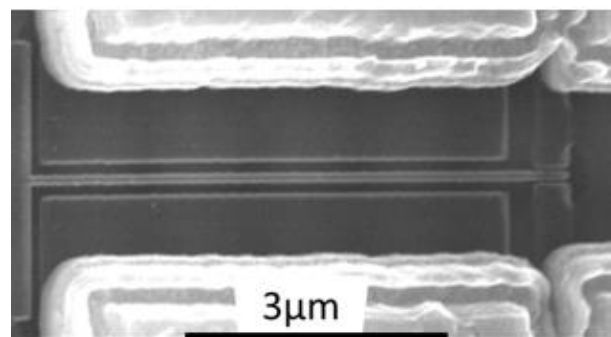


Fig. 106: SEM Top view of released relay showing the results after rinsing and drying

The first step, called pre-release, requires etching the encapsulation layer ($\sim 400\text{nm}$). It is important to notice that the oxide-etching rate under HF vapor depends on the type of precursor used for the oxide deposition method, and also on the deposition method: for instance, silane-based SiO_2 etches 10 times faster than thermal SiO_2 . Moreover, the oxide used to encapsulate the relay leaves voids between the contact

electrodes and the movable electrodes (gap~50nm). Those voids influence the prerelease steps risking to completely releasing some parts of the structure during the pre-release steps. Once the prerelease-etching step is done the sample can be rinsed and dried without risk of stiction. The second release step does not involve a cleaning steps but only an etching under HF vapor. In fact, the thermal oxide does not produce many residues during the HF etching.

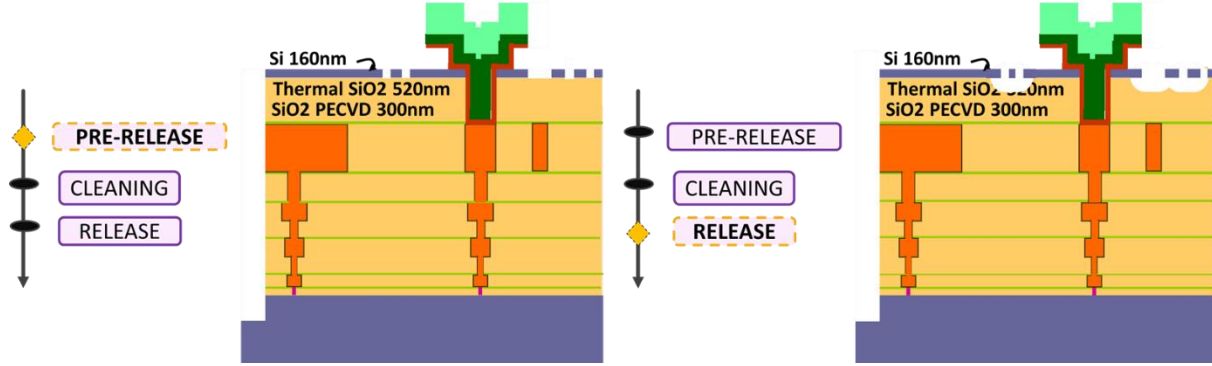


Fig. 107: Cross-sectional schematic view of NEMS above CMOS circuitry showing the release steps.

4.3 FABRICATED DEVICES

Fig. 108 shows the mask set of the fabricated batch for the scope of this thesis. Two main circuits are realized: the NV SRAM and the TCAM see section 2.3.4 for reference. For each NV SRAM there is a non-volatile switch that works at V_{DD} given by the CMOS circuitries. Next to each NV SRAM there is also a stand-alone switch for debugging purpose. Both clamped-clamped beam and cantilever beam structure were used for NV SRAM and TCAM. In the batch are included stand-alone back end of the line switches. The target application for BEOL switches is NV memory for data searching [KATO 2016B], meaning that the functional validation relies on the criteria of non-volatility and reprogrammability.

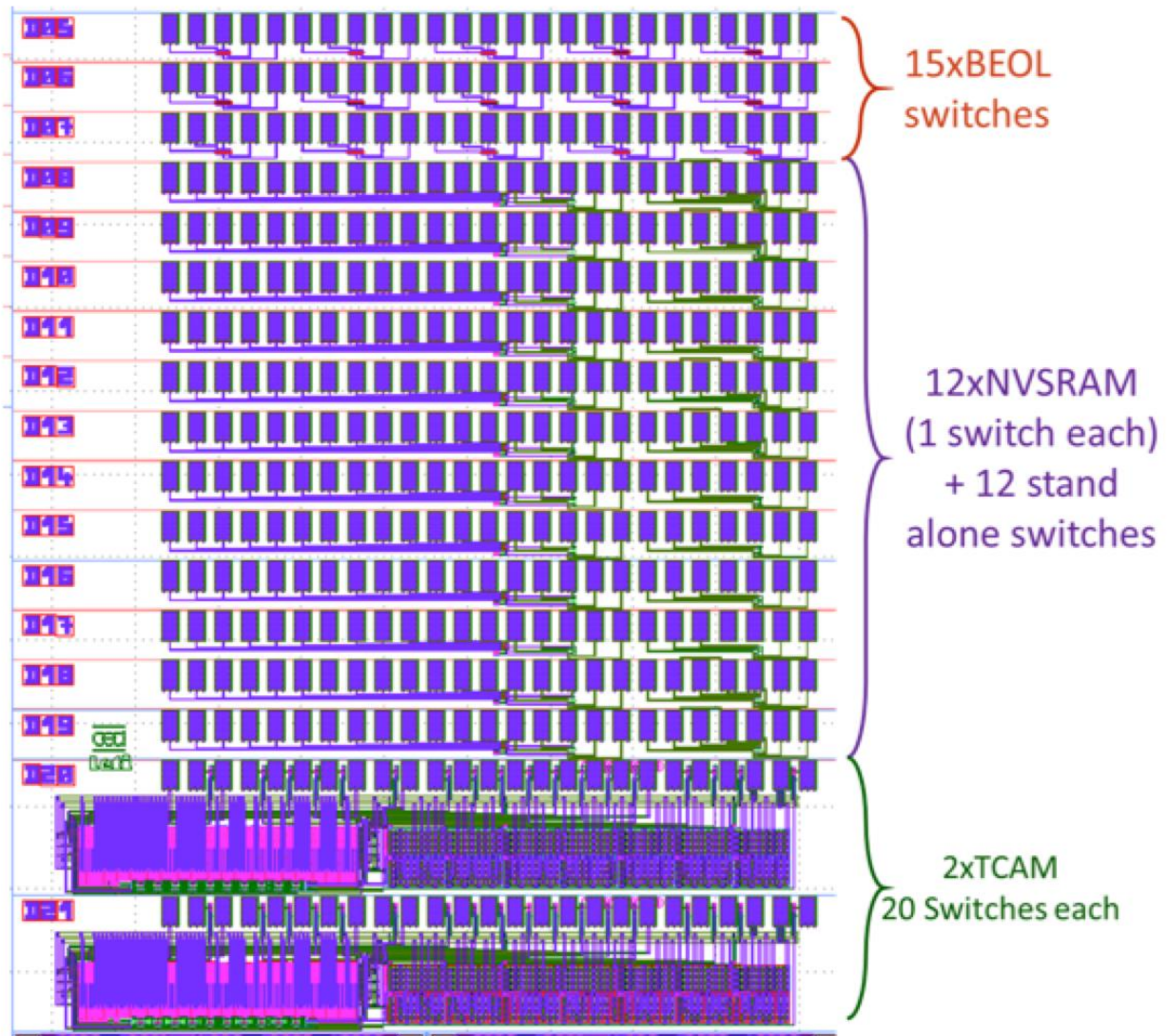


Fig. 108:Mask Set of the designed and fabricated batch during the work of thesis

The c-Si approach allows to define highly doped c-Si NEMS above CMOS without misalignment and almost no interconnections parasitic. However, there is still room for improvement, especially regarding the planarization of the CMOS wafer. If the surface planarization does not fulfill the wafer bonding requirements, side delamination can occur augmenting the risk of copper contaminations. Additionally, the silane based SiO_2 used is not conformal for gaps smaller than 70nm, causing to the release process inhomogeneity.

4.4 I/V CHARACTERIZATION OF MONOCRYSTALLINE STAND-ALONE NEMS RELAY

With the fabrication process described above we fabricated and characterized standalone 5T NV NEM relays. During this projected, we fabricated various wafers in order to overcome the incertitude of the release steps.

The switching behavior of the fabricated and released 5T SPDT monocrystalline NEM relay, monolithically fabricated above a CMOS wafer, is evaluated using an Agilent B1500A semiconductor analyzer equipped with two high-resolution source meter units (HRES-SMU) and three standard SMU.

In the first wafers, the gate voltage is swept from 0V up to a voltage level $V_{DD} > V_{PI}$ and then swept back to 0V. The two contact electrodes voltages $V_{DL} = V_{DR} = 20V$ and the source voltage V_S to 10mV. The gate current I_{GL} , the source current I_B and the drain current I_{DL} and I_{DR} has been recorded. For all the measurements, a current compliant of 500 μ A is chosen in order to reduce the contact degradation. The wafers are exposed to air ambient. Fig. 109 shows source and drain and gate left current and v_s the gate left voltage sweep. The drain voltage is at 20V in order to augment the probability to detect current flowing between drain and source. The measured current value is $I_B = I_{DL} > 1nA$ once it pulls in at 2.45V giving a $R_{ON} = 20G\Omega$. The relay is nonvolatile since it stays in contact when no voltage is applied. However, when the gate voltage is swept from 5V to 0V there is a snap of the beam in the gate at 1V. The beam starts to contact both drain and gate.

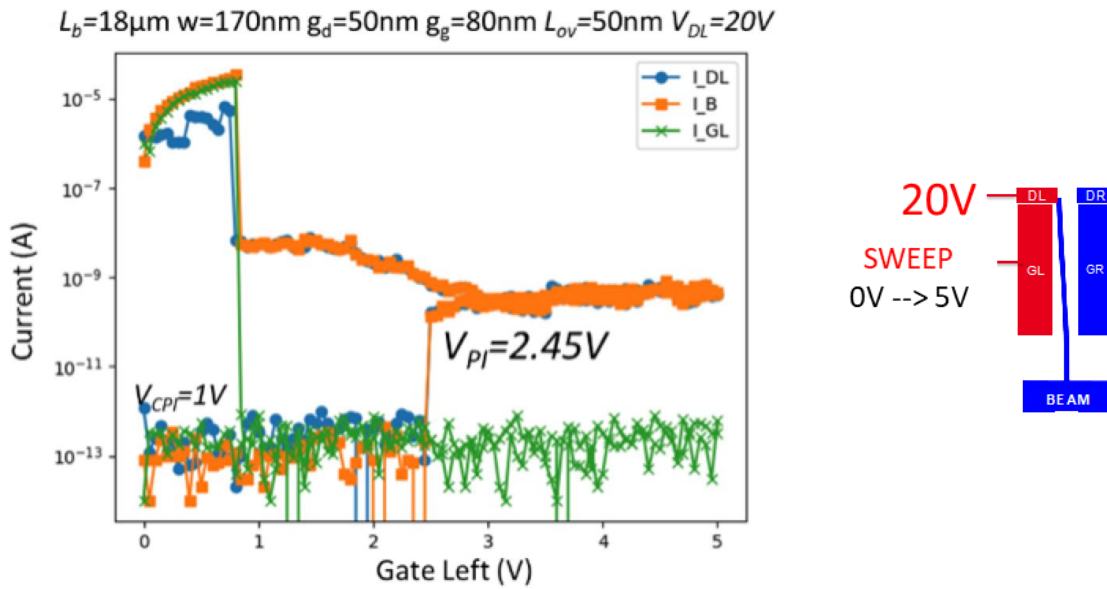


Fig. 109: Measured $I-V_{GR}$ characteristics of the programming operation of a 5-Terminal monocrystalline-silicon relay monolithically fabricated above a CMOS wafer. Showing zero off-state leakage (only noise floor) and abrupt switching behavior at 2.45V.

The Drain I_{ON}/I_{OFF} ratio is $> 10^4$, considering I_{OFF} as the instrumentation noise floor. Some electrical test performed in other relay designs showed the presence of current flowing between the beam and one or two drain and or gate electrodes since the beginning of the gate voltage sweep. It means that the beam had already pulled into the drain, into the gate or both. Some other test results showed no current flowing through the beam. Various factors can be the cause of those results. The beam stiction through one electrode before actuation could have being caused by the VHF release step or by SEM observation. The absence of current can be explained by the fact that the beam had secondary anchoring points formed during the release steps or formed during the SEM observation or because of lithography/etching fabrication failure.

The second wafer had longer VHF released process; the modified process allowed obtaining interesting results. In fact, we were able to switch the relay to the left drain electrode and then to the right electrode. In the first cycle, the gate left voltage (V_{GL}) is swept from 0V up to a voltage level $V_{DD} > V_{PI}$ ($V_{DD}=9V$) and then swept back to 0V. In the first cycle (programming), there was not potential difference between the left drain electrode and the beam in order to not damage the contact. In the second cycle (re-programming), the gate right voltage (V_{GR}) is swept from 0V up to 9V and then swept back to 0V. The drain left voltage is fixed at 20V in order to detect the beam/drain pull-in. The gate current I_{GR} , and the drain current I_{DR} has been recorded. For all the measurements, a current compliant of 500 μ A is chosen in order to reduce the contact degradation.

Fig. 110 shows the drain right current and gate right current vs the gate right voltage sweep. The current value is $I_{DR} > 5nA$ once it pulls in at 2.36V. The relay is nonvolatile since it stays in contact when no voltage is applied. The beam starts to contact both drain and gate at $V_{CPI}=6.45V$ giving a large $V_{CPI}-V_{PI}$ margin (4V). It is possible to observe a hysteresis on the gate right showing that there is not permanent catastrophic pull in. The drain I_{ON}/I_{OFF} ratio is $> 10^4$ considering as I_{OFF} the instrumentation noise floor. This particular relay is designed to work at 1.8V. Therefore, the experimental V_{PI} is 31% larger than the simulated one. The contact resistance of this relay is 4G Ω for an applied drain voltage of 20V. Paragraph 3.5.2.1 shows that the designed SRAM can be re-written if NV element on resistance is below $R_R < 30G\Omega$. The measured resistance is way above the nominal calculated resistance using the Sharvin resistance formula (see 3.4), highlighting the difficulty in estimating the effective contact area.

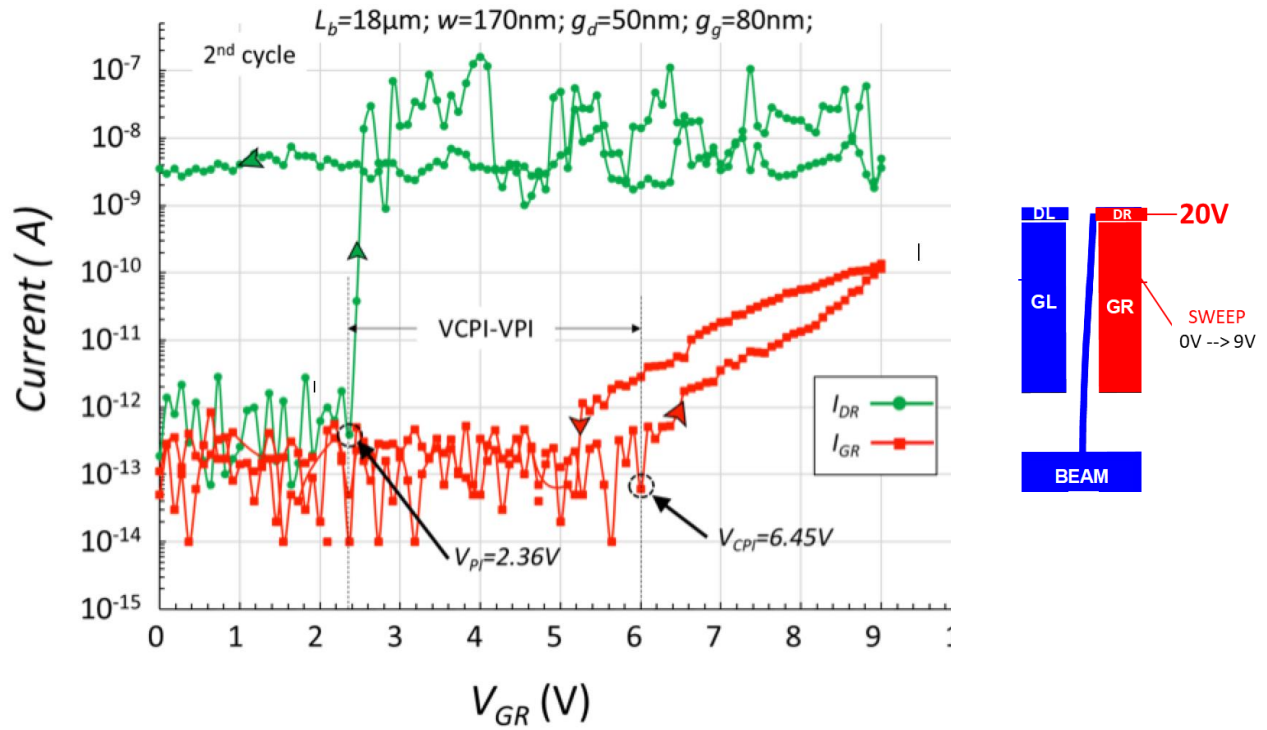


Fig. 110: Measured $I_{D/G}-V_{GR}$ characteristics of the programming operation of a 5-Terminal monocrystalline-silicon relay monolithically fabricated above a CMOS wafer. Showing zero off-state leakage (only noise floor) and abrupt switching behavior at 2.36V and a not-permanent catastrophic pull in at 6.45V

The third wafer had the same release process of the second wafer but with an additional dry-cleaning step. In this third wafer we were able to perform the programming and re-programming operations by switch the beam first to the left drain electrode and then to the right drain electrode. In the first cycle, the gate left voltage (V_{GL}) is swept from 0V up to a voltage level $V_{DD} > V_{PI}$ (9V) and then swept back to 0V. The left drain is biased at 10V in order to detect the beam/drain pull-in. In the second cycle (re-programming), the right gate voltage (V_{GR}) is swept from 0V up to 9V and then swept back to 0V. The right drain is biased at 10V. The gate current I_{GL} and I_{GR} , the source current I_B and the drain current I_{DL} and I_{DR} has been recorded. For all the measurements, a current compliant of 500 μ A is chosen in order to reduce the contact degradation.

