POLITECNICO DI TORINO

Master degree course in Electronic Engineering

Master Degree Thesis

Molecular Nanocomputing: an Engineering Approach from Physics To Circuit Architectures

Supervisors:
Prof. Mariagrazia Graziano
Prof. Gianluca Piccinini
Ph.D. Yuri Ardesi

Candidate:
Chiara Elfi Spano

Ottobre 2020

To those who helped me learn.

The earth never tires,
The earth is rude, silent, incomprehensible at first,
Nature is rude and incomprehensible at first,
Be not discouraged, keep on, there are divine things well envelop'd,
I swear to you there are divine things more beautiful than words can tell.

Walt Whithman

Summary

In the last 50 years *Moore's law* has fueled the develompment of the computing field. Nowadays, skyrocketing fabrication costs and increasing demand for miniaturization, high performance and low power consumption are becoming incredibly challenging, pushing the research effort towards Beyond-CMOS technologies and innovative architectural solutions.

In this scenario, molecular electronics paves the way for smart and cheap technological processes based on self-assembly. The possibility to produce artificial molecules with nano-scaled dimensions and engineered properties, is really attractive, since enables extremely high integration with fewer power issues accomplishing a variety of electronic tasks, including conducting wire, rectification, memory, sensor and switching only by simply changing the type of molecule.

This thesis work investigates molecular nanocomputing solutions based on electronic transport, spacing from physical level to the application abstraction level, with constant attention devoted to providing a physical insight from an Electronics Engineer standpoint.

The first part of this work deepens at the very physical level of abstraction in quantum transport models and theory of conduction at nanoscale. With the aim to provide a practical understanding, theory's outcomes are treated in a "ready-to-use" manner.

The second part is devoted to the device level of abstraction: after a preparatory theoretical part on molecular devices, two different molecules are carefully engineered in a device configuration and characterized by means of ab-initio simulations. Proper figures of merit are extracted.

Finally the aim of the last part is to investigate at architectural level how to overcome the limitations of current computational systems by exploiting the possibilities offered by molecular technology and unconventional architectures. Firstly a molecular implementation of logic (logic gates, Half- Adder, Full-Adder) and memory (SRAM) elements is performed, and then a target Logic-in-Memory application is discussed and implemented. Design, functional verifications, performances analysis are performed in Cadence Virtuoso® both for molecular ambipolar transistors

and FD-SOI MOSFET (28 nm) technologies in order to be compared and demonstrate the benefits and problems, for these specific applications, in migrating from conventional to a molecular technology.

At last the stated goal of the thesis is achieved. It reaches an adequate awareness of how nature of molecular technology is often very different from CMOS established one. As a consequence conventional architectural paradigms cannot really do justice of its overwhelming potentialities. Therefore possible alternative solutions from state-of-art literature are proposed, in conclusion, in view of seeking out a deeper understanding of physics of transport and storage inside molecules and *boldly go where no silicon-based computing system has gone before!*

Contents

Summary v

1 Introduction 1
1.1 Evolution of electronic industry . . . 1
1.1.1 A bit of history . . . 1
1.1.2 Current drivers, limitations and challenges . . . 4
1.2 Molecular electronics: why? . . . 8
1.3 In this thesis: goals and outline . . . 10

I Physical level **12**

2 From microscale to nanoscale 15
2.1 Transport regimes: basics . . . 15
2.1.1 Semi-classical transport . . . 15
2.1.2 Quantum transport . . . 18
2.2 Modeling conduction at nanoscale . . . 30
2.2.1 Density of States (DOS) . . . 31
2.2.2 Conduction in 0D-systems . . . 40
2.3 NEGF formalism: basics . . . 55

II Device level **59**

3 Two-terminals molecular device 61
3.1 State-of-art . . . 62
3.2 Molecular wire: basics . . . 63
3.3 Conduction in molecular wire . . . 67
3.3.1 Simplified modeling based on Landauer formalism . . . 68
3.3.2 Dependence on type of molecule . . . 73

3.3.3 Dependence on the length of molecular channel 80
3.3.4 Dependence on the torsion 83
3.3.5 Dependence on the type of linker 86

4 Three-terminals molecular device 90
4.1 State-of-art 91
4.2 Conduction in MolFET: gating effect 92
4.3 Ambipolar MolFET 94
4.3.1 Methodology and atomistic simulations 95

III Architectural level **105**

5 Circuital modeling of MolFET 107
5.1 From device to circuits 107
5.2 Enriched LUT-based model 110
5.2.1 Functional and static analysis 113
5.2.2 Dynamic performance analysis 117

6 Logic implementation: logic gates, HA, FA 129
6.1 CMOS Static Logic family: a review 130
6.2 MOSFET-based Logic circuits 133
6.2.1 Design and testing 133
6.3 MolFET-based Logic circuits 144
6.3.1 Design and testing 144
6.4 Analysis and comparison 158
6.4.1 Area occupation 158
6.4.2 Power dissipation 159
6.4.3 Speed 164
6.4.4 Conclusions 165

7 Memory implementation: SRAM 167
7.1 SRAM review and metrics 168
7.1.1 Read operation 171
7.1.2 Write Operation 172
7.1.3 Cell robustness 175
7.2 MOSFET-based 6T SRAM 176
7.2.1 Design 176
7.2.2 Testing 181
7.3 MolFET-based 6T SRAM 188
7.3.1 Design and testing 188

8 LIM cell implementation 195
8.1 in-Memory-Computing paradigm . . . 196
8.1.1 Drivers . . . 196
8.1.2 Concept review . . . 197
8.1.3 LiM approach . . . 198
8.2 LiM cell . . . 199
8.2.1 Algorithm study . . . 199
8.2.2 Hardware implementation . . . 201
8.3 MOSFET-based LiM cell . . . 203
8.3.1 Design . . . 204
8.3.2 Testing . . . 208
8.4 MolFET-based LiM cell . . . 210
8.4.1 Design and testing . . . 210

9 Conclusions and outlook 212

A Atomistic simulations in QuantumWise ATK ® 214

B Modeling molFET in Cadence Virtuoso® 226

Bibliography 238

Acknowledgements 245

Chapter 1

Introduction

The aim of this introductory chapter is to discuss the current technological scenario within which this master thesis is placed. In section 1.1, a brief review of electronic industry, of its evolution and current market drivers, limitations and near and far-terms challenges is reported. Finally in section 1.2 the molecular technology with its strengths, subject under investigation of this work, is presented as a promising candidate able to face the above discussed challenges.

1.1 Evolution of electronic industry

1.1.1 A bit of history

An insatiable desire by the masses for consumer electronics and the astonishing developments in microelectronics have fueled over the past five decades the semiconductor industry at peerless rates, definitively changing our lifestyles.
Nothing that is of importance to mankind has been left untouched, from communication and entertainment to medicine or science. Even in the automobile, one of the most mechanical creature of modern technology, electronics is now responsible for 40 percent of its cost [1].

But taking a step back, the very dawn of electronics can be traced back to the invention of **vacuum tube** at the beginning of the 20th century. It made possible signals manipulation, crucial to the development of radio, television, radar, long-distance telephone networks, and above all in 1945 to the first general purpose electronic digital computer: the ENIAC. It contained about 18000 vacuum tubes, which were fragile, very power-demanding, and produced huge amount of heat. Moreover the ENIAC occupied about $170\,m^2$ of floor space, thus it was tremendously bulky. It was therefore inconceivable at that time to imagine that

one day electronic engineering would have dealt so heavily on artificial structures, thousands of times smaller than the diameter of a human hair.

Anyway the **transistor** made its appearance in 1948 and it was a breakthrough. Indeed it provided the same electrical function of the vacuum tube but at a reduced cost, weight, size, and power consumption and with higher reliability.
About ten years later, the feasibility of integrating transistors with resistors and capacitors on a single semiconductor chip was proved, leading together with the invention of the MOSFET to the IC era. Being small in size and light in weight, the **IC** was widely used in a variety of electric appliances and it even helped to make possible the first human space flights of the 1960's!