Fig. 111 shows the I-V measured characteristic for the programing and reprogramming operations. On the top row, there is the programming operation toward the left side. The current value measured on the left drain is $I_B \sim 50pA$ once it pulls in at $\sim 3V$. The relay is nonvolatile since it stays in contact when no voltage is applied. The left drain I_{ON}/I_{OFF} ratio $> 10^3$ considering as I_{OFF} the instrumentation noise floor. In the bottom there is the re-programming operation toward the right side. The current value measured on the left drain is $I_B \sim 50pA$ once it pulls in at 3V. The relay is nonvolatile since it stays in contact when no voltage is applied. The left drain I_{ON}/I_{OFF} ratio $> 10^3$. This particular relay is designed to work at 4.8V.

Therefore, the experimental V_{PI} is smaller than the simulated one. The ON resistance of this relay is $\sim T\Omega$ for an applied drain voltage of 10V. For this relay the ON resistance seems too high to be able to re-write the SRAM with it.

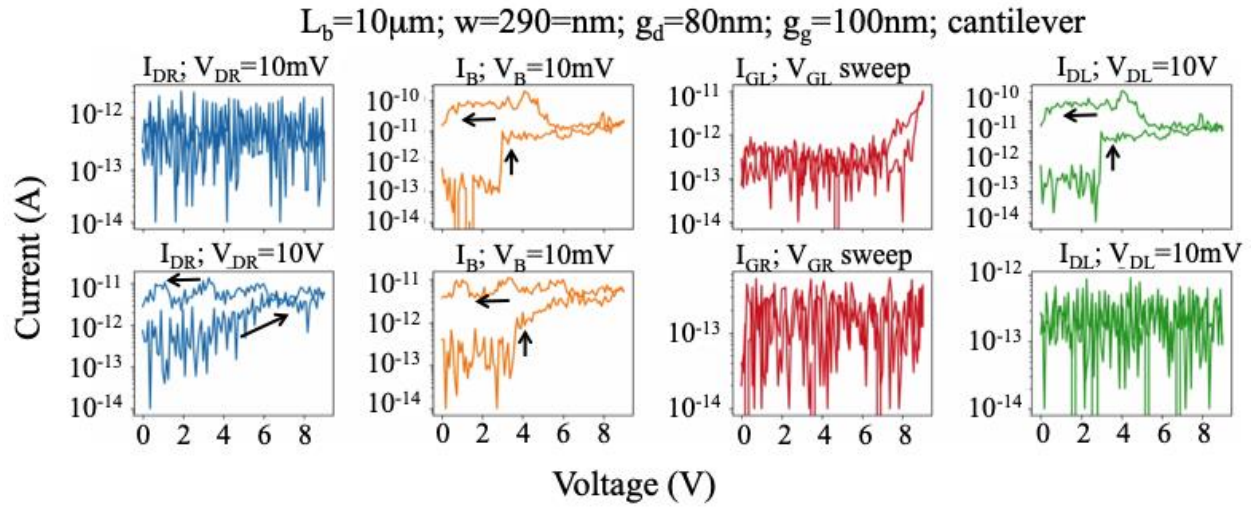


Fig. 111: Measured I - V characteristics of the programming (top) and reprogramming (bottom) operations of a 5-Terminal mono-crystalline-silicon relay monolithically fabricated above a CMOS wafer. Showing programmability, non-volatility and re-programmability.

4.5 CONCLUSION

Chapter 4 describes the process flow used to fabricate NEMS relay above IC. For NEMS relay co-integrated with CMOS, the monolithic approach is more interesting with respect hybrid not monolithic integration because of the lower packaging cost.

The presented process flow describes monolithically highly doped mono-crystalline Si NEMS above IC. The c-Si approach allows to define highly doped c-Si NEMS above CMOS without misalignment. However, it requires a certain manufacturing complexity and it can be an expensive process because of the presence of the SOI wafer, wafer bonding and grinding steps.

In the end of the chapter, the electrical results of the stand-alone mono-crystalline Si relay integrated above CMOS circuitry using M4 to connect the actuation electrodes for the clamped-clamped design are presented.

We demonstrated CMOS co-integrated NEM relays with the intended functionality: programming, non-volatility and reprogramming. The programming is done around 3V for the relay with a nominal V_{PI} of 4.8V. When not supply voltage is applied the beam maintains its position, showing non-volatility. For this design the reprogramming is done at $\sim 4V$ ($< V_{DD}$). For the relay design with a nominal V_{PI} of 1.8V the reprogramming is done at 2.36V ($> V_{DD}$). For both the design endurance is still limited at only 1 cycle of programming + reprogramming operations. The silicon-contacting surface easily wears out and or the beam relays breaks. Additionally, the contact resistance is very high from $\sim G\Omega$ (if drain biased at 20V) to $T\Omega$ (if drain biased at 10V). The high resistance value of highly doped silicon is due to the facts that in nano-contact only few asperities are actually contacting instead of the nominal contact area, and because of the native silicon oxide that is formed on the two contacting surfaces.

5 CONCLUSIONS AND PERSPECTIVES

The work presented in this thesis manuscript is focused on design, modeling, simulation and fabrication of in plane laterally actuated electrostatic NEM relay co-integrated above CMOS circuitry. NEM relay is currently lacking competitiveness in terms of performance, endurance, and feature size compared to MOSFETs. In this thesis, we focused on investigating suitable applications that will increase the energy efficiency of some circuits vs. a MOSFET-only alternative, by adding a NEM relays. We focused on demonstrating the feasibility of those suitable applications. Combining NEM with CMOS technology could become interesting for ultra-low power applications. The proof of concept performed during the work of this thesis relies on hybrid integration and in particular, the integration of c-Si NEMS above CMOS. In this thesis, the challenges of NV-NEMs relays using hybrid monolithic integration are discussed.

5.1 CONCLUSION

Chapter 1 gives a general overview of the status of MOSFET technology and Beyond CMOS devices. Once the transistor size reached the nanoscale, parasitic effects affecting the leakage current started appearing, causing performance degradation and device-to-device variability. Therefore, keeping extending Moore's law requires one to implement innovation strategies, new integration methods, and new computational methods or to find new device solutions. The MOSFET intrinsic limitation is that the subthreshold swing that cannot go below 60mV/dec. at room temperature.

The memory hierarchy has also been affected by the scaling race. Nowadays there is a large gap between the CPU access time and speed, and the out-of-CPU (*i.e.* memory) circuits. CPUs are very fast but need to spend time idling and waiting for the memory input/output transfer to be completed. Additionally, there is a large gap in speed and endurance between volatile (*e.g.* SRAM) and nonvolatile memory (*e.g.* Flash). As a result, novel memory candidates are under investigations.

In general, the beyond CMOS devices are explored with the intent of satisfying the needs of ultra-low power applications for logic and/or for memory.

From the many applications present in literature, we decided to focus on memories in computing applications, and especially on Non-Volatile relays as auxiliary element for CMOS circuitry since those applications have relaxed endurance and speed requirements.

NEM Relays are good candidates for in-computing memory because they have a hysteresis effect that can be controlled to obtain a nonvolatile element, they have a steep subthreshold slope, high R_{ON}/R_{OFF} ratio ($>10^3$), zero current leakage, and very low programming energy compared to the other emerging memories. Additionally, they can be co-integrated with CMOS circuits and they can be compatible with

monolithic 3D integration. NEM relays, compared to other emerging non-volatile memories should not need a selector, which could be interesting in term of footprint and lower cost per bit. Even if NEM switching speed is fundamental limited by the mechanical delay, this fact can be used as advantage in CMOS/NEM memory circuitry to increase the overall endurance. However, in order to progress and be competitive with current emerging memories technologies, efforts will have to be directed towards reducing the footprint and improving the endurance, without increasing the fabrication cost.

An extensive description of NEM relay devices and their state of the art is given in Chapter 2. The NEM relay design is chosen depending on the target application, in our case we used a SPDT relay design with an in-plane actuation motion. In fact, two terminals are needed to encode states '0' or '1' (drains) in the beam position (source) and two additional terminals on each side of the movable electrode (gates) to provide a an electrostatic force to allow movement to the left or to the right. The in-plane actuated motion is easier to fabricate compared to the out-of-plane actuated relay design although there is an area penalty downside. To minimize the area consumption, it is possible to integrate the switch vertically, but as discussed in section 2.2.4 forming stacked air gap interconnects requires an increased level complexity.

Dealing with contacting surfaces at the nano-size requires considering the asperity of the mechanical contacts. The electrical contact between two rough surfaces is a complex physical interaction between thermo-mechanical deformation, current flow and heating between the contacting surfaces, which can only be determined probabilistically. This fact highlights the challenges of knowing the value of the interacting adhesions forces, and of the contact resistance, additionally those values are more likely to vary during cycling. Cycling modifies and degrades the contact surfaces, increasing the contact resistance and inducing voltage shifts.

That is why in Chapter 3, we developed a method for optimizing the relay pull-in operating margin. The pull-in working margin is defined by the difference between the catastrophic pull-in voltage and the pull-in voltage ($V_{CPI} - V_{PI}$). This method takes into account the gate length and position with respect the beam tips. The method is carried on using analytical modeling and it is validated through FEA. Reducing the gate length and shifting towards the tip increases the margin with no or little impact on the pull- in voltage. The optimization of this margin allows robustness of operations. Leveraging this trend helps in counteracting the dispersion and the drift of pull-in characteristic due to cycling, process variability and environmental factors. The relay can be designed with a pull-in working margin build to be optimized for the supply voltage given by the CMOS circuitry. In this chapter, we also carried out a design optimization for non-volatile NEM relays. In relays for reprogrammable memories, an additional source of uncertainty, dispersion and drift is the adhesion force. This study allows designing relays that

will be able to withstand the programming, reprogramming and holding operations. To obtain a reliable and robust relay design it is necessary to maximize the design window defined by the inequality which has the sum between the electrostatic force and the spring restoring force larger than the adhesion force which is larger than the spring restoring force. Broadening that window and centering it on F_{adh} can be carried out independently. However, the parameters enabling broadening this window needs to be the same that keep the overdrive factor p sufficiently low, as it is correlated to the risk of catastrophic pull-in. Therefore, a trade-off between adhesion and $V_{CPI}-V_{PI}$ margin needs to be considered.

Chapter 3 proceeds with a simulations study of the writing (programming/reprogramming) and re-writing operation of a NV-SRAM (1R-8T). The writing operation works in the same manners as an ordinary 6T-SRAM cell. However, since there is a NEM relay connected to the storage node two scenarios can occurs. i) If the time between two writing operation is larger than the system RC delay + NEM mechanical delay the beam follows the SN and pulls into the contact electrode. ii) If the time between two writing operation is smaller than the NEM mechanical delay + RC delay the relay does not move (in our case if it is below 300ns). An ultra-scaled relay is projected to have a delay around 10ns. If write operation occur within less than 300ns, the switch will not follow, increasing the relay life. However, enough time (larger than 300ns in our case) needs to be given before power gating to allow the relay to switch. The RC delay on the writing operation depends on the p/nMOS ON resistance, on the resistance of the resistive path towards the beam and on the beam resistance itself, and on the relay capacitance which is usually very small. The re-write operation requires having the SRAM disabled until a certain time t_{re} in which the signal coming from the relay is injected on the SRAM storage node. The resistive path through the relay plus the re-writing assistant transistor RC delay gives this waiting time t_{re} . Waiting t_{re} before powering back on the 6T-SRAM pre-charges the storage node allowing injecting the data even if the ON resistance of the relay is very large (within a certain limit). Compared to writing the SRAM through the access transistors, the delay associated to this re-write automatic sequence is governed by the t_{re} and the RC delay made by the resistive path through the relay and the re-writing assistance transistor. This RC delay has a higher value compared to the RC delay of the only access transistor connected to the storage node and is highly influenced by the relay ON resistance.

Chapter 4 describes the process flow used to fabricate NEMS relays above a CMOS IC. The monolithic approach is interesting for NEMS co-integrated with CMOS because of the lower packaging cost with respect to hybrid (*i.e.* non monolithic) integration. In this thesis, we fabricate the NEM relay above CMOS following three different approaches, however in this chapter only one approach is described. The two other approaches can be found in the annexes. The first process

flow describes the fabrication of a monolithically, highly doped mono-crystalline Si NEMS above an IC. Thanks to a project carried out here at Leti, we were able to co-integrate NEMS relays and NEMS sensors on the same die, both of them connected to the corresponding CMOS circuit. The c-Si approach allows one to fabricate a highly doped c-Si NEMS above CMOS without misalignment. However, it requires a certain manufacturing complexity and it can be an expensive process.

Finally, the chapter shows results of the electrical characterization of the fabricated SPDT NV NEM relays. We present the electrical results of the stand-alone relay integrated above CMOS circuitry and using M4 to connect the two-actuation electrodes for the clamped-clamped design.

The electrical results sometimes do not show any drain current; we think that the absence of pull in at the given V_{DD} or at list of current flowing from the beam to the drain it might be caused by one or a combination of the three following factors. i) The presence of secondary anchoring points that increase the movable electrode spring constant, hence V_{PI} . ii) High value of contact resistance that does not allow to sense current from the contacting electrode. iii) Residues coming from the etching of the SiO_2 (deposited by SiH_4 -based LPCVD) which could prevent the contact between movable and contact electrodes. Additionally, for the same wafer, we observed an electrical response variability of the same relay design situated in different dies. This confirmed that the release process is not uniform.

Despite the variability, we were able to demonstrate CMOS co-integrated NEM relays with the intended functionality: programming, non-volatility and reprogramming. The programming is done with no catastrophic pull-in at $V_{PI} \sim V_{DD}$. V_{PI} was found to be around 3V for the nominal 4.8V design. When no supply voltage is applied the beam maintains its position, showing non-volatility. The reprogramming is also done without catastrophic pull in at $V_{PI} \sim V_{DD}$ around 2.4V for the nominal 1.8V switch design and around 4V for the nominal 4.8V. Endurance is still limited because of permanent stiction to one electrode or to the substrate or even by beam fracture. Additionally, the silicon-contacting surface easily wears out. The contact resistance is very high from $\sim\text{G}\Omega$ to $1\text{T}\Omega$ depending on the switch design, due to the fact that in nano-contact only few asperities are actually contacting instead of the nominal contact area and because of the native silicon oxide that is formed on the two contacting surfaces. Reducing the contact surface roughness and adding a metal coating layer could be helpful to reduce the contact resistance. However, the addition of a coating layer comes at the expensive of a modified stress gradient for the released structure. Therefore, additionally studies need to be done.

In the annexes, the alternatively two process flows are presented, with the electrical results obtained for only one of the two approaches. The first process flow

describes a polysilicon NEM relay above CMOS obtained by the re-crystallization through laser annealing of low temperature deposited amorphous silicon. The fabrication cost is lower than the c-Si above CMOS co-integration process. However, especially for NEMS relay designed with a beam width value $\leq 50\text{nm}$, having grains smaller than the beam width, to reduce device/device variability, comes at the expenses of too high resistivity values ($>30\text{G}\Omega$). The third process flow describes lateral actuated NEM relay with a vertical structure made by more than two Cu metal layers of a CMOS wafers. The vertical definition should help decreasing the actuation voltage. In this approach NEMS layout and mechanical properties are constrained by BEOL technology and design rules.

5.2 PERSPECTIVES

Today it seems that there is not an application in stand-alone logic for which current volatile NEM relay proof-of-concept performance is good enough to replace low-power CMOS as the basic building block in the circuitry logic. However, NV NEM Relays used as auxiliary element for CMOS could help the circuit to be more energy-efficient and to add circuit functionality.

Compared to other emerging NV memories, NEMS relays still faces many challenges. Additionally, co-integrating it with CMOS circuits requires fab post-processing (for the BEOL relays) or a larger number of additional process steps (c-Si relays).

In our lab, a previous team was working on developing a route to fabricate c-Si NEM sensors above CMOS. We improved this route to demonstrate functional c-Si NV NEM Relays above CMOS.

The fabricated technology indeed allowed a proof-of concept for the feasibility of NEM Relay as NV elements for SRAM and TCAM. The studies on the design optimization margin resulted in consistent characteristics. This result is very important, because it is fundamental for the target application (i.e. NV SRAM and TCAM) that the beam does not pull into the gate, to not encode the opposite state.

The proposed technology is not yet the ideal solution. In fact, highly doped c-Si fabricated NEM relays have a very high contact resistance value and a poor contact endurance, compared to relays having metal as contact/structural material for instance. However, the latter are more complex to scale and/or to co-integrate with CMOS.

In the frame of the targeted applications, we did not find yet a NEM relay structure that has the combinations of the following characteristics: scalable, high endurance, low contact resistance, and that does not require adding a large number of photo-masks and/or a large number of fabrication dedicated modules.

However, in the future NEM relays and NEM sensors could be co-integrate in the same layer. This co-integration could open opportunities to new applications that are still to be explored such as radiation hardening sensing and memory, since they are promising candidate for harsh environment. But again, this will require first the demonstration of functional, reliable, scalable and with a higher endurance NEM relays.

6 ANNEX 1

6.1 POLY-Si NEMS ABOVE CMOS

The complexity of the 3D monolithic cointegration between NEMS and CMOS made by wafer bonding motivates us in looking into other integration options. Here at Leti, the alternately solution of using poly silicon at the place of c-Si has been studied by [OUERGI 2015A]. The motivation is to reduce the cost of the entire process. Using poly silicon will allow using a single wafer to develop the NEMS above CMOS fabrication. Avoiding having an SOI wafers and lowering the number of the total process steps allows lowering the total cost of the post-CMOS fabrication. The poly-Si NEMS process flow allows to get rid of 11 process steps and the cost of the SOI wafer Fig. 112. Additionally, it allows to get rid of complex step such as wafer bonding which are critical for the wafer manufacturing, since it introduces fragility on the wafer side and the consequent needs to trim the side might cause copper exposure on the sidewalls.

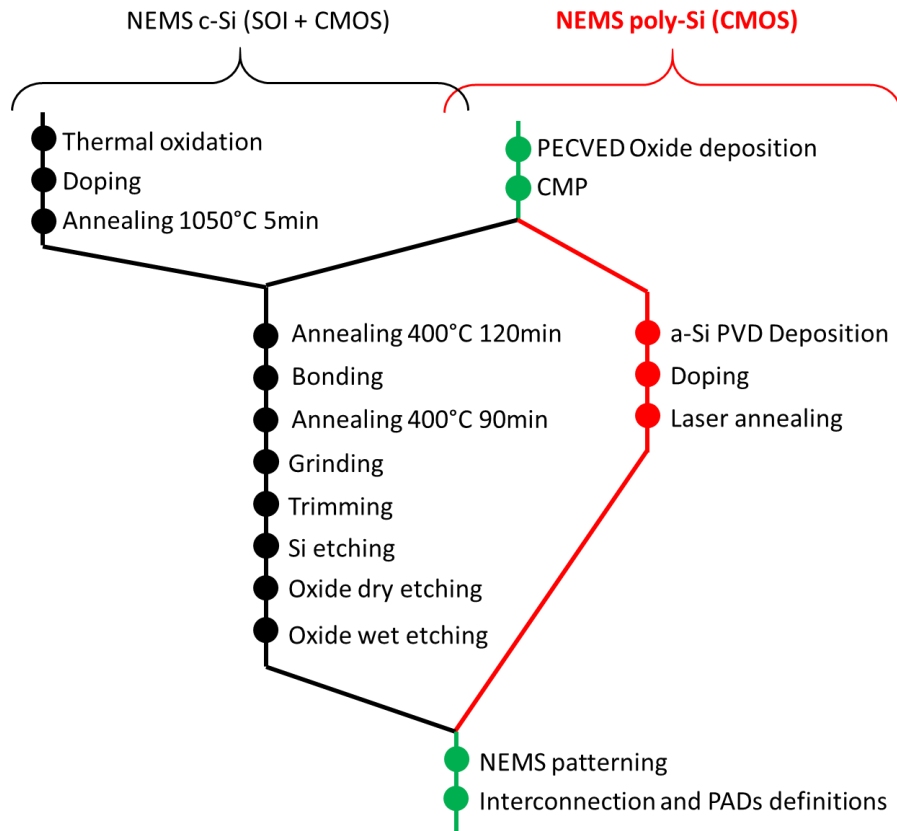


Fig. 112: Fabrication Flow of poly-Si NEMS and c-Si NEMS fabricated on SOI wafer or directly on the CMOS wafer for the 3D monolithic integration with the CMOS

Today, literature does not give a solution to obtain mono c-Si at low temperature besides wafer bonding, above the IC substrate. In fact, once the CMOS substrate is fabricated, the presence of the interconnections limits the allowed thermal budget, making impossible to grow crystalline silicon. However it is possible to fabricate

Poly-Si NEMS above CMOS, with poly-Si obtained by laser annealing of amorphous silicon. Fig. 113 shows the process flow to obtain poly crystalline-doped layer above the CMOS wafer. The CMOS interconnections are an encapsulated by 60nm of SiN. The first step is the deposition of 820nm of oxide. The oxide thickness has been kept the same with respect the oxide on the route explained on section 4.2 in order to avoid redeveloping the interconnections patterning but it could be limited by at the needs of the sacrificial layer thickness. The choice of oxide was made using a trade-off between deposition temperature and residues produced during the HF vapor release steps. Three different silicon oxide types were chosen all deposited by Plasma-enhanced chemical vapor deposition (PECVD): tetraethyl orthosilicate based (TEOS) based SiO_2 , tetraethyl orthosilicate low rate (TEOS LR) based SiO_2 , and silane based SiO_2 .

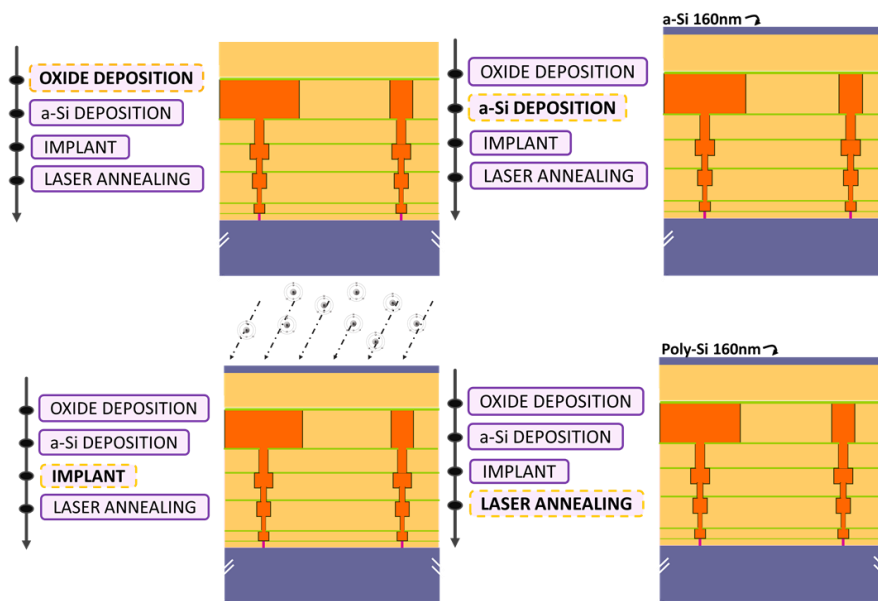


Fig. 113: Cross-sectional schematic view of poly-Si above CMOS process flow.

After the oxide deposition, there is the a-Si deposition. The a-Si is deposited at low temperature (below 400°C) with physical vapor deposition (PVD) method. The thickness deposited is 200nm in order to allow a slight CMP after the laser annealing to reduce the surface topology given by the poly-Si grain at minima. After the a-Si, a dose of $3 \times 10^{15} \text{ at/cm}^3$ of Boron is implanted at 30 keV in order to obtain after annealing a dopant concentration around 5×10^{19} . The last step is the laser annealing which both recrystallize the a-Si into poly-Si and activate the dopant. The laser pulse energy density value needs to be high enough to melt the a-Si, and low enough to not cause a-Si ablation and the pulse durations needs to include two a-Si melting peaks. The first peak will melt the a-Si and the second peak will allow the re-crystallization in grains. The laser pulse duration time chosen is 200ns. The a-Si melted already for an energy density value of 30 J/cm^2 for all the combination of oxides type. However, the second melting peak occurrence varied

in dependency of the oxide type used. For a given laser pulse period, the second melting peak came at 30J/cm^2 for the silane based SiO_2 , and at 35J/cm^2 for the TEOS LR and TEOS. It might be possible that the silane based SiO_2 isolates more the a-Si from the rest of the wafer and the wafer holder, allowing a faster temperature ramp. The sheet resistance decreases with the increment of the laser energy density. It also decreases adding sub-sequential laser pulses for low energy density values see Fig. 114.

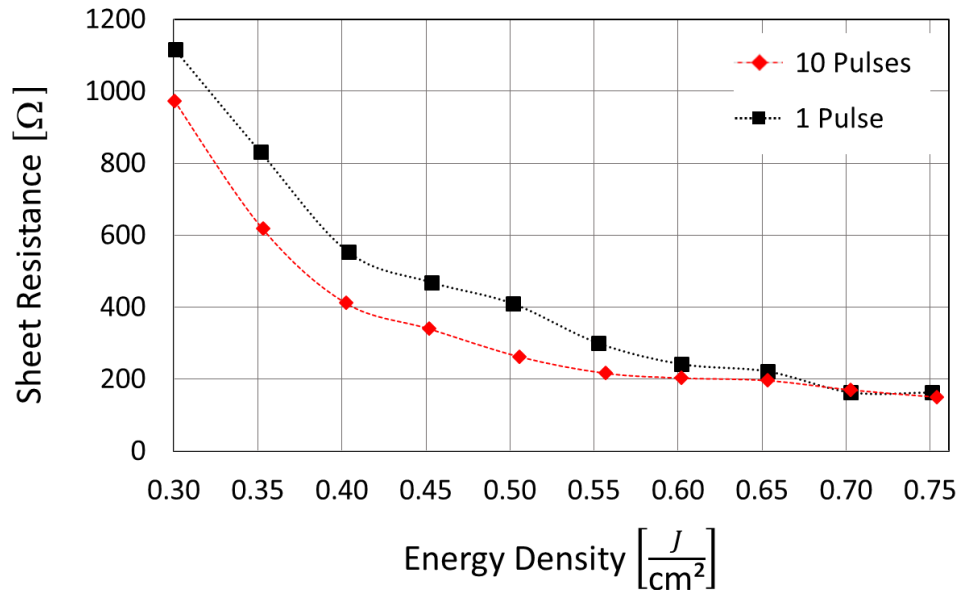


Fig. 114: Sheet resistance dependency on energy density values and number of laser pulses.

Additionally, as the energy density values increases the grains size increase together with surface roughness. The value of the resistivity and the elasticity depends on the grain orientation on the crystalline plane. In poly-Si layer the grains are oriented in random crystalline planes. To reduce electrical and mechanical variability from device to device it is suitable to include, for a given length and width, a maximal number of grains having all possible random orientation. However, the poly-Si NEMS width will be around 50nm, giving an upper limit to the grain size. The down limit is given by the minimum allowed resistance value. The optimal grain size it is obtained by a tradeoff between a minimizing the resistivity value and the device variability. In our batch, we annealed the a-Si at 0.40J/cm^2 with 2 pulses laser pulses. The process continues as the mono-crystalline flow.

The feasibility of poly-Si at low thermal budget with an acceptable sheet resistance has been demonstrated, and in the section below the electrical characterization is presented.

6.1.1 I/V CHARACTERIZATION OF POLYCRYSTALLINE STAND-ALONE NEMS RELAY

The switching behavior of the fabricated 5T SPDT polycrystalline Silicon NEM relay is evaluated using an Agilent B1500A semiconductor analyzer equipped with two high-resolution source measurement units (HRES-SMU) and 3SMU. The gate voltage is swept from 0V up to a voltage level $V_{DD} > V_{PI}$ and then swept back to 0V. The two contact electrodes voltages $V_{DL} = V_{DR} = 5V$ and the source voltage V_S to 10mV. The gate current I_{GL} and I_{GR} , the source current I_B and the drain current I_{DL} and I_{DR} is recorded. The wafers are exposed to air.

All the tested structures have been designed using relay parameters that satisfy the optimized pull-in and catastrophic pull-in margin (see Chapter 3). The simulations are carried out using finite element method (FEM) simulations. Switch ID from 8 to 11 are supposed to have a pull-in voltage of 4.8V and Switch designed from 12 to 17 are supposed to have a pull-in voltage of 1.8V.

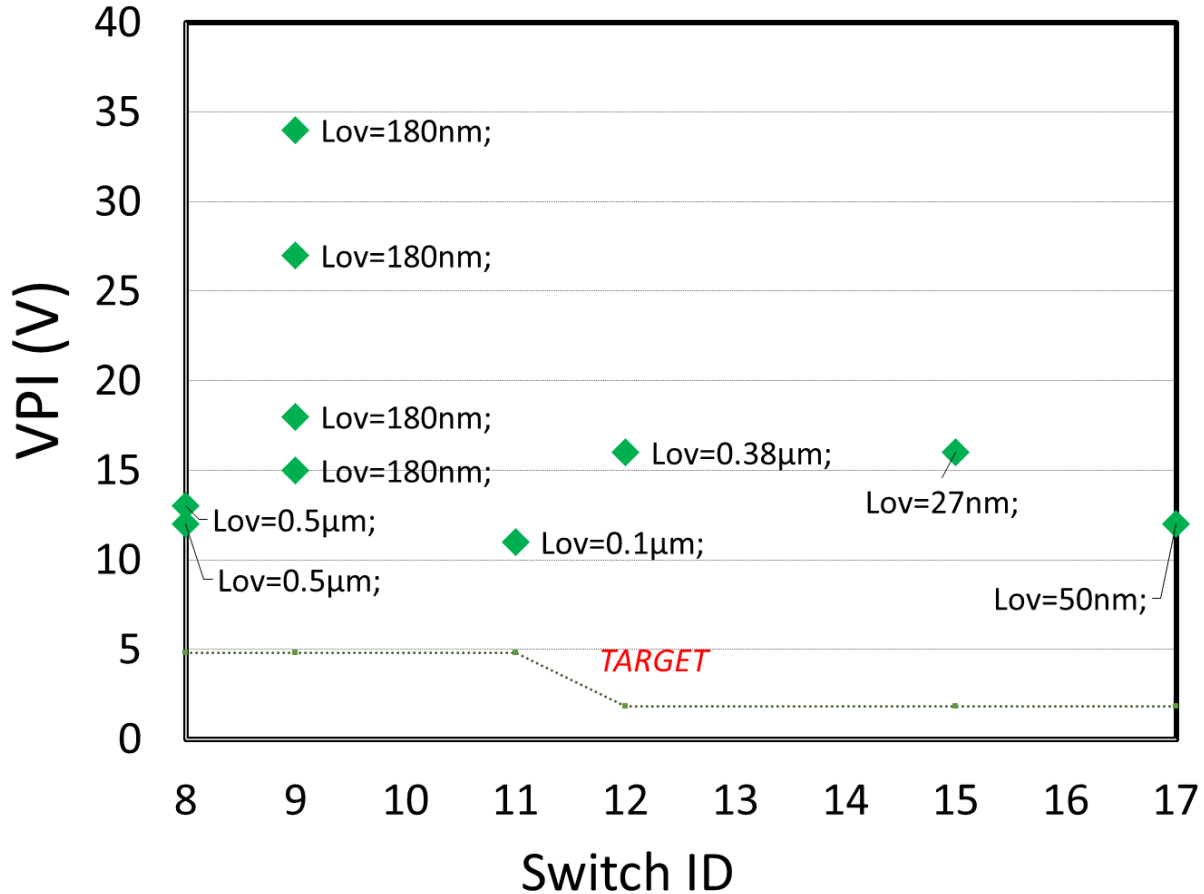


Fig. 115: Measured Pull In Voltage characteristic for different relay design (i.e. Switch ID).