In the late 1960's, many scientists daydreamed the possibility of a computer on a chip, but integrated circuit technology was ready to support such a chip only in 1971, when an Intel team developed such an architecture with just 2300 transistors in an area of only 3x4 mm. It was called the **4004 microprocessor**. With its 4-bit CPU, command register, decoder, decoding control, control monitoring of machine commands, the 4004 was a great invention, used to build the first hand-held calculator, and exploited in the *Pioneer 10*, the first spacecraft to enter the Asteroid Belt. Nonetheless, it led to the birth of the personal computer, an ubiquitous device in our every day life. Anyway, since IC first introduction up to

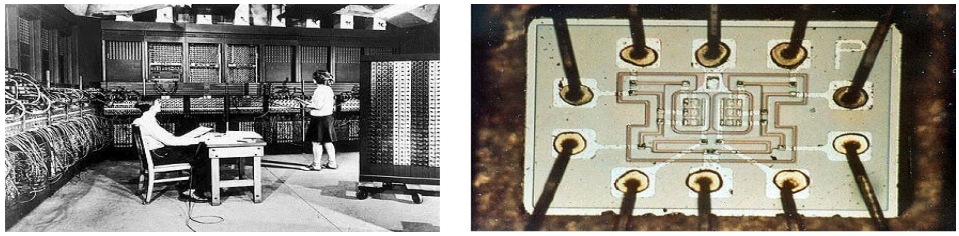


Figure 1.1: On the left: a photo of the ENIAC. On the right: the NOR gate chip implemented with RTL (Resistor-Transistor Logic) used to build Apollo guidance computer.

now, semiconductor industry has grown by leaps and bounds, ruled by the famous trend known as *Moore's Law* which postulates that the number of transistors on IC doubles every eighteen to twenty-four months.

From the Small Scale Integrated (SSI) circuit of 1964 which counted few transistors to the Medium (MSI, 1968) and Large (LSI, 1961) Scale Integrated circuits, until the Very Large (VLSI) and Ultra-Large (ULSI) ones in 1980 and 1984, which counted from 100 thousand to more than 10 million, this empirical prediction has been proved quite evident. Just to have an idea: nowadays Samsung's 1 TB embedded flash memory (eUFS) chip based on Vertical-NAND (V-NAND) technology counts 2 trillion floating-gate MOSFETs [2] and non-memory chip as `Wafer Scale`

`Engine` by Cerebras [3], a deep learning engine, has 1.2 trillion MOSFETs, manufactured using TSMC's 16 nm FinFET process.

Moore's law which implies restless miniaturization of transistors (i.e. higher integration density in a chip) actually is driven by two sound business reasons related to technology:

a) a larger number of IC can be produced per wafer, thus increasing the profits. Of course, this holds true only if at the same time the chips per wafer yield is also maintained;

b) more functionalities can be packed in the same sized die as before, enabling manufacturers to fabricate and sell more capable ICs that get higher prices.

Another way the industry has been raising its profits is through the use of larger wafers. It does not seem obvious since processing larger wafers is more expensive both because it requires new or modified tools capable of handling them and because it requires more processing chemicals to complete the wafer fabrication process. Nevertheless, large wafers have a favourable impact on industry revenues because it means significantly more devices than smaller wafers with an improvement of throughput and overall cost reduction. The current state-of-the-art fab produces wafers of 300 mm (referred to as "12 inch"), that is a standard in leading fabs from 2002 and there are proposals to adopt 450 mm in the near future [4].

These were the factors that caused semiconductor companies to invest heavily in new fabrication techniques and integration of new technologies and materials in order to improve in a IC, both the Front End Of Line (FEOL, technological process at transistor level) and Back End Of Line (BEOL, technological process at interconnections level). This enabled them to keep migrating from one technological node generation to the next, preserving computational performances, reliability, a bearable value of power consumption and a certain degree of compatibility w.r.t. previous generation.

Indeed, concerning FEOL, from '80s to around 2000, CVS (Constant Voltage Scaling) policy applied to a standard bulk CMOS technology marked the "Golden Age of scaling", allowing to reach maximum speed with a strong compatibility with previous nodes. Anyway a point of no return was reached because of unacceptable values of power density and, despite the adoption of different scaling policies (CFS, Constant Field Scaling and Generalized Scaling) and the many technological innovations introduction that shifted standard Bulk CMOS process to advanced one (HKMG, LDD, silicides, conductive spacers, retrograde doping profiles, halo implants, strain silicon technologies...) it was not still enough to effectively go forward next technological nodes, overcoming SCEs (Short Channel Effects) and sub-threshold currents problems. A reassessment of active devices structure was

necessary to continue Moore's Law, that's why UTB (Ultra Thin Body) technologies were introduced. Nowadays they have totally replaced Bulk MOSFET technology. They enabled again effectively down-scaling from 2011 so far, since were able to provide a better electrostatic gate control of the channel and to limit leakage paths, thus keeping under control static power consumption. On UTB Multi-Gate FinFET technology are based the technological nodes from 14 nm [1](2014) to 7 nm (2016). The 5 nm node is on the way of release by the end of 2020, and Samsung recently announced the production of the world's first 3 nm GAA-FET (Gate All-Around) process prototype, saying that it is targeting mass production in 2021.

Finally, concerning BEOL, unfortunately its scaling trend differs a lot from FEOL for obvious physical reasons related to electrical parameters degradation. Nevertheless many achievements were reached thanks to hierarchical scaling, and to huge technological breakthroughs such as Cu technology and Dual Damascene process (to reduce the metal lines resistance per unit length), and advanced low-k dielectrics (for capacitances per unit length).

1.1.2 Current drivers, limitations and challenges

In order to take a stock of the situation, nowadays electronic industry scenario can be summarized with the following key-points from the 2020 edition of the *IEEE International Roadmap for Devices and Systems* (IRDS) [5]:

- the health of the industry keeps on, thanks to constant growth of consumers and number of shipped units.
- Semiconductor R&D investment grows.
- *Moore's Law* tirelessly carries on. Indeed logic products with a minimum metal pitch of 36 nm will be introduced in 2020.
- EUV (Extreme Ultraviolet) lithographic finally appeared into high-volume manufacturing in 2019. It allows to realize lines and spaces of 16 nm with a

[1]Please pay attention that this quantity from 90 nm technological node so far actually does not correspond to any meaningful and measurable quantity related to transistor density on the wafer, neither no more to gate length or FIN width! It's only a silicon manufacturers label. Indeed in the previous nodes the MOSFET's channel lenght and half pitch of the tightest metal layer (i.e. the original definition of node according to NTRS/ITRS) almost coincide, but from 90 nm the geometrical scaling was no more followed and substituted by equivalent scaling, which consisted on practically keeping unchanged the channel length and introducing for example strain to enhance mobility of carriers. Anyway for more interested readers: section 1.2.2 of the 2020 IRDS Executive Summary [5] clarifies how the industry nomenclature has led nowadays to a complete detachment between IC features and technology nodes' names. Indeed, there are companies nowadays announcing the advent of technologies below 1 nm in this decade... clearly insane! Moreover the new node broader definition that IRDS has adopted is reported.

single exposure, leading to a drastic reduction of required number of masks to satisfy design goals.

- IC functionalities will continue to increase for at least 10 more years thanks to new architecture paradigms which exploits both vertical monolithic and heterogeneous integration.
- Neuromorphic architectures supported by Artificial Intelligence and Machine Learning (AI/ML) will keep on in revolutioning how computers operate.
- R&D in Beyond-CMOS devices continues to successfully proceed, thanks to constant improvements of several innovative devices (stacked transistors, tunnel-FETs..).
- Quantum computing (QC) is a newcomer, promising unprecedented increases in functionality.
- the "More than Moore"(MtM) approach is allowing the non-digital functionalities (e.g., RF communication, power control, passive components, sensors, actuators) to migrate from the system board-level into the package (SiP) or onto the chip (SoC). Moreover, the integration of CMOS-based system on chip and non-CMOS based system in package (SiP) technologies within a single package would become increasingly important, leading to the future of heterogeneous integration.
- The first 5G phones are in the market!
- For the first time in years the PC industry saw a sudden demand rise in 2020, due to millions of people suddenly forced to work at home because of Covid-19. Moreover video conferencing exploded.
- The number of data centers has kept on increasing at higher rates for the past 5 years.

The above reported key-points highlight the main market drivers that are and will force the advancement of the electronic industry. From them, a big specific need is evident: make people interact, entertain, work between each other and between devices in a faster, and more effective way at lower cost. This means also move and store a huge amount of data. To satisfy this market requirement a restless race for higher integration, functionality, and performances at ultra low power and cost terms is on-going.

The critical problem, in modern days, is that these features are orthogonal: for e.g. higher integration typically means a more power demanding system, designed and realized at higher costs. Indeed, the golden era in which these needs were easily satisfied are abundantly finished, and many limitations are to be faced or

smartly by-passed. That is why this race is requiring more and more enormous efforts and skyrocketing R&D and fabrication costs.