Fig. 115 shows the value of the pull in event for various relay design. Each switch ID corresponds to a relay design and the multiples points for the same switch ID corresponds to different dies of the same wafers. We proposed various relay design

parameters in order to overcome the incertitude of the adhesion forces value and to obtain non-volatile and reprogrammable relays.

Fig. 115 clearly shows that under 10V it is impossible to measure a pull-in event, no matter the relay design. Therefore, we think that i) the relay effective contact area is very small, and the contact resistance is too high, or ii) there are secondary anchoring points along the beam. These secondary anchoring points increase the pull in voltage or in worst case-scenario prevent the beam-drain contact. Option i) is mainly because the relay designed on the mask is not perfectly re-produced on the wafer and because the silicon surfaces are oxidized. The second option is mainly due to the release process, which is not yet optimized for nano-gaps.

Additionally, for a same device, it is possible to observe various voltage pull in values, and for other devices, it is not possible to measure pull-in events. Same as before, we think that this is due to secondary anchored points created during the release process, or even during SEM observations in which the movable electrode moves and contacts the substrate and/or the gate electrodes in different points, increasing the pull in voltage value. Its variability can also depends on the current value, in fact, depending on the effective contacting area, the resistance could be too high and the current cannot be detected. Unfortunately, without the possibility to observe the mechanical movement during the electrical test, it is difficult to draw a definite conclusion. *In-situ* SEM electrical testing might be useful to understand if the problem comes from secondary anchoring point, from the impossibility to contact the drain without first contacting the gate or from a too high contact resistance.

For some relay, it was impossible to see current flowing to the drain. This is happening on the relays design having already a very small drain/movable electrode nominal overlap ($L_{ov} < 50\text{nm}$). In fact, the sequence of lithography and etch steps rounded the tip of the drain and the tip of the movable electrode, thereby decreasing the nominal overlap area. Looking at the example in Fig. 116, the tip of both drain and movable electrode are rounded, decreasing the nominal overlap area L_{ov} , increasing the drain beam gap g_d . The reduction of the contact area because of rounded tip can cause several issues:

- Increment of the contact resistance.
- Decrement of the adhesion force.
- Increment of the contact gap
- Contact failure (the beam is unable to contact the drain without first contacting the gate or in the worst-case scenario the beam is unable to contact the drain).

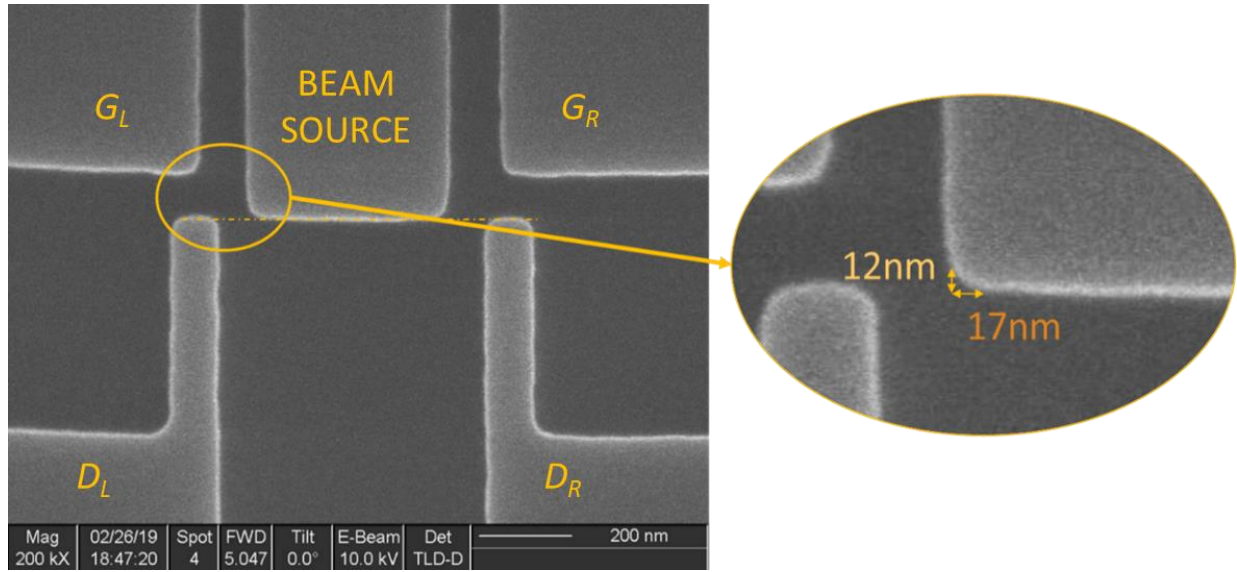


Fig. 116: SEM Top View image of a 5-Terminal poly-silicon relay having the drain and the movable electrode tip rounded.

6.1.1.1 DC CHARACTERISTIC

Fig. 117 shows I_B - V_{GL} curve measured with $V_D=5V$. The voltage sweep is applied to the left gate. In the OFF state, there is no leakage; the level of leakage seen represents the noise level in measurement setup ($I_{noise} < 10^{-12}$). The relay presents a very steep behavior with sharp ON/OFF behavior. The current on the source reaches the value of the current compliance ($1\mu A$) once the relay is ON. The V_{PI} voltage is 12.5V and there is not pull out voltage. Therefore, the relay is nonvolatile.

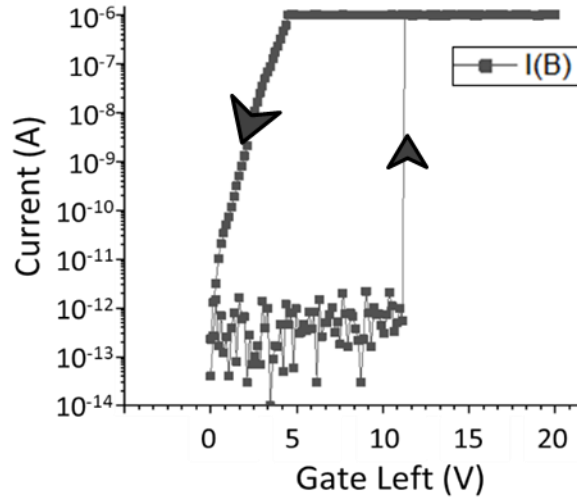


Fig. 117: Measured I_B - V_{GL} characteristics of 5-Terminal poly-silicon relay showing zero off-state leakage and abrupt switching behavior.

However, if we look at the dependence of the other currents on gate voltage, as shown in Fig. 118, the scenario is a little bit different. There is no current measured

on the drain electrode $I_{DL}=I_{noise}$. Unfortunately, the beam contacting the gate is a “catastrophic pull in” for relay as NV element for SRAM or TCAM. The catastrophic pull in could causes functional failure encoding the opposite state. Fig. 118 right shows the reprogram operation to right electrode. In this case, the gate right is swept at higher voltages with respect the gate left on the previous writing operation.

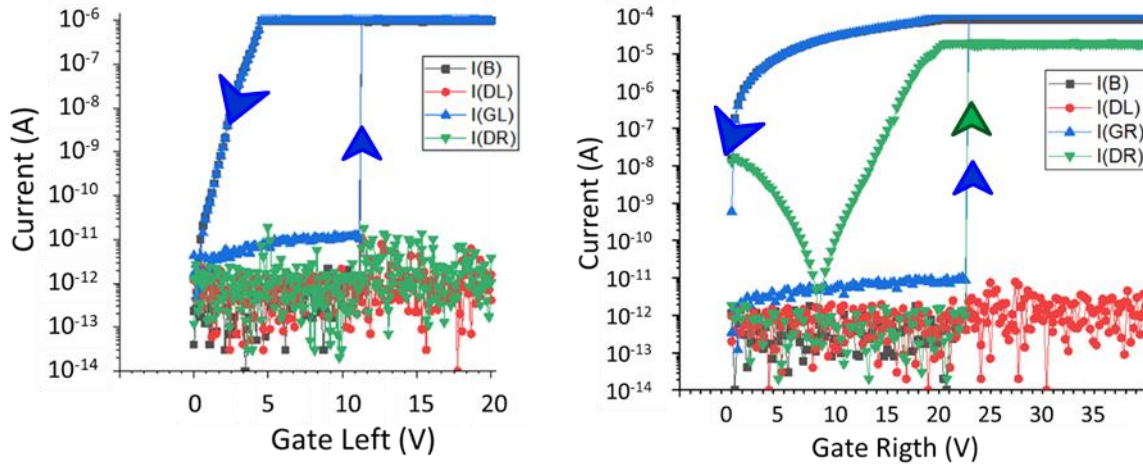


Fig. 118: Left) Measured $I-V_{GL}$ characteristics of the programming operation of a 5-Terminal poly-silicon relay leakage floor for the gate left and catastrophic pull in event. Right) Measured $I-V_{GR}$ characteristics of the reprogramming operation of 5-Terminal poly-silicon relay showing pull in and catastrophic pull in on the right side.

6.2CU BEOL NEMS

During the course of this thesis, in collaboration with UC Berkeley, we designed vertical NEMS on the back end of the line of ST Microelectronics 130nm HCMOS9A technology node. The scope of this stand-alone switches is to be nonvolatile and reprogrammable. The structure was fabricated in ST 130nm technology node. The minimum space, and the minimum width were given by the metal interconnection design rules. At Leti and Berkeley we worked for the release of this structures. Here the release process will be described.

A basic knowledge of the HCMOS9A technology is needed in order to develop and adapt a post-CMOS process to release the structure (ref. page. 33 of [ETIENNE 2018]). In our batch the BEOL goes to M4 and it uses Cu double damascene process with standard Fluorinated SiO₂ Inter Metal dielectrics [CMP 2019]. The interconnections are done with a dual-damascene process. The oxide etching thickness is controlled by a layer of SiN that acts as etch stop layer during the interconnect structure patterning. The BEOL switch interconnection are defined in M4, the contact electrodes are defined in M3, the actuation electrodes are defined both in M3 and M2 and the movable electrode is anchored in M1. A total stack of almost 3μm of fluorinated SiO₂ needs to be etched to release the structure. The

NEMS structure release process is usually done with HF vapor, however while dealing with NEMS structure on defined on the BEOL it might be not possible to use HF vapor. Even if copper does not etch under HF vapor, the metal barrier, which allows avoiding copper diffusion, might be etched under HF vapor (Ta for example etches under HF while TaN is degraded but not completely etched). Additionally, the SiN etch stop layer does not etch under HF vapor, while LPCVD Stoichiometric Si_3N_4 , and PECVD Nitride reacts and swell. Therefore, to release the NEMS it is necessary to perform a RIE to etch the entire stack, with a final isotropic etching step to undercut the structure. In the section below the BEOL switch release process flow is described.

6.2.1 SWITCH RELEASE

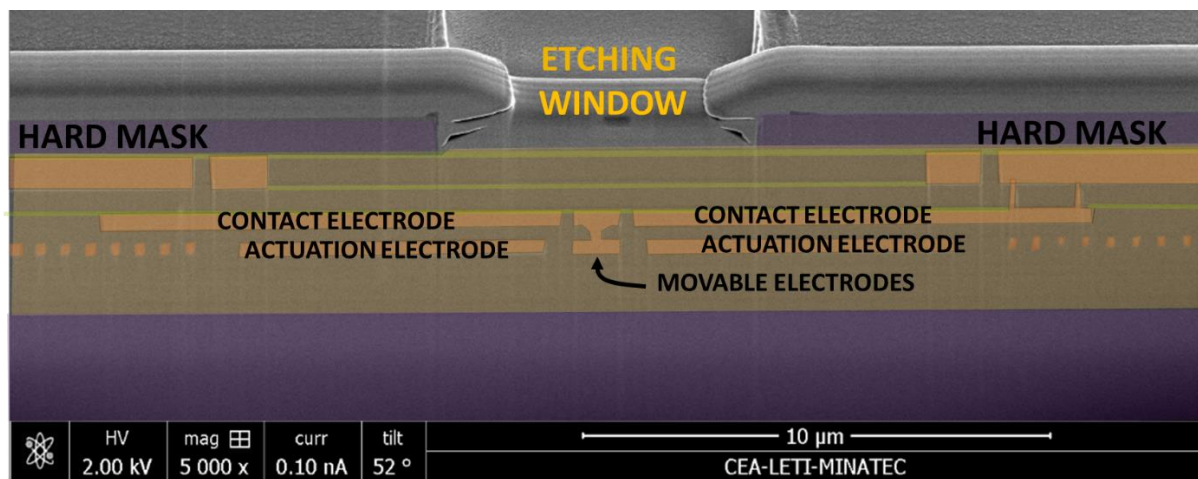


Fig. 119 Cross sectional tilted SEM view of the BEOL relay before release. The a-Si hard mask is deposited and etched to open the release window.

The process starts with a deposition of a hard mask to allow the release of the relay. The mask used is a-Si 800nm. A first iteration of photolithography, etching and stripping is done to open an area above the M4 alignment mark. Then, a second cycle of photolithography, etching and stripping is done to open a window on the area above the NEMS, which will release the “movable” structure. It is important that the area is also free of dummies; otherwise, it could cause the release of unwanted metal that could create shorts.

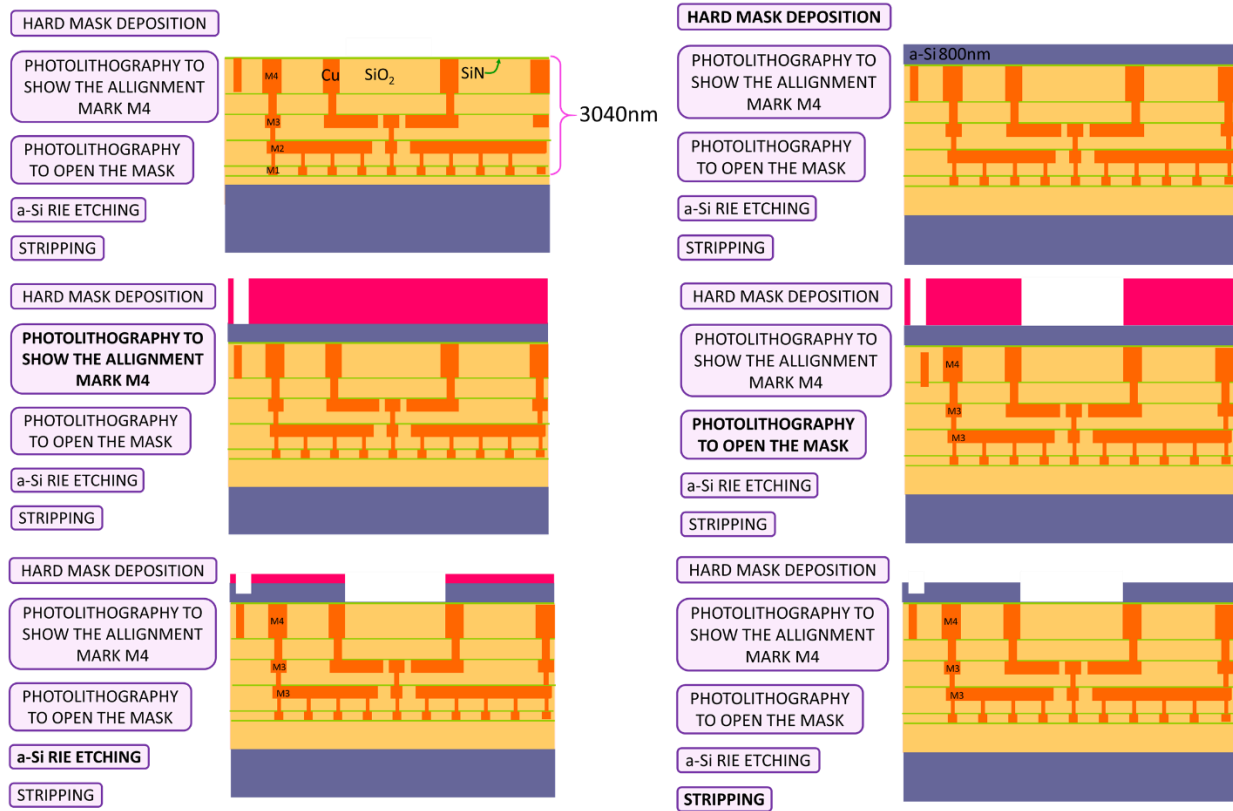


Fig. 120: Cross-sectional schematic view of NEMS relay defined on the CMOS interconnection showing the pre-release steps: hard mask deposition and aligned lithography for window opening.

The final release is performed at the UC Berkeley facility. The step is divided in cycles of dry etching and wet cleaning. The cleaning steps are necessary to remove the number of residues formed during the SiO_2 fluoride etching. The HF vapor cannot be used to etch the stack because of the presence of SiN , which swell once in contact with HF. The last step is a low power dry etching to undercut the oxide, allowing for complete release of the anchors of the structures.

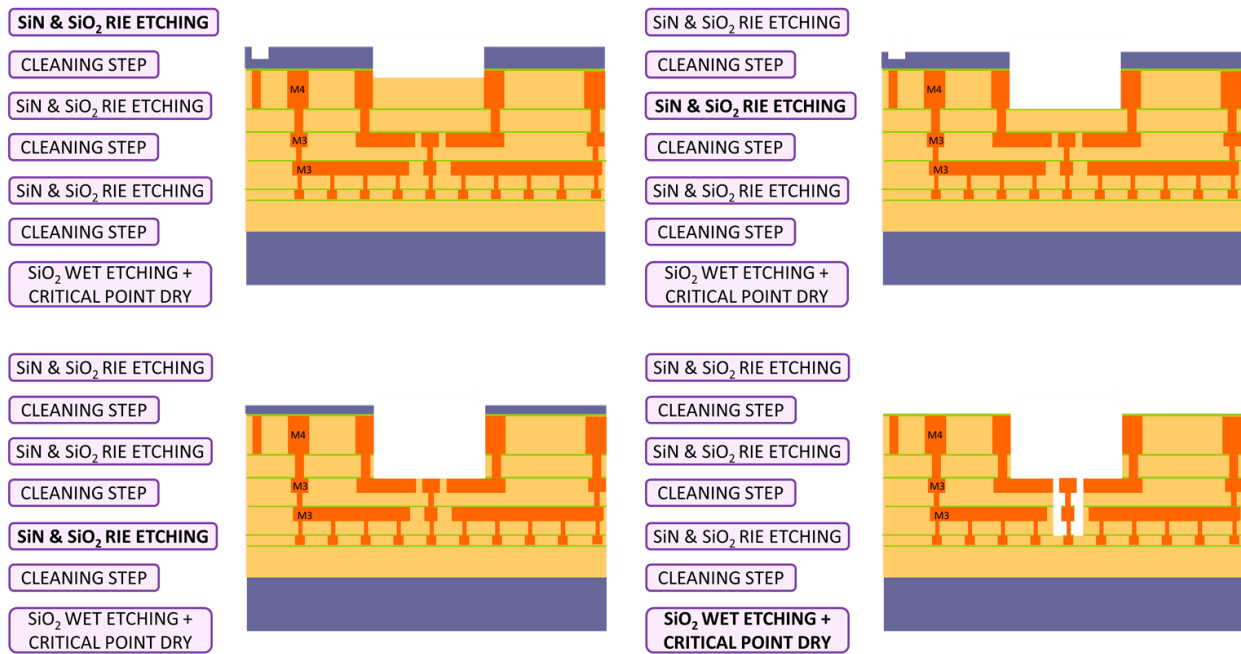


Fig. 121: Cross-sectional schematic view of NEMS relay defined on the CMOS interconnection showing the release steps: the etching is separated in cycles to not overheat the samples

The post-CMOS approach allows the NEMS fabrication without any modification of the standard baseline.

During this thesis, three post-CMOS fabrication methods have been studied, here below the advantages and limits have been listed.

<i>Post-CMOS</i>	ADVANTAGES	LIMITS
BEOL NEMS	Can be fabricated by CMOS foundry	Design depending on CMOS technology used
	Only 2 additional mask levels (for NEMS release)	
	No need to fabricate additional NEMS / CMOS interconnections	Materials limited by CMOS technology
c-Si NEMS	c-Si can be used as structural material	Grinding can weaken the wafer side=> development needed.
	Low parasitic capacitances	Requires high planarity of the CMOS substrates
	The high temperature process can be done before bonding	
	The NEMS relay can be co-integrate with NEMS Sensor and CMOS	~ Additional 5 mask level (for NEMS patterning & interconnection)
		Circuit area given by NEMS area
Poly-Si NEMS	The NEMS relay can be co-integrate with NEMS Sensor and CMOS	Assembly thermal budget <450°C ~2h
		Grain size and
	Low parasitic capacitances	~ Additional 5 mask level (for NEMS patterning & interconnection)

The poly-Si NEMS is an interesting approach for NEMS relays combined with NEMS sensors above CMOS. However, especially for NEMS relay designed with a beam width value $\leq 50\text{nm}$, having grains smaller than the beam width, to reduce device/device variability, could come at the expenses of too high resistivity values.

The BEOL NEMS approach is fabricated within the CMOS interconnections, it requires only few post processing steps to release the structure. However, it is the CMOS technology itself that limits the NEMS relay design and fabrication (*e.g.* NEMS materials, minimum gap/width/thickness).

BIBLIOGRAPHY

- [AKARVARDAR 2007A] Akarvardar K, Elata D, Parsa R, Wan G C, Yoo K, Provine J, Peumans P, Howe R T, Wong H-P, “Design Considerations for Complementary Nanoelectromechanical Logic Gates,” in 2007 IEEE International Electron Devices Meeting, pp. 299–302, 2007a.
- [ALHALABI 2017] Alhalabi R, Pendina G D, Prejbeanu I I, Nowak E, “High speed and high-area efficiency non-volatile look-up table design based on magnetic tunnel junction,” in 2017 17th Non-Volatile Memory Technology Symposium (NVMTS), pp. 1–4, 2017.
- [APPLEBY 2014] Appleby D J R, Ponon N K, Kwa K S K, Zou B, Petrov P K, Wang T, Alford N M, O’Neill A, “Experimental Observation of Negative Capacitance in Ferroelectrics at Room Temperature,” *Nano Lett.*, vol. 14, no. 7, pp. 3864–3868, 2014.
- [ARUTT 2017] Arutt C N, Alles M L, Liao W, Gong H, Davidson J L, Schrimpf R D, Reed R A, Weller R A, Bolotin K, Nicholl R, Pham T T, Zettl A, Qingyang D, Hu J, Li M, Alphenaar B W, Lin J-T, Shurva P D, McNamara S, Walsh K M, Feng P X-L, Hutin L, Ernst T, Homeijer B D, Polcawich R G, Proie R M, Jones J L, Glaser E R, Cress C D, Bassiri-Gharb N, “The study of radiation effects in emerging micro and nano electro mechanical systems (M and NEMs),” *Semicond. Sci. Technol.*, vol. 32, no. 1, p. 013005, 2017.
- [AZZAZ 2015] Azzaz M, Benoist A, Vianello E, Garbin D, Jalaguier E, Cagli C, Charpin C, Bernasconi S, Jeannot S, Dewolf T, Audoit G, Guedj C, Denorme S, Candelier P, Fenouillet-Beranger C, Perniola L, “Benefit of Al₂O₃/HfO₂ bilayer for BEOL RRAM integration through 16kb memory cut characterization,” in 2015 45th European Solid State Device Research Conference (ESSDERC), pp. 266–269, 2015.
- [BACKUS 1977] Backus J, “Can Programming Be Liberated from the von Neumann Style? A Functional Style and Its Algebra of Programs,” *ACM Turing Award Lect.*, vol. 21, no. 8, p. 29, 1977.
- [BAO 2000] Bao M, *Micro Mechanical Transducers: Pressure Sensors, Accelerometers and Gyroscopes*. Elsevier, 2000.
- [BEHBAHANI 2012] Behbahani A, Wood B, Benson D, Berner A, Hegwood B, Dejager J, Rhoden W, Ohme B, Sloat J, Harmon C, “Technology Requirements and Development for Affordable High-Temperature Distributed Engine Controls,” San Diego, p. 18, 2012.
- [BEIU 2014] Beiu V, Ibrahim W, Tache M, Liu T J K, “On ultra-low power hybrid NEMS-CMOS,” in 14th IEEE International Conference on Nanotechnology, pp. 201–206, 2014.
- [BERTHELON 2018] Berthelon, “CEMES - CNRS - Thèse Rémy Berthelon.”: <http://www.cemes.fr/These-Remy-Berthelon-1065?lang=fr.>, 2018.

- [BERTRAND 2011] Bertrand D, Sàrl H, “From biology to NEMS: the importance of new sensor developments,” p. 26, 2011.
- [BLACKMORE 1998] Blackmore D, Zhou J G, “Fractal Analysis of Height Distributions of Anisotropic Rough Surfaces,” *Fractals*, vol. 06, no. 01, pp. 43–58, 1998.
- [BORREGUERO 2017] Borreguero M R, “Micro and nano-electro-mechanical devices in the cmos back end: relays and resonators,” *Universitat Autònoma de Barcelona*, 2017.
- [BOTA 2017] Bota S A, Verd J, Barceló J, Gili X, Alorda B, Torrens G, Benito C D, Segura J, “Cantilever NEMS relay-based SRAM devices for enhanced reliability,” in 2017 12th International Conference on Design Technology of Integrated Systems In Nanoscale Era (DTIS), pp. 1–6, 2017.
- [CAO 2012] Cao J, Vitale W A, Ionescu A M, “Self-assembled nano-electro-mechanical tri-state carbon nanotube switches for reconfigurable integrated circuits,” in 2012 IEEE 25th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 188–191, 2012.
- [CATALAN 2015] Catalan G, Jiménez D, Gruverman A, “Negative capacitance detected: Ferroelectrics,” *Nat. Mater.*, vol. 14, no. 2, pp. 137–139, 2015.
- [CAUBET 2006] Caubet V, Beylier C, Borel S, Renault O, “Mechanisms of isotropic and selective etching between SiGe and Si,” *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct. Process. Meas. Phenom.*, vol. 24, no. 6, pp. 2748–2754, 2006.
- [CHARBUILLET 2006] Charbuillet C, Skotnicki T, Villaret A, “IMOS transistor,” US20060220086A1, 2006.
- [CHEN 2012A] Chen C, Lee W S, Parsa R, Chong S, Provine J, Watt J, Howe R T, Wong H-P, Mitra S, “Nano-Electro-Mechanical relays for FPGA routing: Experimental demonstration and a design technique,” in 2012 Design, Automation Test in Europe Conference Exhibition (DATE), pp. 1361–1366, 2012a.
- [CHEN 2012] Chen Chen, Lee S, Provine J, Chong S, Parsa R, Daesung Lee, Howe R T, Wong H-S P, Mitra S, “Nano-Electro-Mechanical (NEM) relays and their application to FPGA routing,” in 17th Asia and South Pacific Design Automation Conference, Sydney, Australia, pp. 639–639, 2012.
- [CHEN 2010] Chen C, Wong H-S P, Mitra S, Parsa R, Patil N, Chong S, Akarvardar K, Provine J, Lewis D, Watt J, Howe R T, “Efficient FPGAs using nanoelectromechanical relays,” in Proceedings of the 18th annual ACM/SIGDA international symposium on Field programmable gate arrays - FPGA '10, Monterey, California, USA, p. 273, 2010.
- [CHEN 2008] Chen F, Kam H, Markovic D, Liu T J K, Stojanovic V, Alon E, “Integrated Circuit Design with NEM Relays,” in 2008 IEEE/ACM

- International Conference on Computer-Aided Design, pp. 750–757, 2008.
- [CHEN 2014] Chen I-R, “Novel Material Integration for Reliable and Energy-Efficient NEM Relay Technology,” p. 112, 2014.
- [CHEN 2012B] Chen I-R, Hutin L, Park C, Lee R, Nathanael R, Yaung J, Jeon J, Liu T-J K, “Scaled Micro-Relay Structure with Low Strain Gradient for Reduced Operating Voltage,” ECS Trans., vol. 45, no. 6, pp. 101–106, 2012b.
- [CHEN 1987] Chen J, Chan T Y, Chen I C, Ko P K, Hu C, “Subbreakdown drain leakage current in MOSFET,” IEEE Electron Device Lett., vol. 8, no. 11, pp. 515–517, 1987.
- [CHEN 2012C] Chen Y, Nathanael R, Jeon J, Yaung J, Hutin L, Liu T J K, “Characterization of Contact Resistance Stability in MEM Relays With Tungsten Electrodes,” J. Microelectromechanical Syst., vol. 21, no. 3, pp. 511–513, 2012c.
- [CHEN 2015] Chen Y P, “Reliability Studies of Micro-Relays for Logic Applications,” University of California, Berkeley, Berkeley, 2015.
- [CHER 2003] Cher Ming Tan, Zhenghao Gan, Xiaofang Gao, “Temperature and stress distribution in the SOI structure during fabrication,” IEEE Trans. Semicond. Manuf., vol. 16, no. 2, pp. 314–318, 2003.
- [CHEVALLIER 2010] Chevallier C J, Siau C H, Lim S F, Namala S R, Matsuoka M, Bateman B L, Rinerson D, “A 0.13 μ m 64Mb multi-layered conductive metal-oxide memory,” in 2010 IEEE International Solid-State Circuits Conference - (ISSCC), pp. 260–261, 2010.
- [CHIU 2010A] Chiu P, Chang M, Sheu S, Lin K, Chiang P, Wu C, Lin W, Lin C, Hsu C, Chen F T, Su K, Kao M, Tsai M, “A low store energy, low VDDmin, nonvolatile 8T2R SRAM with 3D stacked RRAM devices for low power mobile applications,” in 2010 Symposium on VLSI Circuits, pp. 229–230, 2010a.
- [CHIU 2012] Chiu P F, Chang M F, Wu C W, Chuang C H, Sheu S S, Chen Y S, Tsai M J, “Low Store Energy, Low VDDmin, 8T2R Nonvolatile Latch and SRAM With Vertical-Stacked Resistive Memory (Memristor) Devices for Low Power Mobile Applications,” IEEE J. Solid-State Circuits, vol. 47, no. 6, pp. 1483–1496, 2012.
- [CHOI 2015] Choi W Y, Kim Y J, “Three-Dimensional Integration of Complementary Metal-Oxide-Semiconductor-Nanoelectromechanical Hybrid Reconfigurable Circuits,” IEEE Electron Device Lett., vol. 36, no. 9, pp. 887–889, 2015.
- [CHONG 2009] Chong S, Akarvardar K, Parsa R, Yoon J B, Howe R T, Mitra S, Wong H S P, “Nanoelectromechanical (NEM) relays integrated with CMOS SRAM for improved stability and low leakage,” in 2009 IEEE/ACM International Conference on Computer-Aided Design - Digest of Technical Papers, pp. 478–484, 2009.
- [CHOWDHURY 2013] Chowdhury F K, Pourzand H, Tabib-Azar M, “Investigation of

- contact resistance evolution of Ir, Pt, W, Ni, Cr, Ti, Cu and Al over repeated hot-contact switching for NEMS switches,” in 2013 IEEE 26th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 445–448, 2013.
- [CLINTON 2018] Clinton, Michael, Singh, Rajinder, Tsai, Marty, Zhang, Shayan, “A 5GHz 7nm L1 cache memory compiler for high-speed computing and mobile applications - IEEE Conference Publication,” 2018.
- [CMP 2019] CMP, “IC 130nm HCMOS9A - CMP: Circuits Multi-Projets,” IC STMicroelectronics 130nm CMOS High Voltage 4 ML HCMOS9A. <https://mycmp.fr/datasheet/ic-130nm-hcmos9a.>, 2019.
- [COLINGE 2016] Colinge J-P, Greer J C, Greer J, Nanowire Transistors: Physics of Devices and Materials in One Dimension. Cambridge University Press, 2016.
- [CUNNINGHAM 2011] Cunningham S, Kupnik M, “Wafer Bonding,” pp. 817–877, 2011.
- [CZAPLEWSKI 2009] Czaplewski D A, Patrizi G A, Kraus G M, Wendt J R, Nordquist C D, Wolfley S L, Baker M S, de Boer M P, “A nanomechanical switch for integration with CMOS logic,” J. Micromechanics Microengineering, vol. 19, no. 8, p. 085003, 2009.
- [DAAMEN 2007] Daamen R, “The evolution of multi-level air gap integration towards 32 nm node interconnects,” Microelectron. Eng., vol. 84, no. 2177–2183, 2007.
- [DAAMEN 2015] Daamen R, Bancken P, Emur Badaroglu D, Michelon J, Nguyen V, Verheijden G J A M, Humbert A, Waeterloos J, Yang A, Cheng J K, Chen L, Martens T, Hoofman R, “Multi-Level Air Gap Integration for 32/22nm Nodes using a Spin-on Thermal Degradable Polymer and a SiOC CVD Hard Mask.” IITC2007, 2015.
- [DADGOUR 2011] Dadgour H F, Hussain M M, Cassell A, Singh N, Banerjee K, “Impact of scaling on the performance and reliability degradation of metal-contacts in NEMS devices,” in 2011 International Reliability Physics Symposium, pp. 3D.3.1-3D.3.10, 2011.
- [DELEONIBUS 2018] Deleonibus S, “Emerging Devices for Low-Power and High-Performance Nanosystems: Physics, Novel Functions, and Data Processing,” CRC Press. <https://www.crcpress.com/Emerging-Devices-for-Low-Power-and-High-Performance-Nanosystems-Physics/Deleonibus/p/book/9789814800112>, 2018.
- [DELLAERT 2015] Dellaert D, Doutrelaigne J, “Compact thermally actuated latching MEMS switch with large contact force,” Electron. Lett., vol. 51, no. 1, pp. 80–81, 2015.
- [DENNARD 1974] Dennard R H, Gaensslen F H, Yu H-N, Rideout V L, Bassous E, Leblanc A R, “Design of Ion-Implanted MOSFET’s with Very Small Physical Dimensions,” Proc. IEEE, vol. 87, no. 4, p. 11, 1974.
- [DEWALD 1966] Dewey G, Chu-Kung B, Boardman J, Fastenau J M, Kavalieros J, Kotlyar R, Liu W K, Lubyshev D, Metz M, Mukherjee N, Oakey P,