Currently, three main limitations can be outlined: (A.) the so called "Silicon brick Wall"; (B.) the "Power Wall"; and (C.) the "Memory Wall"; They are briefly discussed in the following.

(A.) The Silicon-brick Wall. This type of wall is related to transistor miniaturization. The reasons behind this limit are basically of three types: economical, technological and physical. Concerning the economical reason, the issue is the following. Fifty years of aggressive scaling have pushed the device dimensions close to the atomic range, making the manufacturing process based on top-down lithographic approach, characterized by very high costs, no more affordable. Indeed, modern ICs are becoming more and more engineering marvels requiring hundreds of careful processing steps performed in ultraclean environments. This led to a situation in which the reduction of the transistor size no longer implied the reduction of the cost per unit. That is why only three ICs semiconductor foundries are now survived (TMSC, Samsung, Intel) in the further step from one technological node to the next one. On the other hand, there is a technological reason. It is related to the effort spent in conceiving new reliable and effective nanodevice geometries (as much as possible immune by short-channel effects (SCEs), detrimental quantum effects and leakage currents), and in the development of technologies and lithography tools able to pattern increasingly smaller line and space widths in silicon. By the way, both the 2020 IRDS and ASML predict that the ability to reduce these features would reach final limits around 7-8 nm by the end of this decade [5]. At last, the third reason: the physical one. Even if the previous issues will be somehow overcome, there will still be a fundamental physical barrier that will limit hopelessly the scaling trend, and so the *Moore's law* as traditionally is intended. There will be inevitably a no-return point at which silicon devices could not overcome physical science barriers and become smaller than one atom! Therefore, to further push forward the *Moore's law*, an equivalent scaling approach must be adopted instead of the traditional geometrical one. It consists on making possible to play an increased role in performance increase and power and cost reduction thanks to the boost provided by innovative paradigms and ideas in devices, architectures, packaging techniques, etc. Typically researchers, outline in this direction, two main paths: Beyond-CMOS and More-Than-Moore. The first consists on exploring and developing completely different type of electronic devices, also profoundly different from transistors in both their physical structure and their operating principles. Examples are Tunnel-FETs, Carbon-Nanotube FETs (CNTs), spintronics devices, Field-Coupling Nanocomputing (FCN) nanodevices, Molecular-FETs, and so on. The latter consists on increasing functionality thanks to new architectural paradigms which exploits both vertical monolithic and heterogeneous integration.

In any case, even if it would be absurdly possible to scale endlessly down transistor dimensions, it would not be beneficial at all, due to unbearable power density increase on a chip. This implication leads to the following "wall".

(B.) The Power Wall. This wall is related to nowadays inability to significantly reduce dissipated power and above all ICs power density. The power problem is a consequence of the breakdown of *Dennard's scaling rule*, which states that smaller transistors leads to higher performing circuits but keeping the power density constant. The breakdown occurred when static power consumption suddenly became a major problem due to sub-micrometer transistor scaling. Indeed below the micron barrier they showed their true nature: they are never truly off and leakages become really relevant. This enormous increase of static power as collateral effect of transistor down-scaling, led to a complete stop of the power benefits engineers took for granted for many technology generations. The operating frequency trend slew down and stopped to few GHz in the last years. Nowadays, it has become increasingly difficult to dissipate the heat generated by a CPU running at high speed, because the more transistors are packed into a chip, the greater the power density that must be dissipated. The power density of modern CPUs is approximately 150 W/cm^2. Just to have an idea: the power density at the surface of the sun is approximately 6000 W/cm^2 which is radiated by heating itself to 6000 K. However the CPUs must be maintained at approximately room temperature! Therefore the heat load of CPUs has pushed fan coolers to practical limits, but beyond air cooling there is water cooling, that may be capable of remove several hundred Watts from a 1 cm^2 sized chip at more expensive costs. Anyway beyond water cooling, there is no yet known solution, thus the problem is serious and the only practical solution that was and it is still adopted up to now in order to avoid the burn of the chip is to keep switched off the majority of the transistors on a chip. This strategy essential to satisfy power constraints, nullifies by the way, the big effort in transistors miniaturization of technologist and device engineers. In conclusion power dissipation is the most visible problem that led the electronics industry in the so called "dark silicon" era. This issue cannot be simply solved by the multicore revolution due to inefficiencies that appear when the number of cores scales (*Amdahl's law*). That is why a great effort is now devoted in conceiving new design techniques at circuital, architectural and even at software level in order to reduce as much as possible power dissipation in electronic systems.

(C.) The Memory Wall. It is related to the discrepancy in Von-Neumann computer architecture, between logic and memory speed. Indeed the memory represents a bottleneck for the operating-frequency of a system because is not able to provide data as fast as CPUs are able to compute them. This is simply due to the historical different rates of progress between processing units and memories:

speed of logic circuits doubles every 2 years, according to Moore's law, while memory technology increases its speed at much slower rate. Indeed, basically, current memory technology can be vast or fast, but not both at the same time. From an architectural standpoint, it is possible to reduce this gap exploiting the hierarchy of memories, by making closer successively smaller but faster memories toward the CPU. Frequently used or useful data are automatically transferred from slow memories to faster ones. This provides, most of the time, an illusion of a larger, faster and cheaper memory system. Moreover, this strategy has implications also on power consumption: since it brings useful data closer to the CPU reduces data transfers over long wires that have to be charged/discharged. However, this trick cannot be a definitive and universal solution because the memory gap is increasing exponentially. Furthermore, there are several applications where the principle of data locality does not hold: sparse matrix algorithms, sorting, data mining, and, in general, every situation where we need to go over vast data sets performing very little computation. Therefore innovative solutions which no more rely on separation between logic and memories are becoming more and more popular at research level and are also briefly discussed in III, chapter 8.

These strong limitations in order to be overcome need revolutionary change of perspectives both at device and architectural levels. Anyway, in order to be accepted by the market need to have cheaper fabrication costs and manufacturing process compatible enough with the previous traditional generations. There are several ideas and implementations at research level that may overcome the above discussed walls, at both levels of abstractions. There can be solutions more or less feasible and reliable, each one with its own trade-off in terms of R&D and fabrication costs, expected performances, power density, speed and so on. Among several, the path of molecular electronics, topic of this thesis, could be a promising viable solution. A brief explaination of the reasons why is reported in the following section.

1.2 Molecular electronics: why?

As previously outlined, there are mainly three limitations in the current electronic industry scenario. Molecular-electronics can, in principle, be a innovative alternative in solving the first two issues: technological and economical difficulties in miniaturization and power dissipation. In the following is briefly discussed why and how.

The dawn of molecular electronics can be traced back to the early 1970s. Since then this field attracted many researchers around the world for its intriguing idea which consists on exploiting single or small groups of molecules in device-based structures, as the fundamental units for electronic components. In other words, it aims at using individual molecules, which constitutes the smallest stable structures

imaginable [6], [7], to perform functions in electronic circuitry now performed by traditional solid-state, and thus to complement or partly replace them in the future [8].
Despite the constant interest by scientific community on this field, it is only in the past two decades that significant experimental and theoretical barriers have finally been overcome mainly thanks to the introduction of scanning tunneling microscope (STM)[9] and then of the atomic force microscope (AFM)[10]. Experiments with single-molecule junctions, have become more robust, reliable and reproducible [11]. Meanwhile, theoretical methods based on Non-Equilibrium Green's functions (NEGF)[12] have been developed, allowing researchers to investigate the fundamental physics of single molecules under non-equilibrium conditions and to discover several novel effects. At the same time, a variety of potential practical applications have also been proposed, extending beyond the electronic devices predicted in the early days of the field. Devices such as wires [13], transistors [14], [15], [16], diodes [17] [18], storage element [19], sensors and so on.

Beside the evident and great advantage in device **integration** [20], [21] that may lead to extend the Moore's law to unforseen limits, thanks to molecule dimensions ranging around few Angstrom (hundreds of times smaller than the smallest features obtainable with semiconductor technology), the actual revolutionizing feature of molecular electronics concerns the manufacturing process which it exploits. The problem with modern ICs is that they are engineering marvels, requiring hundreds of careful processing steps performed in ultraclean environments and by means of million dollars lithographic instrumentation. This top-down approach had become economical unsustainable for many fabs, so that only three semiconductor ICs are nowadays survived. A change of perspective can be provided by nature: a bottom-up approach based on ***self-assembly*** [22] consists on letting nanodevices and circuits directly self-assemble from their molecular constituents. Therefore microscopic components are assembled together to build complex macroscopic systems: this is exactly the opposite to the traditional top-down approach aiming at down scaling the sizes of components that build the systems. This bio-inspired process is really smart, extremely cheap and relatively easy since is a thermodynamically favorable process, i.e. it is energetically favorable for the interacting entities to form some organized aggregate structure. This means also that is a self-extinguished process and so there is no need to be controlled or monitored: it is enough to suggest the direction of assembly by means of chosen environment parameters (like temperature, pressure, type of substrate...) and the costituent parts in the process create their patterns assembling themselves.
Self-assembly may lead to a substantial reduction of fabrication costs, paving the way to a new era of nano-electronics, characterized by the possibility to produce artificial molecules with nano-scaled dimensions and engineered properties.