- Pillarisetty R, Radosavljevic M, Then H W, Chau R, “Fabrication, characterization, and physics of III–V heterojunction tunneling Field Effect Transistors (H-TFET) for steep sub-threshold swing,” in 2011 International Electron Devices Meeting, pp. 33.6.1-33.6.4, 2011.
- [DIAZ LLORENTE 2018] DIAZ LLORENTE C, “Caractérisation des transistors a effet tunnel à basse température et des architectures innovantes pour l’intégration 3D,” 2018.
- [Do 2013] Do A T, Gopal J K, Singh P, Li C G, Yeo K S, Kim T T, “Design and array implementation a cantilever-based non-volatile memory utilizing vibrational reset,” in 2013 Proceedings of the European Solid-State Device Research Conference (ESSDERC), pp. 284–287, 2013.
- [DONG 2011] Dong C, Chen C, Mitra S, Chen D, “Architecture and performance evaluation of 3D CMOS-NEM FPGA,” in International Workshop on System Level Interconnect Prediction, pp. 1–8, 2011.
- [DURAFFOURG 2015] Duraffourg L, Arcamone J, Nanosystèmes électromécaniques. ISTE Editions, 2015.
- [DURAND 2008] Durand C, Casset F, Ancey P, Judong F, Talbot A, Quenouillère R, Renaud D, Borel S, Florin B, Buchaillot L, “Silicon on nothing MEMS electromechanical resonator,” *Microsyst. Technol.*, vol. 14, 2008.
- [ETIENNE 2018] Etienne N, Jean-Christophe C, “LETI AND CMP ANNOUNCE WORLD’S FIRST MULTI-PROJECT WAFER SERVICE WITH INTEGRATED SILICON OXRAM - CMP: Circuits Multi-Projets,” <https://mycmp.fr/about-us/announcements/leti-and-cmp-announce-world-s-first-multi-project-wafer-service-with-integrated.html>, 2018.
- [FARIBORZI 2010] Fariborzi H, Spencer M, Karkare V, Jeon J, Nathanael R, Wang C, Chen F, Kam H, Pott V, Liu T J K, Alon E, Stojanović V, Marković D, “Analysis and demonstration of MEM-relay power gating,” in IEEE Custom Integrated Circuits Conference 2010, pp. 1–4, 2010.
- [FENOUILLET-BERANGER 2014] Fenouillet-Beranger C, Previtali B, Batude P, Nemouchi F, Cassé M, Garros X, Tosti L, Rambal N, Lafond D, Dansas H, Pasini L, Brunet L, Deprat F, Grégoire M, Mellier M, Vinet M, “FDSOI bottom MOSFETs stability versus top transistor thermal budget featuring 3D monolithic integration,” in 2014 44th European Solid State Device Research Conference (ESSDERC), pp. 110–113, 2014.
- [FISCHER 2015] Fischer K, Agostinelli M, Allen C, Bahr D, Bost M, Charvat P, Chikarmane V, Fu Q, Ganpule C, Haran M, Heckscher M, Hiramatsu H, Hwang E, Jain P, Jin I, Kasim R, Kosaraju S, Lee K S, Liu H, McFadden R, Nigam S, Patel R, Pelto C, Plekhanov P, Prince M, Puls C, Rajamani S, Rao D, Reese P, Rosenbaum A, Sivakumar S, Song B, Uncuer M, Williams S, Yang M, Yashar P, Natarajan S, “Low-k interconnect stack with multi-layer air gap and

- tri-metal-insulator-metal capacitors for 14nm high volume manufacturing,” in 2015 IEEE International Interconnect Technology Conference and 2015 IEEE Materials for Advanced Metallization Conference (IITC/MAM), pp. 5–8, 2015.
- [FLANDRE 1999] Flandre D, Colinge J P, Chen J-J, Ceuster D D, Eggermont J, Ferreira L L D, Gentinne B, Jespers P G A, Viviani A, Raskin J P, Vorst A V, “Fully-Depleted SOI CMOS Technology for Low-Voltage Low-Power Mixed Digital/Analog/Microwave Circuits,” 1999.
- [FUJIKI 2014] Fujiki J, Xu N, Hutin L, Chen I, Qian C, Liu T K, “Microelectromechanical Relay and Logic Circuit Design for Zero Crowbar Current,” IEEE Trans. Electron Devices, vol. 61, no. 9, pp. 3296–3302, 2014.
- [GAMARRA 2014] Gamarra M, Luis J, “NEMS/MEMS integration in submicron CMOS Technologies,” Ph.D. Thesis, Universitat Autònoma de Barcelona, 2014.
- [GIRBAU 2003] Girbau D, Lazaro A, Pradell L, “RF MEMS switches based on the buckle-beam thermal actuator,” in 33rd European Microwave Conference Proceedings (IEEE Cat. No.03EX723C), vol. 2, pp. 651–654 vol.2, 2003.
- [GLICKMAN 2011] Glickman M, Tseng P, Harrison J, Niblock T, Goldberg I B, Judy J W, “High-Performance Lateral-Actuating Magnetic MEMS Switch,” J. Microelectromechanical Syst., vol. 20, no. 4, pp. 842–851, 2011.
- [GODOY 2001] Godoy A, López-Villanueva J A, Jiménez-Tejada J A, Palma A, Gámiz F, “A simple subthreshold swing model for short channel MOSFETs,” Solid-State Electron., vol. 45, no. 3, pp. 391–397, 2001.
- [GOPALAKRISHNAN 2002] Gopalakrishnan K, Griffin P B, Plummer J D, “I-MOS: a novel semiconductor device with a subthreshold slope lower than kT/q ,” in Digest. International Electron Devices Meeting, pp. 289–292, 2002.
- [GOSELE 1998] Gosele U, “SEMICONDUCTOR WAFER BONDING,” p. 27, 1998.
- [GOSSET 2003A] Gosset L G, Arnal V, Brun P, Broekaart M, Monget C, Casanova N, Rivoire M, Oberlin J-C, Torres J, “Integration of SiOC air gaps in copper interconnects,” Microelectron. Eng., vol. 70, no. 2, pp. 274–279, 2003a.
- [GOSSET 2003B] Gosset L G, Arnal V, Prindle C, Hoofman R, Verheijden G, Daamen R, Broussous L, Fusalba F, Assous M, Chatterjee R, Torres J, Gravesteijn D, Yu K C, “General review of issues and perspectives for advanced copper interconnections using air gap as ultra-low K material,” in Proceedings of the IEEE 2003 International Interconnect Technology Conference (Cat. No.03TH8695), pp. 65–67, 2003b.

- [GOVOREANU 2011] Govoreanu B, Kar G S, Chen Y, Paraschiv V, Kubicek S, Fantini A, Radu I P, Goux L, Clima S, Degraeve R, Jossart N, Richard O, Vandeweyer T, Seo K, Hendrickx P, Pourtois G, Bender H, Altimime L, Wouters D J, Kittl J A, Jurczak M, “10×10nm²Hf/HfO_xcrossbar resistive RAM with excellent performance, reliability and low-energy operation,” in 2011 International Electron Devices Meeting, pp. 31.6.1-31.6.4, 2011.
- [GROGG 2014] Grogg D, Ayala C L, Drechsler U, Sebastian A, Koelmans W W, Bleiker S J, Fernandez-Bolanos M, Hagleitner C, Despont M, Duerig U T, “Amorphous carbon active contact layer for reliable nanoelectromechanical switches,” in 2014 IEEE 27th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 143–146, 2014.
- [GROGG 2013] Grogg D, Drechsler U, Knoll A, Duerig U, Pu Y, Hagleitner C, Despont M, “Curved in-plane electromechanical relay for low power logic applications,” *J. Micromechanics Microengineering*, vol. 23, no. 2, p. 025024, 2013.
- [GROSSI 2018] Grossi A, Vianello E, Zambelli C, Royer P, Noel J, Giraud B, Perniola L, Olivo P, Nowak E, “Experimental Investigation of 4-kb RRAM Arrays Programming Conditions Suitable for TCAM,” *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 26, no. 12, pp. 2599–2607, 2018.
- [GROSSI, 2016] Grossi, Alessandro, “Emerging Non Volatile Memories Reliability,” University of Ferrara, 2016.
- [GUPTA 2009] Gupta T, “The Copper Damascene Process and Chemical Mechanical Polishing,” in *Copper Interconnect Technology*, T. Gupta, Ed. New York, NY: Springer New York, pp. 267–300, 2009.
- [HALAWANI 2019] Halawani Y, Mohammad B, Abu-Lebdeh M, Al-Qutayri M, Al-Sarawi S F, “ReRAM-based In-Memory Computing for Search Engine and Neural Network Applications,” *IEEE J. Emerg. Sel. Top. Circuits Syst.*, pp. 1–1, 2019.
- [HARRISON 2014] Harrison K L, Dalvi C, Asheghi M, Howe R T, “Analysis of asperity dominated contacts in nanoelectromechanical relays using thin films,” in *Fourteenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, pp. 1256–1260, 2014.
- [HE 2013] He T, Yang R, Ranganathan V, Rajgopal S, Tupta M A, Bhunia S, Mehregany M, Feng P X-, “Silicon carbide (SiC) nanoelectromechanical switches and logic gates with long cycles and robust performance in ambient air and at high temperature,” in 2013 IEEE International Electron Devices Meeting, pp. 4.6.1-4.6.4, 2013.
- [HENRY 2013] Henry M B, Nazhandali L, “NEMS-Based Functional Unit Power-Gating: Design, Analysis, and Optimization,” *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 60, no. 2, pp. 290–302, 2013.

- [HENRY 2011] Henry M, Srivastav M, Nazhandali L, “A case for NEMS-based functional-unit power gating of low-power embedded microprocessors,” in 2011 48th ACM/EDAC/IEEE Design Automation Conference (DAC), pp. 872–877, 2011.
- [HOOFMAN, 2006] Hoofman, Romano, “Alternatives to Low-k nanoporous materials: dielectric air-gap integration,” Solid State Technology, 2006.
- [INC 1997] Inc I M G, InfoWorld. InfoWorld Media Group, Inc., 1997.
- [IRDS 2017] IRDS, “INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS 2017 EDITION BEYOND CMOS.”, 2017.
- [ISSCC 2017] ISSCC, “ISSCC 2017 TRENDS.” International Solid-State Circuits Conference, 2017.
- [JANG 2008] Jang W W, Lee J O, Yoon J-B, Kim M-S, Lee J-M, Kim S-M, Cho K-H, Kim D-W, Park D, Lee W-S, “Fabrication and characterization of a nanoelectromechanical switch with 15-nm-thick suspension air gap,” Appl. Phys. Lett., vol. 92, no. 10, p. 103110, 2008.
- [JEON 2012] Jeon J, Hutin L, Jevtic R, Liu N, Chen Y, Nathanael R, Kwon W, Spencer M, Alon E, Nikolic B, Liu T J K, “Multiple-Input Relay Design for More Compact Implementation of Digital Logic Circuits,” IEEE Electron Device Lett., vol. 33, no. 2, pp. 281–283, 2012.
- [JEON 2010] Jeon J, Pott V, Kam H, Nathanael R, Alon E, Liu T J K, “Perfectly Complementary Relay Design for Digital Logic Applications,” IEEE Electron Device Lett., vol. 31, no. 4, pp. 371–373, 2010.
- [JEONG-OEN LEE 2010] Jeong-Oen Lee, Min-Wu Kim, Seung-Deok Ko, Jun-Bo Yoon, “Annealing effect on contact characteristics in TiN based 3-terminal NEM relays,” in 10th IEEE International Conference on Nanotechnology, pp. 258–261, 2010.
- [JOHNSON 2004] Johnson R W, Evans J L, Jacobsen P, Thompson J R, Christopher M, “The changing automotive environment: high-temperature electronics,” IEEE Trans. Electron. Packag. Manuf., vol. 27, no. 3, pp. 164–176, 2004.
- [JONES 2017] Jones S, “Intel Manufacturing Day: Nodes must die, but Moore’s Law lives!,” Semiwiki.com. <https://www.semiwiki.com/forum/content/>, 2017.
- [JOSHI 2010] Joshi V, Khieu C, Smith C G, Schepens C, Csaszar F, Lacey D, Nagata T, Renault M, Kampen R V, Knipe R, Yost D, “A CMOS compatible back end MEMS switch for logic functions,” in 2010 IEEE International Interconnect Technology Conference, pp. 1–3, 2010.
- [JURCZAK, 2015] Jurczak, Gosia, “Advances and Trends of RRAM technology,” Semicon Taiwan, 2015.
- [KAM 2011] Kam H, Liu T J K, Stojanovi? V, Markovic D, Alon E, “Design,

- Optimization, and Scaling of MEM Relays for Ultra-Low-Power Digital Logic,” IEEE Trans. Electron Devices, vol. 58, no. 1, pp. 236–250, 2011.
- [KAM 2009A] Kam H, Pott V, Nathanael R, Jeon J, Alon E, Liu T-J K, “Design and reliability of a micro-relay technology for zero-standby-power digital logic applications,” in 2009 IEEE International Electron Devices Meeting (IEDM), pp. 1–4, 2009a.
- [KAM 2009B] Kam H, “MOSFET Replacement Devices for Energy-Efficient Digital Integrated Circuits | EECS at UC Berkeley,” University of California, Berkeley, Berkeley, 2009b.
- [KAM 2015] Kam H, Chen F, Micro-Relay Technology for Energy-Efficient Integrated Circuits | Hei Kam | Springer, Springer. , 2015.
- [KANG 2001] Kang J K, Musgrave C B, “The mechanism of HF/H₂O chemical etching of SiO₂,” J. Chem. Phys., vol. 116, no. 1, pp. 275–280, 2001.
- [KANG 2007] Kang S, Cho W Y, Cho B, Lee K, Lee C, Oh H, Choi B, Wang Q, Kim H, Park M, Ro Y H, Kim S, Ha C, Kim K, Kim Y, Kim D, Kwak C, Byun H, Jeong G, Jeong H, Kim K, Shin Y, “A 0.1- μ m 1.8-V 256-Mb Phase-Change Random Access Memory (PRAM) With 66-MHz Synchronous Burst-Read Operation,” IEEE J. Solid-State Circuits, vol. 42, no. 1, pp. 210–218, 2007.
- [KARAM 2015] Karam R, Puri R, Ghosh S, Bhunia S, “Emerging Trends in Design and Applications of Memory-Based Computing and Content-Addressable Memories - IEEE Journals & Magazine,” Proc. IEEE, vol. 103, no. 8, pp. 1311–1330, 2015.
- [KATO 2016A] Kato K, Stojanović V, Liu T J K, “Embedded Nano-Electro-Mechanical Memory for Energy-Efficient Reconfigurable Logic,” IEEE Electron Device Lett., vol. 37, no. 12, pp. 1563–1565, 2016a.
- [KATO 2016B] Kato K, Stojanović V, Liu T J K, “Non-Volatile Nano-Electro-Mechanical Memory for Energy-Efficient Data Searching,” IEEE Electron Device Lett., vol. 37, no. 1, pp. 31–34, 2016b.
- [KENNETH 2004] Kenneth A, “Radiation Effects on Electronics 101,” Radiat. Eff., p. 69, 2004.
- [KIM 2011] Kim M, Song Y, Yoon J, “Fast and robust cantilever switch with suppressed bouncing for ic applications,” in 2011 IEEE 24th International Conference on Micro Electro Mechanical Systems, pp. 1229–1232, 2011.
- [KIM 2012] Kim S H, “Germanium-Source Tunnel Field Effect Transistors for Ultra-Low Power Digital Logic:,” Defense Technical Information Center, Fort Belvoir, VA, 2012.
- [KIM 2015] Kim Y J, Choi W Y, “Nonvolatile Nanoelectromechanical Memory Switches for Low-Power and High-Speed Field-Programmable Gate Arrays,” IEEE Trans. Electron Devices, vol. 62, no. 2, pp. 673–679,

2015.