The second breakthrough of molecular devices is that is a promising technology not only for highly dense but also **low power** computing applications. Indeed

currents at stakes are very low, maximum of the order of few µA. Also the intrinsic device capacitances to be charge/discharge in order to switch states in logic circuits, are of the order of few aF, differently from current CMOS ones that are of the order of few fF, three order of magnitude greater! As a consequence the dynamic power at stakes are much lower.

Another great advantage of molecular electronics is the huge conformational **flexibility** in functionalities of devices [8] that can be engineered and controlled through synthetic chemistry also simply changing the type of molecule in the channel, or the surrounding components such as the reservoir electrodes or anchoring groups. The huge diversity at disposal may also lead to the emergence of distinctive functionalities and the discovery of new physical phenomena (NDR, Kondo effect, QI...) [23], [24], [25], [26], typically not accessible with traditional materials/approaches.
However on the other hand, even if molecular electronics is born more than five decades ago, is a discipline that is still at an infancy level, due to the difficulties in manipulate entities of the order of Angstrom like molecules, in measure significant parameters and in support process variations that are really relevant especially in the molecular synthesis phase.

In conclusion, molecular devices can play an important role in emerging future nanoelectronics enabling extremely high integration with fewer power issues at very cheap fabrication costs, accomplishing a variety of electronic tasks, including conducting wire, rectification, memory, sensor and switching via chemical synthesis.

Concerning the future of this field an its eventual commercial introduction a reminder from one of the pioneer of molecular electronics, James Tour, has to be considered [27]: “the advances in the semiconductor industry came through Herculean efforts involving thousands of person-years of work and trillions of dollars of investment, hence any direct frontal assault by a new technology on the semiconductor industry will fail”. Indeed,in his book (“Molecular Electronics”, 2003) he asserts that an introduction of molecular devices into the market is possible if the fundamental and financial barriers the semiconductor industry is facing will drive it to consider other options and if a specific insertion and investment strategy plan is adopted. Indeed, most companies spend four times more on cost reduction programs of present products than R&D costs for developing new products. New technologies and techniques would be welcomed only if they reduce the cost of an existing product still maintaining or increasing the product’s performance.

1.3 In this thesis: goals and outline

This thesis is organized in three parts in addition to the introductory and conclusive chapters. Each part is related to three different abstraction levels ordered from the bottom level, the physical ground to the top one, the architectural abstraction

level, passing through the device one.

- Part I provides the theoretical background for the thesis. In particular, in chapter 2 a deepening into quantum transport is reported. The aim is to to provide a "ready-to-use" insight useful at the higher abstraction level: the device one.

- Part II provides basic concepts useful to investigate, engineer and finally characterize molecular devices. In particular in chapter 3 some basic concepts related to molecular wires are presented from a theoretical and simulative points of views, providing also some experimental and fabrication processes related overviews referring to the state-of-art literature. Then in chapter 4 basic concepts of conduction in molecular transistors, preparatory for engineer and characterize new devices with atomistic simulations, are reported. Finally a full characterization aided by atomistic simulations performed with QuantumWise ATK software [28] of two ambipolar molecular transistors, engineered ad-hoc for digital computing applications, is discussed.

- Part III provides firstly in chapter 5 the basics to understand the circuital modeling for molecular transistors: some possible methodologies that could be adopted in order to assemble more than one molecular device in a more complex circuit, are discussed. Then in the next chapters 6, 7 8, logic (logic gates, Half- Adder, Full-Adder), memory (SRAM) elements and a target Logic-in-Memory application are implemented with molecular transistor technology, exploiting the results of the previous part. Design, functional verifications, performances analysis are performed in Cadence Virtuoso® both for molecular ambipolar transistors and FD-SOI MOSFET (28nm) technologies in order to be compared and demonstrate the benefits and problems, for these specific applications, in migrating from conventional to a molecular technology.

Part I

Physical level

Conventional electronic devices operate by controlling the flow of carriers through a channel. Over the years, the basics of semiconductor physics were clarified and oversimplified into the operative knowledge of device engineers. Usually they have concerned themselves only with macroscopic description of devices and the complexities of microscopic have been hidden in the macroscopic parameters like mobility, diffusion coefficients and lifetime etc. The interpretation of electrons and holes as semiclassical particles with an effective mass has resulted to be usually adequate. For most devices, the drift diffusion model provided a simple and adequate description of carrier transport. Also in typical Electronic Engineering (EE) undergraduate courses there has never been much incentive from an applied standpoint to understand how macroscopic parameters are calculated from first-principles.

However, today things are changing: electronic devices are approaching the molecular scale, the materials properties can be engineered by strain or sizing effects thanks to quantum confinement. Devices contain a countable number of carriers and dopants and are sensitive at the atomistic scale to the structure. Moreover, in addition to conventional devices like MOSFETs, which have been downscaled to nanometer sizes, new devices based on carbon nanosheet or nanotubes, or on organic molecules are being investigated. Therefore, in this scenario, a deeper exploration of quantum theory has thus become not only important even for the device engineer, but also a necessity for facing present and future challenges in nanoelectronics that is actually requiring more and more interdisciplinarity. A change of perspective is needed. To describe carrier transport in nanoscale devices, engineers have to think about charge carriers as quantum mechanical entities rather than as semiclassical particles, and they should learn how to create their knowledge's scenario at the atomistic scale rather than at a continuum one.

Moreover, more in particular, concerning the field of molecular electronics, features like conductance switching, rectification and negative differential resistance (NDR) and many novel effects have been ascertained; thus, recently a lot of research is focused on designing molecules with specific functionalities. In this field, it has become essential to model and understand the quantum physics that dominates the properties of such extremely small devices. The usual methods exploited in investigating the behavior of molecular devices, are the Non-Equilibrium Green's Function (NEGF) formalism, the Density Functional Theory (DFT), and some Semi-Empirical methods like the one based on the Extended-Hückel theory (EHT). Designing molecular devices is only possible after having a complete theoretical model that can explain and predict the charge transport behavior through molecular devices.

Therefore, in this perspective, the aim of the first part of this thesis is to provide a *practical* overview on how to model conduction at nanoscale without loosing into mathematical formalisms and thus forgetting the physical insights. Indeed, the recipient of this treatment is intended to be a typical Electronic Engineering master student facing with nanoelectronics for the first time. However notice that

a certain degree of accuracy and analytical rigor is still kept.
The aim is to to provide a “ready-to-use” insight useful at the higher abstraction level: the device one.

Chapter 2

From microscale to nanoscale

In this chapter a simplified treatment of quantum transport modeling, mainly inspired by [29], [30] is reported. Firstly, in 2.1.1, basics concepts of conduction in conventional micro-structures (bulk-systems or 3D-systems) are briefly reviewed in order to highlight the differences w.r.t. the nano-structures (2D, 1D, 0D -systems), on which the analysis is focused in the next sections.
This chapter is intended to be a comprehensive and simplified treatment on how to model conduction at nano-scale with a focus on 0D-systems. In order to pursue this aim the following outline is adopted. The Density Of States (DOS) are firstly derived for each structure in order to obtain the density of carriers available for conduction and thus finally the current as function of bias. The latter is derived in subsection 2.2.2. Then a more advanced formalism based on Non-Equilibrium Green's Functions (NEGF) on which are based the majority of atomistic simulators, is very briefly described.

2.1 Transport regimes: basics

In this section, firstly a review of the semi-classical description of carriers is provided. Then the reasons why this traditional approach no more holds at the nano-scale, are explained. Next, semi-classical ballistic transport is discussed, and finally basics of quantum transport are reported.

2.1.1 Semi-classical transport

In bulk-systems (a.k.a. 3D-systems), electrons are treated as semi-classical particles that flow through a channel under the influence of an electric field and randomly scatter within the crystal lattice. Until these assumptions hold true (i.e.

interpretation of carriers as semi-classical particles and presence of many scattering events), the drift-diffusion model is still adequate enough. Differently, in nano-structures, both of these assumptions loose of validity and it becomes necessary to include quantum mechanical phenomena. More, in particular, in these traditional bulk-systems, a semi-classical treatment holds mainly because of three feautures:

- **Periodicity of the crystal lattice.** Thanks to the large scale dimensions of the bulk-systems, the crystal lattice can be considered, with a very good approximation, periodic in all directions (x, y, z) (figure 2.1). Indeed, in these systems, the typical sizes of the active part of the device (i.e. the channel) that is responsible for the conduction, is ranging between few µm to hundreds of nm.

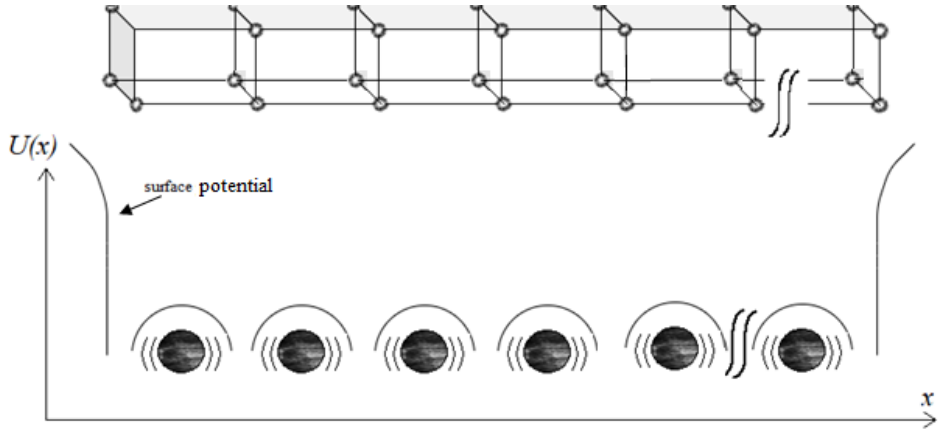


Figure 2.1: The 1D potential energy of the crystal lattice, periodic function of the space. It results from the periodic placement of the ions lattice within the crystal. The distance between one ion and the next is the lattice constant a.

  This periodicity is deeply exploited to model the behaviour of carriers and obtain overall important parameters of the device, such as resistance of the channel (R_{CH}) or electron/hole mobility ($\mu_{e/h}$). In other words, in bulk structures, average quantities of the whole carriers population are exploited to model transport. This is different from nano-scaled devices (1D, 0D - systems), in which is a single electron to give an individual contribution to the conduction.
  Therefore in bulk-systems, for e.g. electrical conduction, dicteted by *Ohm's law*, can be written by means of macroscopic quantities like the electrical

conductance G. In particular, for a rectangular conductor with area cross-section A and length L, is defined as:

$$G = \frac{\sigma \cdot A}{L} \tag{2.1}$$

where $\sigma = qn\mu_n$ is the electrical conductivity of the material with n electron density, q the electron charge, and μ_n the electron mobility. The electron mobility in turn is defined as $\mu_n = q\tau_n/m_n^*$, with $m*_n$ the effective electron mass and $\tau_n = \lambda_{MFP}/v_{th}$, where v_{th} is the thermal velocity and λ_{MFP} the mean free path (i.e. the avarage distance crossed by an electron before it scatters).

- **Non-coherent transport.** When an electron is injected in the channel from the source electrode (S) and moves towards the drain (D) one, it looses completely its initial information that it acquired in the electrode (energy, phase and direction of the momentum k). That is because in its motion it frequently scatter from various perturbing potentials (due to defects, ionized impurities, lattice vibrations, etc.). This exchange of energy with the crystal lattice after its own mean free path λ_{MFP}, happens in order to always make the electron being in thermal equilibrium with the system and occupying the minimum energetic stable state in the conduction band.

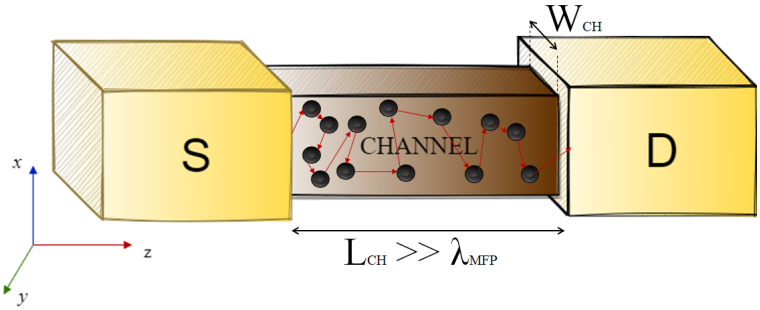


Figure 2.2: A sketch of diffusive charge transport in bulk systems, dominated by scattering events. The size of conductor L_{CH} is much greater than the electron free mean path.

The result is that carriers do not travel in a straight line, but undergo a random walk through the channel when a small bias is applied, hopping from one trajectory to another as shown in figure 2.2 and changing the direction, intensity and sign of their motion.
In the case of bulk-systems, as it is evident from figure 2.2, scattering events

occur because the electron mean free path is much lower than the critical length of the channel ($\lambda_{MFP} << L_{CH}$).

- **Unconfinement.** Electrons and holes are not confined in any direction. They have all the three directions (x, y, z) as degree of freedom, so that they can freely move everywhere without constraints within the material. That is because of the critical sizes of the channel (length, thickness, width) that are very large if compared with the *De Broglie* wavelength of electrons near the Fermi energy λ_{Fermi} (fig. 2.2).

2.1.2 Quantum transport

If nano-structures are considered instead of bulk-systems, then some of the previous assumptions loose of validity and to properly model conduction it becomes necessary to abandon classical concepts and to include quantum mechanical phenomena. Indeed, by shrinking the dimensions of the channel, carriers turn to be confined and their two-fold quantum nature of wave-particles must be taken into account. Remember that the quantum confinement is assured in nanoscaled devices where the λ_{Fermi} (that for Fermi energy of few eV is around 2-3 nm) is comparable with the critical sizes of the conductor. Moreover:

- in the direction of confinement the potential cannot be anymore considered periodic, hence macroscopic quantities like resistance or current expressed by means of average parameters loose of meanings. Therefore important questions arise: how to modify standard transport models based on the so established drift-diffusion, in order to model the conduction at nanoscale? And how these macroscopic quantities will be modeled? A detailed analysis is provided in the next sections, where starting from DOS of the different type of confined systems the current under an applied bias will be derived for 0D-systems.

- in systems where the electrons are confined in 2 or all the three dimensions of space (1D and 0D -systems) the transport becomes coherent or quasi-coherent. This type of regime is also said ballistic because the motion of the carrier is like the motion of a bullet. Indeed, it does not undergo to scattering and move directly from source to drain without changing its properties because of no exchange of energy with the system occurs since $\lambda_{MFP} \simeq L_{CH}$. In other words momentum and kinetic energy are conserved.

- in 1D and especially in 0D -systems, the interfaces between the contact electrodes and the channel becomes essential to model the transport. This issue will be detailed in the following.

In the following a brief overview on the types of nanoscale systems is reported.

- **2D-systems (a.k.a Quantum wells).** They are systems like for e.g. heterostructures, HEMTs, CN sheets and in general all the structures made of many layers assembled in a nanoscaled way, where the thickness of the channel is of the order of few nanometers and so comparable with λ_{Fermi}. Indeed, downscaling towards nanosizes one dimension (for e.g. the thickness as in figure 2.3) leads to the confinement of the carriers in one direction (x), so that they turn to have only two degrees of freedom (y, z).

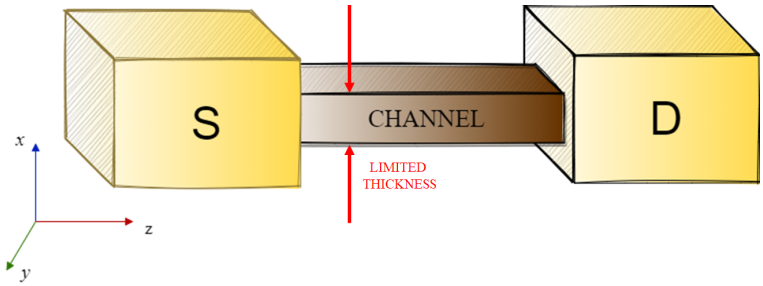


Figure 2.3: An example of 2D-system where the thickness is limited to few nanometers. The direction of confinement is x, a transversal one, orthogonal to the direction of transport (z).