- [Ko 2017] Ko E, Lee J W, Shin C, “Negative Capacitance FinFET With Sub-20-mV/decade Subthreshold Slope and Minimal Hysteresis of 0.48 V,” *IEEE Electron Device Lett.*, vol. 38, no. 4, pp. 418–421, 2017.
- [Kohl 2000] Kohl P A, Bhusari D M, Wedlake M, Case C, Klemens F P, Miner J, Byung-Chan Lee, Gutmann R J, Shick R, “Air-gaps in 0.3 μm electrical interconnections,” *IEEE Electron Device Lett.*, vol. 21, no. 12, pp. 557–559, 2000.
- [Kuehnel 1994] Kuehnel W, Sherman S, “A surface micromachined silicon accelerometer with on-chip detection circuitry,” *Sens. Actuators Phys.*, vol. 45, no. 1, pp. 7–16, 1994.
- [Kuon 2007] Kuon I, Tessier R, Rose J, “FPGA Architecture: Survey and Challenges,” *Found. Trends® Electron. Des. Autom.*, vol. 2, no. 2, pp. 135–253, 2007.
- [Kwon 2017] Kwon H S, Kim S K, Choi W Y, “Monolithic Three-Dimensional 65-nm CMOS-Nanoelectromechanical Reconfigurable Logic for Sub-1.2-V Operation,” *IEEE Electron Device Lett.*, vol. 38, no. 9, pp. 1317–1320, 2017.
- [Lee 2012] Lee and S, Provine J, Chong S, Parsa R, Howe and R T, Wong H-P, Mitra S, “Nano-Electro-Mechanical (NEM) relays and their application to FPGA routing,” in *17th Asia and South Pacific Design Automation Conference*, pp. 639–639, 2012.
- [Lee 2016] Lee C, “Reliability and failure analysis of MEMS/NEMS switches,” in *2016 IEEE 23rd International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, pp. 408–413, 2016.
- [Lee 2010A] Lee D, Lee W S, Provine J, Lee J-, Yoon J-, Howe R T, Mitra S, Wong H-P, “Titanium nitride sidewall stringer process for lateral nanoelectromechanical relays,” in *2010 IEEE 23rd International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 456–459, 2010a.
- [Lee 2010B] Lee D, Pott V, Kam H, Nathanael R, Liu T J K, “AFM characterization of adhesion force in micro-relays,” in *2010 IEEE 23rd International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 232–235, 2010b.
- [Lee 2009A] Lee J, Kim M, Ko S, Kang H, Bae W, Kang M, Kim K, Yoo D, Yoon J, “3-terminal nanoelectromechanical switching device in insulating liquid media for low voltage operation and reliability improvement,” in *2009 IEEE International Electron Devices Meeting (IEDM)*, pp. 1–4, 2009a.
- [Lee 2013] Lee J O, Song Y-H, Kim M-W, Kang M-H, Oh J-S, Yang H-H, Yoon J-B, “A sub-1-volt nanoelectromechanical switching device,” *Nat. Nanotechnol.*, vol. 8, no. 1, pp. 36–40, 2013.
- [Lee 2017] Lee M H, Chen P-, Fan S-, Chou Y-, Kuo C-, Tang C-, Chen H-, Gu S-, Hong R-, Wang Z-, Chen S-, Liao C-, Chen K-, Chang S T, Liao

- M-, Li K-, Liu C W, “Ferroelectric Al:HfO₂ negative capacitance FETs,” in 2017 IEEE International Electron Devices Meeting (IEDM), pp. 23.3.1-23.3.4, 2017.
- [LEE 2015] Lee M H, Chen P-, Liu C, Chu K, Cheng C-, Xie M-, Liu S-, Lee J-, Huang S-, Liao M-, Tang M, Li K-, Chen M-, “Prospects for ferroelectric HfZrO_x FETs with experimentally CET=0.98nm, SS_{for}=42mV/dec, SS_{rev}=28mV/dec, switch-off I_t ;0.2V, and hysteresis-free strategies,” in 2015 IEEE International Electron Devices Meeting (IEDM), pp. 22.5.1-22.5.4, 2015.
- [LEE 2004] Lee S W, Lee D S, Morjan R E, Jhang S H, Sveningsson M, Nerushev O A, Park Y W, Campbell E E B, “A Three-Terminal Carbon Nanorelay,” *Nano Lett.*, vol. 4, no. 10, pp. 2027–2030, 2004.
- [LEE 2009B] Lee T-, Speer K M, Fu X A, Bhunia S, Mehregany M, “Polycrystalline silicon carbide NEMS for high-temperature logic,” in TRANSDUCERS 2009 - 2009 International Solid-State Sensors, Actuators and Microsystems Conference, pp. 900–903, 2009b.
- [LEE 2010C] Lee T-H, Bhunia S, Mehregany M, “Electromechanical Computing at 500°C with Silicon Carbide,” *Science*, vol. 329, no. 5997, pp. 1316–1318, 2010c.
- [LEITE 2012] Leite F L, Bueno C C, Da Róz A L, Ziemath E C, Oliveira O N, “Theoretical models for surface forces and adhesion and their measurement using atomic force microscopy,” *Int. J. Mol. Sci.*, vol. 13, no. 10, pp. 12773–12856, 2012.
- [LEWIS 2015] Lewis A P, McBride J W, Jiang L, “Evolution of Voltage Transients During the Switching of a MEMS Relay With Au/MWCNT Contacts,” *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 5, no. 12, pp. 1747–1754, 2015.
- [LI 2010] Li G H, Laboriante I, Liu F, Shavezipur M, Bush B, Carraro C, Maboudian R, “Measurement of adhesion forces between polycrystalline silicon surfaces via a MEMS double-clamped beam test structure,” *J. Micromechanics Microengineering*, vol. 20, no. 9, p. 095015, 2010a.
- [LI 2017] Li H, Wu T F, Mitra S, Wong H-S P, “Device-architecture co-design for hyperdimensional computing with 3d vertical resistive switching random access memory (3D VRRAM) - IEEE Conference Publication,” in 2017 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), 2017.
- [LI 2015] Li K, Chen P, Lai T, Lin C, Cheng C, Chen C, Wei Y, Hou Y, Liao M, Lee M, Chen M, Sheih J, Yeh W, Yang F, Salahuddin S, Hu C, “Sub-60mV-swing negative-capacitance FinFET without hysteresis,” in 2015 IEEE International Electron Devices Meeting (IEDM), pp. 22.6.1-22.6.4, 2015.
- [LI 2018] Li R, Fariborzi H, “Ultra-Low Power Data Converters with BEOL NEM Relays,” in 2018 IEEE 61st International Midwest

- Symposium on Circuits and Systems (MWSCAS), pp. 627–630, 2018.
- [*Li 2010*] Li X, Yin L, Hu J, Lang L, Liu J, Xia Y, Fang D, Zhang H, “Electro-thermally actuated RF MEMS switch for wireless communication,” in 2010 IEEE 5th International Conference on Nano/Micro Engineered and Molecular Systems, pp. 497–500, 2010b.
- [*LIEBENS 2019*] Liebens M, Jourdain A, Vos J D, Vandeweyer T, Miller A, Beyne E, Li S, Bast G, Stoerring M, Hiebert S, Cross A, “In-Line Metrology for Characterization and Control of Extreme Wafer Thinning of Bonded Wafers,” IEEE Trans. Semicond. Manuf., vol. 32, no. 1, pp. 54–61, 2019.
- [*LIU 2010*] Liu J, Draghi L, Noman M, Bain J A, Schlesinger T E, Fedder G K, “High current low contact resistance platinum-coated CMOS-MEMS probes,” in 2010 IEEE 23rd International Conference on Micro Electro Mechanical Systems (MEMS), pp. 1143–1146, 2010.
- [*LIU 2012*] Liu T J K, Hutin L, Chen I R, Nathanael R, Chen Y, Spencer M, Alon E, “Recent progress and challenges for relay logic switch technology,” in 2012 Symposium on VLSI Technology (VLSIT), pp. 43–44, 2012.
- [*LIU 2017*] Liu T J K, Sikder U, Kato K, Stojanovic V, “There’s plenty of room at the top,” in 2017 IEEE 30th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 1–4, 2017a.
- [*LU 2014*] Lu H, “Tunnel Field-Effect Transistors: State-of-the-Art,” IEEE J. Electron Devices Soc., vol. 2, no. 4, pp. 44–49, 2014.
- [*LU 2018*] Lu Y, Shu J, Sun L, Mutlu O, “Improving the Performance and Endurance of Persistent Memory with Loose-Ordering Consistency,” IEEE Trans. Parallel Distrib. Syst., pp. 1–1, 2018.
- [*LUMBANTOBING 2004*] Lumbantobing A, Kogut L, Komvopoulos K, “Electrical contact resistance as a diagnostic tool for MEMS contact interfaces,” J. Microelectromechanical Syst., vol. 13, no. 6, pp. 977–987, 2004.
- [*LY 2018*] Ly D R B, Giraud B, Noel J-P, Grossi A, Castellani N, Sassine G, Nodin J-F, Molas G, Fenouillet-Beranger C, Indiveri G, Nowak E, Vianello E, “In-depth Characterization of Resistive Memory-Based Ternary Content Addressable Memories,” in 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, pp. 20.3.1-20.3.4, 2018.
- [*MANOHAR 2012*] Manohar S K, Venkatasubramanian R, Balsara P T, “Hybrid NEMS-CMOS DC-DC Converter for Improved Area and Power Efficiency,” in 2012 25th International Conference on VLSI Design, pp. 221–226, 2012.
- [*MASZARA 1988*] Maszara W P, Goetz G, Caviglia A, McKitterick J B, “Bonding of silicon wafers for silicon-on-insulator,” J. Appl. Phys., vol. 64, no. 10, pp. 4943–4950, 1988.

- [MAYER 2006] Mayer F, Royer C L, Carval G L, Tabone C, Clavelier L, Deleonibus S, “Co-integration of 2 mV/dec Subthreshold Slope Impact Ionization MOS (I-MOS) with CMOS,” 2006 Eur. Solid-State Device Res. Conf., pp. 303–306, 2006.
- [MERLIJN VAN SPENGEN 2003] Merlijn Van Spengen W, Abstract MEMS reliability from a failure mechanisms perspective. , 2003.
- [MIKOLAJICK 2018] Mikolajick T, Schroeder U, Slesazeck S, “Hafnium oxide based ferroelectric devices for memories and beyond,” in 2018 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), pp. 1–2, 2018.
- [MITTAL 2019] Mittal S, “A Survey of ReRAM-Based Architectures for Processing-In-Memory and Neural Networks,” Mach. Learn. Knowl. Extr., vol. 1, no. 1, pp. 75–114, 2019.
- [MOORE 1965] Moore G E, “Cramming More Components Onto Integrated Circuits,” Electronics, vol. 38, no. 8, pp. 82–85, 1965.
- [MUHLSTEIN 2001] Muhlstein C L, Brown S B, Ritchie R O, “High-cycle fatigue of single-crystal silicon thin films,” J. Microelectromechanical Syst., vol. 10, no. 4, pp. 593–600, 2001.
- [MÜLLER 1994] Müller G, Krötz G, Niemann E, “SiC for sensors and high-temperature electronics,” Sens. Actuators Phys., vol. 43, no. 1, pp. 259–268, 1994.
- [MUÑOZ-GAMARRA 2014] Muñoz-Gamarrá J L, Vidal-Álvarez G, Torres F, Uranga A, Barniol N, “CMOS-MEMS switches based on back-end metal layers,” Microelectron. Eng., vol. 119, pp. 127–130, 2014.
- [MUÑOZ-GAMARRA 2016] Muñoz-Gamarrá J L, Uranga A, Barniol N, “CMOS-NEMS Copper Switches Monolithically Integrated Using a 65 nm CMOS Technology,” Micromachines, vol. 7, no. 2, p. 30, 2016.
- [NATARAJAN 2014] Natarajan S, Agostinelli M, Akbar S, Bost M, Bowonder A, Chikarmane V, Chouksey S, Dasgupta A, Fischer K, Fu Q, Ghani T, Giles M, Govindaraju S, Grover R, Han W, Hanken D, Haralson E, Haran M, Heckscher M, Heussner R, Jain P, James R, Jhaveri R, Jin I, Kam H, Karl E, Kenyon C, Liu M, Luo Y, Mehandru R, Morarka S, Neiberg L, Packan P, Paliwal A, Parker C, Patel P, Patel R, Pelto C, Pipes L, Plekhanov P, Prince M, Rajamani S, Sandford J, Sell B, Sivakumar S, Smith P, Song B, Tone K, Troeger T, Wiedemer J, Yang M, Zhang K, “A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size,” in 2014 IEEE International Electron Devices Meeting, pp. 3.7.1-3.7.3, 2014.
- [NATHANAEL 2009] Nathanael R, Pott V, Kam H, Jeon J, Liu T K, “4-terminal relay technology for complementary logic,” in 2009 IEEE International Electron Devices Meeting (IEDM), pp. 1–4, 2009.
- [NATHANAEL 2012] Nathanael R, “Nano-Electro-Mechanical (NEM) Relay Devices and Technology for Ultra-Low Energy Digital Integrated Circuits,” p.

134, 2012.

- [O'BRIEN 2001] O'Brien G, Monk D J, Lin L, "MEMS cantilever beam electrostatic pull-in model," in Design, Characterization, and Packaging for MEMS and Microelectronics II, vol. 4593, pp. 31–41, 2001.
- [OLLIER 2008] Ollier E, Andreucci P, Duraffourg L, Colinet E, Durand C, Casset F, Ernst T, Hentz S, Labarthe S, Marcoux C, Nguyen V, Renaud D, Mile E, Renaux P, Mercier D, Robert P, Ancey P, Bouchaillot L, "NEMS based on top-down technologies: from stand-alone NEMS to VLSI NEMS amp; NEMS-CMOS integration," in 2008 IEEE International Conference on Electron Devices and Solid-State Circuits, pp. 1–6, 2008.
- [OSOBA 2018] Osoba B, Saha B, Almeida S F, Patil J, Brandt L E, Roots M E D, Acosta E, Wu J, Liu T K, "Variability Study for Low-Voltage Microelectromechanical Relay Operation," IEEE Trans. Electron Devices, vol. 65, no. 4, pp. 1529–1534, 2018.
- [OSOBA 2016] Osoba B, Saha B, Dougherty L, Edgington J, Qian C, Niroui F, Lang J H, Bulovic V, Wu J, Liu T J K, "Sub-50 mV NEM relay operation enabled by self-assembled molecular coating," in 2016 IEEE International Electron Devices Meeting (IEDM), pp. 26.8.1–26.8.4, 2016.
- [OUERGHI 2015] Ouerghi I, Sansa M, Ludurczak W, Duraffourg L, Benedetto K, Besombes P, Moffitt T, Adams B, Larmagnac D, Gergaud P, Poulain C, Vidaña A I, Ladner C, Fabbri J M, Muiyard D, Rodriguez G, Rabille G, Pollet O, Brianceau P, Kerdiles S, Hentz S, Ernst T, "Polysilicon nanowire NEMS fabricated at low temperature for above IC NEMS mass sensing applications," in 2015 IEEE International Electron Devices Meeting (IEDM), pp. 18.3.1–18.3.4, 2015a.
- [PALMA 2013] Palma G, "Nouvelles Architectures Hybrides: Logique / Mémoires Non-Volatiles et technologies associées.," Université de Grenoble, 2013.
- [PAN 2015] Pan Y, Yu F, Rutgers J J, "Fully-polymeric NEM relay for flexible, transparent, ultra-low power electronics and sensors," in 2015 28th IEEE International Conference on Micro Electro Mechanical Systems (MEMS), pp. 940–943, 2015.
- [PARK 2008] Park S, Allen S A B, Kohl P A, "Air-Gaps for High-Performance On-Chip Interconnect Part II: Modeling, Fabrication, and Characterization," J. Electron. Mater., vol. 37, no. 10, pp. 1534–1546, 2008.
- [PARSA 2013] Parsa R, Lee W S, Shavezipur M, Provine J, Maboudian R, Mitra S, Wong H S P, Howe R T, "Laterally Actuated Platinum-Coated Polysilicon NEM Relays," J. Microelectromechanical Syst., vol. 22, no. 3, pp. 768–778, 2013.
- [PARSA 2011] Parsa R, Shavezipur M, Lee W S, Chong S, Lee D, Wong H S P, Maboudian R, Howe R T, "Nanoelectromechanical relays with