- **1D-systems (a.k.a Quantum wires).** They are systems like for e.g. Nanowires, CNT, where two dimensions are of the order of few nanometers and so comparable with λ_{Fermi}. Indeed, downscaling towards nanometer sizes two dimension (for e.g. the thickness and the width of the channel as in figure 2.4) leads to the confinement of the carriers in two directions (x,y), so that they turn to have only one degree of freedom (z).

  In these type of 1-dimensional conductors, each electron starts to individually contribute to conduction, differently from 3D, 2D -systems.

- **0D-systems (a.k.a Quantum dots).** By scaling all the three dimensions, a complete quantum confinement of carriers is obtained. Now carriers have no degrees of freedom. Typical quantum dots are for e.g. molecules, nanocrystals, etc. In these type of systems, since $\lambda_{MFP} \simeq L_{CH}$, ballistic or quasi-ballistic transport will occur.

  Moreover, in 0D-systems, the quality of the interfaces (S-dot, dot-D) will play a relevant role, deeply influencing the conduction. Indeed, depending on the coupling strengths of the dot between S/D electrodes there will be completely different type of conduction, giving also rise to different type of devices. In particular according to the time interval during which the electron

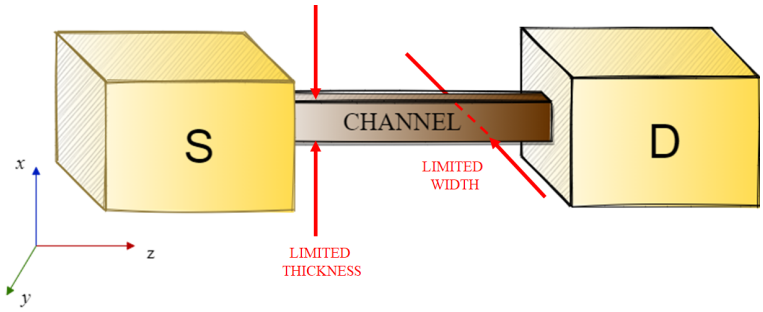


Figure 2.4: An example of 1D-system where the thickness and the width of the channel are limited to few nanometers. The direction of confinement are the transversal directions x and y, orthogonal to the direction of transport (z).

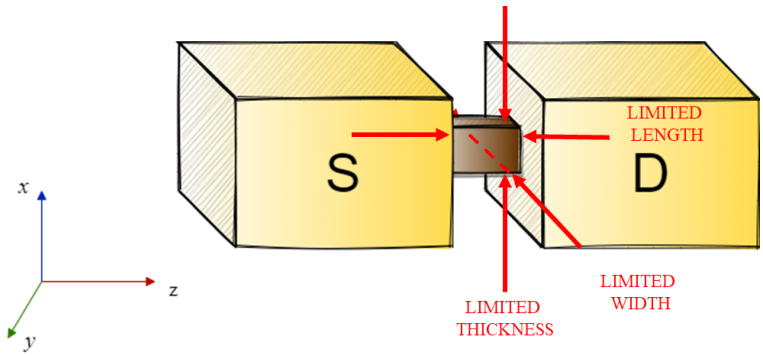


Figure 2.5: An example of 0D-system where the thickness, the width and the length of the channel are limited to few nanometers. The direction of confinement are x,y,z.

stay confined in the dot, called *escape time*, defined as $\tau = \hbar/\gamma$, where γ is the so called *coupling factor*, different types of coupling can be obtained:

- if $\tau \to \infty$: there is no coupling at all between the reservoirs (S/D) and the channel (dot). There is no exchange of matter with the external world and the dot is considered a closed system, in which electrons stay confined forever. Obviously this is a limit case and therefore a non-realistic one.
- if $\tau \simeq$ fs: the system is in the so called *strong coupling* or *Self-Consistent Field* (SCF) regime, where the eigenvalues of the system, i.e. the allowed

discrete energy levels of the dot, are broadened due to the strong influence of the electrodes. So the so called *broadening* of the energy levels Γ occurs and transfer of electrons is easier. Therefore there is exchange of matter with the external world and the dot is considered an open system.

- $\tau \simeq$ns: the system is in the so called *weak coupling* or *Coulomb Blockade* (CB) regime, where the discrete energy levels of the dot are no more broadened due to the very weak influence of the electrodes. No *broadening* occurs and electrons are more confined in the dot and it is more difficult for them to escape.

So, summarizing: the more is the coupling strength γ, the higher is the broadening of the energy levels Γ and at the end the smaller is the escape time τ.

The quantum particle in a box

At nanoscale, a quantum mechanical view of the electron in its wave-particle dualism must be adopted. It is no more interpreted as a particle of finite dimension which scatters with ions of the lattice like a ball in classical mechanics. The uncertainty principle holds and the exact position and velocity of an electron in space cannot be known. A different paradigm is needed: the position of the electron in space is described by a probability density function, the wavefunction $\varphi(x, y, z)$ which can be determined by solving the so well known Schrödinger equation. Schrödinger equation models any quantum system. Given a certain potential energy distribution (frozen in a certain time instant) describing the system U(x,y,z), is it possible to derive the allowed energetic level of the system and the associated wavefunctions $\varphi(x, y, z)$ by solving the stationary Schrödinger equation:

$$\hat{H}\varphi = E\varphi \tag{2.2}$$

with $\hbar = h/2\pi$ ($h = 6.62\text{x}10^{34}$ Js, Plank constant), and $\hat{H}$ Hamiltonian operator defined as the sum of the potential energy operator and the kinetic energy operator:

$$\hat{H} = \hat{U} + \hat{T} \tag{2.3}$$

with $\hat{U} = U(x, y, z)$ and $\hat{T} = -\frac{\hbar^2\nabla}{2m^*}$. So that 3D-Schrödinger equation (2.2) can be rewritten as:

$$\left[U(x, y, z) - \frac{\hbar^2\nabla}{2m^*}\right]\varphi(x, y, z) = E\varphi(x, y, z) \tag{2.4}$$

By solving 2.4, it is possible to find the allowed energetic states E_i (eigenvalues of the system) and the corresponding probability distribution to find an electron φ_i (eigenstates of the system). Therefore by feeding the Schrödinger equation with

the potential energy profile of the system, it provides in output the eigenvalues and eigenstates of the system $\{E_i, \varphi_i\}$, thus it completely characterizes the quantum system.

Schrödinger equation can be analytically solved only for certain idealized particular potential profile distributions. For more complex cases a numerical approach becomes mandatory. In the following the idealized case of a 1D-finite square well is considered, in order to practically show how to solve it analytically. Then these results will be exploited in the next section to derive the density of states of 2D/1D/0D -systems.

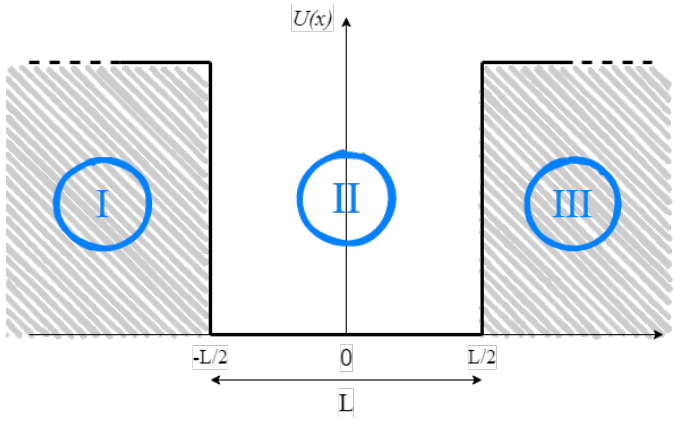


Figure 2.6: Potential energy distribution of 1-dimensional finite square well of length L. The finite heigth of the well is U_0.

- **1D Finite Square well.** The considered potential energy distribution $U(x)$ shown in figure 2.6 is an even function of x that can be described piece-wisely as follows:

$$U(x) = \begin{cases} U_0, & x < -L/2 \\ 0, & -L/2 \leq x \leq L/2 \\ U_0, & x > L/2 \end{cases}$$

In this way is it possible to solve Schrödinger equation piecewisely, i.e. for each region in an independent way. And so:

I. Supposing $E < U_0$ (since the focus is on looking for bounded states):

$$\left[U_0 - \frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial x^2}\right]\varphi_I(x) = E\varphi_I(x) \tag{2.5}$$

by rewriting it:

$$\frac{\partial^2}{\partial x^2}\varphi_I(x) = -(E - U_0)\varphi_I(x)\frac{2m^*}{\hbar^2} \tag{2.6}$$

a II order homogeneous Ordinary Differential Equation (ODE) with constant coefficients is obtained. This type of ODE has exponential solutions like:

$$\varphi_I(x) = Ae^{\lambda_1 x} + Be^{\lambda_2 x} \tag{2.7}$$

with $\lambda^2 = \frac{2m^*}{\hbar^2}(U_0 - E) \rightarrow \lambda_{1,2} = \sqrt{\frac{2m^*}{\hbar^2}(U_0 - E)} = \pm\alpha$ with $\alpha \in \mathbb{R}$ if $E < U_0$. Thus equation 2.7 can be rewritten as:

$$\varphi_I(x) = Ae^{\alpha x} + Be^{-\alpha x} \tag{2.8}$$

II. In the central region $U(x) = 0$, thus electrons are free to move, and the equation 2.4 becomes:

$$\frac{\partial^2}{\partial x^2}\varphi_{II}(x) = -E \cdot \frac{2m^*}{\hbar^2}\varphi_{II}(x) \tag{2.9}$$

again is obteined a II order homogeneous (ODE) with constant coefficients whose solutions are of the type:

$$\varphi_{II}(x) = Ce^{\lambda_1 x} + De^{\lambda_2 x} = Ce^{jkx} + De^{-jkx} \tag{2.10}$$

since $\lambda^2 = -\frac{2m^*}{\hbar^2}E \rightarrow \lambda_{1,2} = \pm j\sqrt{\frac{2m^*}{\hbar^2}E} = \pm jk$ with k wavenumber. Notice that the wavefunctions obtained have the typical expression of plane waves. This confirms the fact that in this region electrons are free to move. However it has to be remembered that the electrons are confined within the well, even if into it are free.

III. Since the potential profile is symmetric, the wavefunction of this region can be expressed as in the first:

$$\varphi_{III}(x) = Ee^{\alpha x} + Fe^{-\alpha x} \tag{2.11}$$

with $\alpha_{1,2} = \pm\sqrt{\frac{2m^*}{\hbar^2}(U_0 - E)}$.

Therefore in summary:

$$\varphi(x) = \begin{cases} Ae^{\alpha x} + Be^{-\alpha x}, & x < -L/2 \\ Ce^{jkx} + De^{-jkx}, & -L/2 \leq x \leq L/2 \\ Ee^{\alpha x} + Fe^{-\alpha x}, & x > L/2 \end{cases}$$

and in order to ensure limited and non divergent wavefunction, that is not physically possible (it must be $\in \mathbb{L}^2$) the following conditions must be imposed:

- 1. $B = 0$, otherwise the term $Be^{-\alpha x} \to +\infty$ for $x \to -\infty$;
- 2. $E = 0$, otherwise the term $Ee^{\alpha x} \to +\infty$ for $x \to +\infty$.

And therefore the expression of the wavefunction becomes:

$$\varphi(x) = \begin{cases} k_1 e^{\alpha x}, & x < -L/2 \\ k_2 cos(kx) + k_3 sin(kx), & -L/2 \leq x \leq L/2 \\ k_4 e^{-\alpha x}, & x > L/2 \end{cases}$$

since the following labels has been adopted:

- $A = k_1$,
- $F = k_4$,
- $C + D = k_2$,
- $(jC - jD) = k_3$.

since: $\varphi_{II} = C(cos(kx)+jsin(kx))+D(cos(kx-jsin(kx)) = (C+D)cos(kx)+(jC - jD)sin(kx) = k_2 cos(kx) + k_3 sin(kx)$.

Now to solve the problem, the values of unknowns k_i must be found. They can be derived with two different approaches:

- **Approach 1.** By enforcing the continuity of the wavefunction and its first derivative at the boundaries in order to get a proper solution ($\varphi \in \mathbb{C}^1$):

  (1) $\varphi_I(-L/2) = \varphi_{II}(-L/2)$ (3) $\varphi'_I(-L/2) = \varphi'_{II}(-L/2)$
  (2) $\varphi_{II}(L/2) = \varphi_{III}(L/2)$ (4) $\varphi'_{II}(L/2) = \varphi'_{III}(L/2)$

  From (1) $\to k_1 e^{-\alpha L/2} = k_2 cos(-kL/2) + k_3 sin(-kL/2)$
  $= k_2 cos(kL/2) - k_3 sin(kL/2)$
  From (2) $\to k_2 cos(kL/2) + k_3 sin(kL/2) = k_4 e^{\alpha L/2}$
  From (3) $\to \alpha k_1 e^{-\alpha L/2} = k_2 k sin(kL/2) + k_3 k cos(kL/2)$
  From (4) $\to -k_2 k sin(kL/2) + k_3 k cos(kL/2) = \alpha k_4 e^{\alpha L/2}$

  And therefore the following homogeneous system with 4 equations and 4 unknowns (k_1, k_2, k_3, k_4) is obtained:

$$A = \begin{bmatrix} e^{-\alpha L/2} & cos(-kL/2) & -sin(kL/2) & 0 \\ 0 & cos(kL/2) & sin(kL/2) & -e^{\alpha L/2} \\ \alpha e^{-\alpha L/2} & -k sin(kL/2) & -k cos(kL/2) & 0 \\ 0 & k sin(kL/2) & k cos(kL/2) & -\alpha e^{\alpha L/2} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = 0$$

To find the non trivial solutions of Schrödinger equation, and thus the non trivial eigenvalues of this system (i.e. the allowed energy values different from 0 eV), the condition $det(A) = 0$ must be enforced.
Anyway it is a very long and annoying procedure. Therefore the second approach explained in the following, will be adopted.

– **Approach 2.** It is a simpler and straightforward approach. Starting again from 2.1.2, it is possible to exploit the fact that $U(x)$ is an even function thanks to which it can be shown that the wavefunctions $\varphi(x)$, solutions of Schrödinger equation, can be either even or odd functions. Therefore, the two cases can be considered separately:

  * For the even wavefunctions $\varphi_{even}(x)$: $k_1 = k_4$ and $k_3 = 0$, so 2.1.2:

$$\varphi_{even}(x) = \begin{cases} k_1 e^{\alpha x}, & x < -L/2 \\ k_2 cos(kx), & -L/2 \leq x \leq L/2 \\ k_1 e^{-\alpha x}, & x > L/2 \end{cases}$$

  and by enforcing the continuity conditions on $\varphi(x)$:

  (1) $\varphi_I(-L/2) = \varphi_{II}(-L/2) \longrightarrow k_1 e^{-\alpha L/2} = k_2 cos(kL/2)$

  (2) $\varphi'_I(-L/2) = \varphi'_{II}(-L/2) \longrightarrow \alpha k_1 e^{-\alpha L/2} = k_2 k sin(kL/2)$

  then by enforcing $det(A) = 0$:

$$\begin{vmatrix} e^{-\alpha L/2} & -cos(kL/2) \\ \alpha e^{-\alpha L/2} & -ksin(kL/2) \end{vmatrix} = 0$$

  it yields: $\alpha = ktan(kL/2)$. Now it is possible to find graphically the non trivial eigenvalues $\{E_i\}$ solutions of Schrödinger equation. Indeed, by multiplying both sides by $L/2$, labeling $x = kL/2$, $y = \alpha L/2$, and by defining a circle of radius $R^2 = x^2 + y^2$, then the following non linear system of two equations is obtained:

$$\begin{cases} y = xtan(x) \\ x^2 + y^2 = R^2 \end{cases}$$

with:

$$\begin{aligned} x^2 + y^2 &= (kL/2)^2 + (\alpha L/2)^2 \\ &= (k^2 + \alpha^2)(L/2)^2 \\ &= \big[\frac{2m^*}{\hbar^2}E + \frac{2m^*}{\hbar^2}(U_0 - E\big](L/2)^2 \\ &= \frac{2m^*}{\hbar^2}U_0(L/2)^2 \\ &= R^2 \end{aligned} \tag{2.12}$$

By solving this non linear system graphically (figure 2.7) w.r.t. for e.g. x, the intersection (x_i) with $i = 1,2,...,n$ are obtained, from which the even eigenvalues E_i are easily derived since it yields:

$$\begin{aligned} E_i &= \big(\frac{2x_i}{L}\big)^2 \frac{\hbar^2}{2m^*} \\ &= \frac{2x_i}{m^* L}\hbar^2 \end{aligned} \tag{2.13}$$