- decoupled electrode and suspension,” in 2011 IEEE 24th International Conference on Micro Electro Mechanical Systems, pp. 1361–1364, 2011.
- [PAUL 2014] Paul S, Bhunia S, Computing with Memory for Energy-Efficient Robust Systems. New York: Springer-Verlag, 2014.
- [PAWASHE 2013] Pawashe C, Lin K, Kuhn K J, “Scaling Limits of Electrostatic Nanorelays,” IEEE Trans. Electron Devices, vol. 60, no. 9, pp. 2936–2942, 2013.
- [PENNY 2017] Penny C, Gates S, Peethala B, Lee J, Priyadarshini D, Nguyen S, McLaughlin P, Liniger E, Hu C-, Clevenger L, Hook T, Shobha H, Kerber P, Seshadri I, Chen J, Edelstein D, Quon R, Bonilla G, Paruchuri V, Huang E, “Reliable airgap BEOL technology in advanced 48 nm pitch copper/ULK interconnects for substantial power and performance benefits,” in 2017 IEEE International Interconnect Technology Conference (IITC), pp. 1–4, 2017.
- [PESCHOT 2015] Peschot A, Qian C, King Liu T-J, “Nanoelectromechanical Switches for Low-Power Digital Computing,” Micromachines, vol. 6, pp. 1046–1065, 2015.
- [PHILIPPE 2016] Philippe J, “Technology development and analysis of a multiphysic system based on NEMS co-integrated with CMOS for mass detection application,” University of Grenoble Alpes, 2016.
- [PHILIPPE 2014] Philippe J, “Technologie de fabrication et analyse de fonctionnement d’un système multi-physique de détection de masse à base de NEMS co-intégrés CMOS,” thesis, Grenoble, 2014.
- [POLLET 2013] Pollet O, Segaud R, Marcoux C, Crecy F de, “Eliminating stiction in NEMS and MEMS release: parameter optimization for an HF vapor process operating at room temperature and ambient pressure,” in Micromachining and Microfabrication Process Technology XVIII, vol. 8612, p. 86120B, 2013.
- [PROKOPOVICH 2011] Prokopovich P, Starov V, “Adhesion models: From single to multiple asperity contacts,” Adv. Colloid Interface Sci., vol. 168, no. 1, pp. 210–222, 2011.
- [QIAN 2015A] Qian C, Peschot A, Chen I, Chen Y, Xu N, Liu T K, “Effect of Body Biasing on the Energy-Delay Performance of Logic Relays,” IEEE Electron Device Lett., vol. 36, no. 8, pp. 862–864, 2015a.
- [QIAN 2015B] Qian C, Peschot A, Connelly D J, Liu T J K, “Energy-delay performance optimization of NEM logic relay,” in 2015 IEEE International Electron Devices Meeting (IEDM), pp. 18.1.1-18.1.4, 2015b.
- [QIAN 2017] Qian C, Peschot A, Osoba B, Ye Z A, Liu T J K, “Sub-100 mV Computing With Electro-Mechanical Relays,” IEEE Trans. Electron Devices, vol. 64, no. 3, pp. 1323–1329, 2017a.
- [QIN 2018] Qin T, Bleiker S J, Rana S, Niklaus F, Pamunuwa D, “Performance Analysis of Nanoelectromechanical Relay-Based Field-

- Programmable Gate Arrays,” IEEE Access, vol. 6, pp. 15997–16009, 2018.
- [QU 2016] Qu H, “CMOS MEMS Fabrication Technologies and Devices,” Micromachines, vol. 7, no. 1, p. 14, 2016.
- [RAMEZANI 2014] Ramezani M, Cosemans S, Coster J D, Rottenberg X, Rochus V, Osman H, Tilmans H A C, Severi S, Meyer K D, “Submicron three-terminal SiGe-based electromechanical ohmic relay,” in 2014 IEEE 27th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 1095–1098, 2014.
- [RAMEZANI 2015] Ramezani M, Severi S, Moussa A, Osman H, Tilmans H A C, Meyer K D, “Contact reliability improvement of a poly-SiGe based nano-relay with titanium nitride coating,” in 2015 Transducers - 2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), pp. 576–579, 2015.
- [RANA 2014] Rana S, Qin T, Bazigos A, Grogg D, Despont M, Ayala C L, Hagleitner C, Ionescu A M, Canegallo R, Pamunuwa D, “Energy and Latency Optimization in NEM Relay-Based Digital Circuits,” IEEE Trans. Circuits Syst. Regul. Pap., vol. 61, no. 8, pp. 2348–2359, 2014.
- [REDDICK 1995] Reddick W M, “Silicon surface tunnel transistor,” Appl. Phys. Lett., vol. 67, no. 4, pp. 494–496, 1995.
- [REZVANIAN 2007] Rezvanian O, Zikry M, Brown C, Krim J, “Surface Roughness, Asperity Contact and Gold RFMEMS Switch Behavior,” J. Micromechanics Microengineering, vol. 17, p. 2006, 2007.
- [RIETHMULLER 1988] Riethmuller W, Benecke W, “Thermally excited silicon microactuators,” IEEE Trans. Electron Devices, vol. 35, no. 6, pp. 758–763, 1988.
- [RIVEROLA 2018] Riverola M, Uranga A, Torres F, Barniol N, “Fabrication and characterization of a hammer-shaped CMOS/BEOL-embedded nanoelectromechanical (NEM) relay,” elsevier., 2018.
- [RIVEROLA 2016] Riverola M, Vidal-Álvarez G, Sobreviela G, Uranga A, Torres F, Barniol N, “Dynamic Properties of Three-Terminal Tungsten CMOS-NEM Relays Under Nonlinear Tapping Mode,” IEEE Sens. J., vol. 16, no. 13, pp. 5283–5291, 2016.
- [SAGE 2018] Sage E, Sansa M, Fostner S, Defoort M, Gély M, Naik A K, Morel R, Duraffourg L, Roukes M L, Alava T, Jourdan G, Colinet E, Masselon C, Brenac A, Hentz S, “Single-particle mass spectrometry with arrays of frequency-addressed nanomechanical resonators,” Nat. Commun., vol. 9, 2018.
- [SAHA 2017A] Saha B, Peschot A, Osoba B, Ko C, Rubin L, Liu T-J K, Wu J, “Reducing adhesion energy of micro-relay electrodes by ion beam synthesized oxide nanolayers,” APL Mater., vol. 5, no. 3, p. 036103, 2017a.

- [SAHA 2017B] Saha S, Kumar U S, Baghini M S, Goel M, Rao V R, “A Nano-Electro-Mechanical Switch Based Power Gating for Effective Stand-by Power Reduction in FinFET Technologies,” IEEE Electron Device Lett., vol. 38, no. 5, pp. 681–684, 2017b.
- [SALAHUDDIN 2008] Salahuddin S, Datta S, “Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices,” Nano Lett., vol. 8, no. 2, pp. 405–410, 2008.
- [SANKAR 2017] Sankar S, Kumar U S, Goel M, Baghini M S, Rao V R, “Considerations for Static Energy Reduction in Digital CMOS ICs Using NEMS Power Gating,” IEEE Trans. Electron Devices, vol. 64, no. 3, pp. 1399–1403, 2017.
- [SENG 2012] Seng C B, “LOW POWER HIGH PERFORMANCES ANALYSIS OF IMPACT IONIZATION MOSFET (IMOS) DEVICE,” p. 7, 2012a.
- [SHAVEZIPUR 2013] Shavezipur M, Lee W S, Harrison K L, Provine J, Mitra S, Wong H-P, Howe R T, “Laterally actuated nanoelectromechanical relays with compliant, low resistance contact,” in 2013 IEEE 26th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 520–523, 2013.
- [SHIEH 1998] Shieh B, Saraswat K C, McVittie J P, List S, Nag S, Islamraja M, Havemann R H, “Air-gap formation during IMD deposition to lower interconnect capacitance,” IEEE Electron Device Lett., vol. 19, no. 1, pp. 16–18, 1998.
- [SI 2015] Si L, Wang X, Xie G, Sun N, “Nano-adhesion and friction of multi-asperity contact: a molecular dynamics simulation study,” Surf. Interface Anal., vol. 47, no. 9, pp. 919–925, 2015.
- [SIKDER 2017] Sikder U, Liu T J K, “Design optimization for NEM relays implemented in BEOL layers,” in 2017 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), pp. 1–3, 2017.
- [SILVANO 2012] Silvano C, “Memory Hierarchy: Advanced Concepts,” p. 68, 2012.
- [SMITH 1995] Smith J H, Montague S, Sniegowski J J, Murray J R, McWhorter P J, “Embedded micromechanical devices for the monolithic integration of MEMS with CMOS,” in Proceedings of International Electron Devices Meeting, Washington, DC, USA, pp. 609–612, 1995.
- [SONG 2018] Song Z T, Cai D L, Li X, Wang L, Chen Y F, Chen H P, Wang Q, Zhan Y P, Ji M H, “High Endurance Phase Change Memory Chip Implemented based on Carbon-doped Ge₂Sb₂Te₅ in 40 nm Node for Embedded Application,” in 2018 IEEE International Electron Devices Meeting (IEDM), pp. 27.5.1-27.5.4, 2018.
- [SPIELER 1997] Spieler H, “Introduction to radiation-resistant semiconductor devices and circuits,” in AIP Conference Proceedings, Argonne, Illinois (USA), vol. 390, pp. 23–49, 1997.

- [TAKEUCHI 2005] Takeuchi H, Wung A, Sun X, Howe R T, King T-J, “Thermal budget limits of quarter-micrometer foundry CMOS for post-processing MEMS devices,” IEEE Trans. Electron Devices, vol. 52, no. 9, pp. 2081–2086, 2005.
- [TAYLOR 1998] Taylor W P, Brand O, Allen M G, “Fully integrated magnetically actuated micromachined relays,” J. Microelectromechanical Syst., vol. 7, no. 2, pp. 181–191, 1998.
- [TIMOSHENKO 1940] Timoshenko S, Strength Of Materials By Timoshenko Part I Elementary Theory and Problems, D. VAN NOSTRAND Company, Inc. , 1940.
- [TOMIOKA 2012] Tomioka K, Yoshimura M, Fukui T, “Steep-slope tunnel field-effect transistors using III–V nanowire/Si heterojunction,” in 2012 Symposium on VLSI Technology (VLSIT), pp. 47–48, 2012.
- [TROLIER-MCKINSTRY 2004] Trolier-McKinstry, Muralt P, “Thin Film Piezoelectrics for MEMS: Special Issue on Electroceramics in Micro-Electro-Mechanical Systems,” J. Electroceramics, vol. 12, 2004.
- [UDESHI 2004] Udeshi and K, Park J, Gianchandani Y B, “A bi-stable electro-thermal RF switch for high power applications,” in 17th IEEE International Conference on Micro Electro Mechanical Systems. Maastricht MEMS 2004 Technical Digest, pp. 797–800, 2004.
- [USAI 2017A] Usai G, Hutin L, Muñoz-Gamarra J L, Ernst T, Vinet M, Feng P X-, “Design considerations for optimization of pull-in stability margin in electrostatic N/MEM relays,” in 2017 IEEE International Conference on IC Design and Technology (ICICDT), pp. 1–4, 2017a.
- [USAI 2017B] Usai G, Hutin L, Sikder U, Muñoz-Gamarra J L, Ernst T, Liu T J K, Vinet M, “Balancing pull-in and adhesion stability margins in non-volatile NEM switches,” in 2017 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), pp. 1–2, 2017b.
- [VENKATASUBRAMANIAN 2013] Venkatasubramanian R, Manohar S K, Balsara P T, “NEM Relay-Based Sequential Logic Circuits for Low-Power Design,” IEEE Trans. Nanotechnol., vol. 12, no. 3, pp. 386–398, 2013.
- [VINCENT 2011] Vincent M, “Etude des mécanismes de défaillance du contact électrique dans un micro-interrupteur en technologie MEMS,” Institut National Polytechnique de Grenoble, 2011.
- [VRINCEANU 2002] Vrinceanu I D, Danyluk S, “Measurement of residual stress in single crystal silicon wafers,” in 2002 Proceedings. 8th International Advanced Packaging Materials Symposium (Cat. No.02TH8617), pp. 297–301, 2002.
- [WANG 2017] Wang C, Liu Y, Suga T, “A Comparative Study: Void Formation in Silicon Wafer Direct Bonding by Oxygen Plasma Activation with and without Fluorine,” ECS J. Solid State Sci. Technol., vol. 6, no. 1, pp. P7–P13, 2017.

- [WANG 2004] Wang F, Okada L, “Air gap dual damascene process and structure,” US20040232552A1, 2004.
- [WATSON 2015] Watson J, Castro G, “A review of high-temperature electronics technology and applications,” J. Mater. Sci. Mater. Electron., vol. 26, no. 12, pp. 9226–9235, 2015.
- [WEIDENMUELLER 2008] Weidenmueller U, Alves H, Schnabel B, Icard B, Le-Denmat J, Manakli S, Pradelles J, “New alignment marks for improved measurement maturity,” in 24th European Mask and Lithography Conference, pp. 1–7, 2008.
- [WONG 2018] Wong H S P, Ahn C, Cao J, Chen Y H, Fong S W, eryilmaz S B, Incorvia J A, Okabe K, “Stanford Memory Trends.”, 2018.
- [WU 2008] Wu L, Xie H, “A large vertical displacement electrothermal bimorph microactuator with very small lateral shift,” Sens. Actuators Phys., vol. 145–146, pp. 371–379, 2008.
- [WU 2010] Wu Y, Zhang C, Ding G, “Vertically bidirectional bistable microrelay with magnetostatic and thermal actuations,” in 2010 IEEE 23rd International Conference on Micro Electro Mechanical Systems (MEMS), pp. 1139–1142, 2010.
- [XIANG 2012] Xiang H, Komvopoulos K, “Evolution of sidewall adhesion in surface micromachines due to repetitive impact loading,” J. Appl. Phys., vol. 111, no. 5, p. 054507, 2012.
- [XU 2014] Xu N, Sun J, Chen I R, Hutin L, Chen Y, Fujiki J, Qian C, Liu T J K, “Hybrid CMOS/BEOL-NEMS technology for ultra-low-power IC applications,” in 2014 IEEE International Electron Devices Meeting, pp. 28.8.1-28.8.4, 2014a.
- [YAN 1998] Yan W, Komvopoulos K, “Contact analysis of elastic-plastic fractal surfaces,” J. Appl. Phys., vol. 84, no. 7, pp. 3617–3624, 1998.
- [YEAP 2013] Yeap G, “Smart mobile SoCs driving the semiconductor industry: Technology trend, challenges and opportunities,” in 2013 IEEE International Electron Devices Meeting, pp. 1.3.1-1.3.8, 2013.
- [YU 2016] Yu S, Chen P, “Emerging Memory Technologies: Recent Trends and Prospects,” IEEE Solid-State Circuits Mag., vol. 8, no. 2, pp. 43–56, 2016.
- [ZAGHLOUL 2013] Zaghoul U, Piazza G, “10–25 NM piezoelectric nano-actuators and NEMS switches for millivolt computational logic,” in 2013 IEEE 26th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 233–236, 2013.
- [ZAGHLOUL 2014] Zaghoul U, Piazza G, “Sub-1-volt Piezoelectric Nanoelectromechanical Relays With Millivolt Switching Capability,” IEEE Electron Device Lett., vol. 35, no. 6, pp. 669–671, 2014.
- [ZAMBELLI 2015] Zambelli C, Grossi A, Olivo P, Walczyk C, Wenger C, “RRAM Reliability/Performance Characterization through Array

Architectures Investigations,” in 2015 IEEE Computer Society Annual Symposium on VLSI, Montpellier, France, pp. 327–332, 2015.

[ZHA 2018]

Zha Y, Li J, “CMA: A Reconfigurable Complex Matching Accelerator for Wire-Speed Network Intrusion Detection,” IEEE Comput. Archit. Lett., vol. 17, no. 1, pp. 33–36, 2018.

[ZHOU 2007]

Zhou Y, Thekkel S, Bhunia S, “Low power FPGA design using hybrid CMOS-NEMS approach,” in Proceedings of the 2007 international symposium on Low power electronics and design (ISLPED ’07), pp. 14–19, 2007.

[2013 ITRS]

“2013 ITRS,” International Technology Roadmap for Semiconductors. <http://www.itrs2.net/2013-itsr.html>.