∗ for the odd wavefunctions $\varphi_{odd}(x)$, an analogous procedure is followed. Thus, $k_1 = -k_4$ and $k_2 = 0$, so 2.1.2:

$$\varphi_{odd}(x) = \begin{cases} k_1 e^{\alpha x}, & x < -L/2 \\ k_3 sin(kx), & -L/2 \leq x \leq L/2 \\ -k_1 e^{-\alpha x}, & x > L/2 \end{cases}$$

and by enforcing the same continuity conditions on $\varphi(x)$ as previosly:

(1) $\varphi_I(-L/2) = \varphi_{II}(-L/2) \longrightarrow k_1 e^{-\alpha L/2} = -k_3 sin(kL/2)$

(2) $\varphi'_I(-L/2) = \varphi'_{II}(-L/2) \longrightarrow \alpha k_1 e^{-\alpha L/2} = k_3 k cos(kL/2)$

then by enforcing $det(A) = 0$:

$$\begin{vmatrix} e^{-\alpha L/2} & sin(kL/2) \\ \alpha e^{-\alpha L/2} & -k cos(kL/2) \end{vmatrix} = 0$$

it yields: $\alpha = -k cotan(kL/2)$. Now it is possible to find graphically the non trivial odd eigenvalues $\{E_i\}$ solutions of Schrödinger equation. Indeed, by multiplying both sides by $L/2$, labeling $x = kL/2$, $y = \alpha L/2$, and by defining a circle of radius $R^2 = x^2 + y^2$, then

in this case the following non linear system of two equations is obtained:

$$\begin{cases} y = -xcotan(x) \\ x^2 + y^2 = R^2 \end{cases}$$

By solving this non linear system graphically (figure 2.7) w.r.t. for e.g. x, the intersection (x_i) with $i = 1,2,...,n$ are obtained, from which in this case the odd eigenvalues E_i are easily derived in the same way as reported previously.

Therefore the solutions of the two systems (even and odd) are given by the intersection between the circumference $x^2 + y^2 = R^2$ and the curves $y = xtan(x)$ and $y = -xcotan(x)$. Moreover notice that since k and α are both $\in \mathbb{R}$ and positive (because $E < U_0$), also x and y $\in \mathbb{R}$ are positive, therefore only the intersections on the first quadrant of (x,y) plane are solutions.

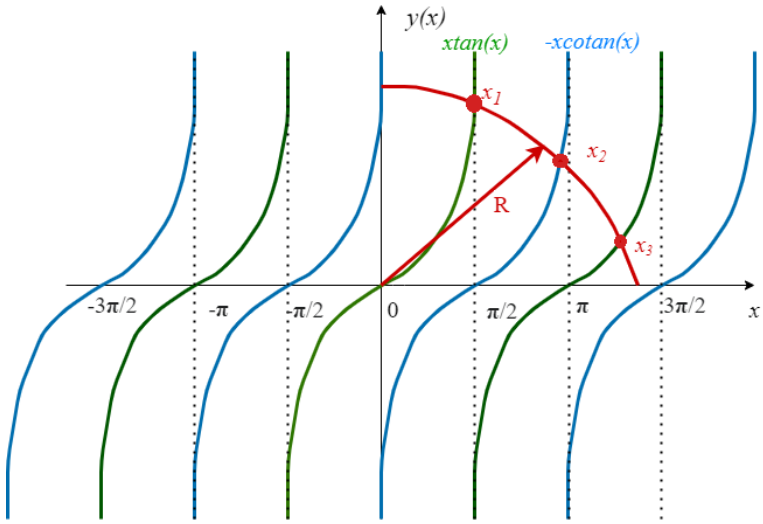


Figure 2.7: Graphical representation of the non trivial solutions of the considered 1D finite square well. The red curve is the circle of radius $R = \sqrt{\frac{2m^*}{\hbar^2}U_0}(L/2)$. Notice that in the figure there are only three allowed non trivial solutions (x_1, x_2, x_3)

Finally, there are some considerations important to be remarked.

If the length of the well L is increased, the radius R increases proportionally. This practically means more intersection points and thus more energy levels E_i. The same happens increasing the height of the well U_0: R increases but with a slower rate due to the square root. In the limit of infinite square well ($U_0 \to \infty$) the radius $R \to \infty$ and there are infinite solutions, thus an infinite number of energy levels placed at $\pi/2, \pi, 3\pi/2...$, that leads in the limit to a continous band of energy levels (3D-system case).

- **1D Infinite Square well.** The considered potential energy distribution $U(x)$ shown in figure 2.8 can be described piece-wisely as follows:

$$U(x) = \begin{cases} +\infty, & x < -L/2 \\ 0, & -L/2 \leq x \leq L/2 \\ +\infty, & x > L/2 \end{cases}$$

Also in this case the stationary Schrödinger equation $\hat{H}\varphi(x) = E\varphi(x)$ can be

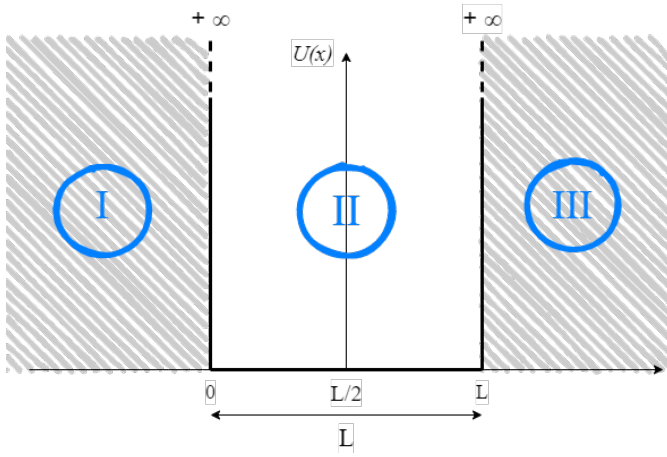


Figure 2.8: Potential energy distribution of 1-dimensional infinite square well of length L.

solved separately for each region. However, the II region is the only region in which the probability to find electrons is not null. Therefore the Schrödinger equation is solved only in this region, where it can be rewritten as:

$$-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial x^2}\varphi_{II}(x) = E\varphi_{II}(x) \tag{2.14}$$

by reordering the terms the following II order homogeneous ODE with constant coefficient is obtained:

$$\varphi''_{II}(x) + \alpha^2 \varphi_{II}(x) = 0 \tag{2.15}$$

with $\alpha^2 = -\frac{2m^*}{\hbar^2}E \rightarrow \alpha_{1,2} = \pm j\sqrt{\frac{2m^*}{\hbar^2}}E = \pm jk$. From basics of calculus, the general solution of this type of ODE is:

$$\begin{aligned} \varphi_{II}(x) &= Ae^{\alpha_1 x} + Be^{\alpha_2 x} \\ &= Ae^{jkx} + Be^{-jkx} \end{aligned} \tag{2.16}$$

Now, since electrons are confined in the quantum well, at its boundaries the probability to find them is null, i.e. $\varphi(0) = 0$, and so the previous equation 2.16 becomes $A + B = 0 \rightarrow A = -B$. Therefore it can be rewritten as:

$$\begin{aligned} \varphi_{II}(x) &= Ae^{jkx} - Ae^{-jkx} \\ &= A\big(e^{jkx} - e^{-jkx}\big) \\ &= A\big(cos(kx) + jsin(kx) - cos(kx) + jsin(kx)\big) \\ &= A2jsin(kx) \end{aligned} \tag{2.17}$$

with A, coefficient that can be derived by the normalization condition imposed on the wavefunction: $\int_{-\infty}^{+\infty} |\varphi(x)|^2 \, dx = 1$. Finally, by imposing the condition of null probability density also at the right boundary $\varphi(L) = 0$, from the previous equation is possible to derive k an thus the energy values:

$$\begin{aligned} \varphi_{II}(L) = A2jsin(kL) = 0 &\longrightarrow sin(kL) = 0 \\ &\longrightarrow k = \frac{n\pi}{L}, \qquad \text{with n} \in \mathbb{N}/0 \end{aligned} \tag{2.18}$$

Thus:

$$k_n = \frac{n\pi}{L} \tag{2.19}$$

$$E_n = \frac{\hbar^2 k^2}{2m^*} \tag{2.20}$$

In summary, it can be useful to comment the following figure (2.9) which compare the two idealized cases previously analyzed. Concerning the ground state E_0 in both cases the probability density to find electrons is focused in the middle of the well. But for a finite square well this probability is not negligible also outside the well, differently from the infinite square well case where this probability is null. Thus, the exponential decay of the wavefunction typical of the finite well,

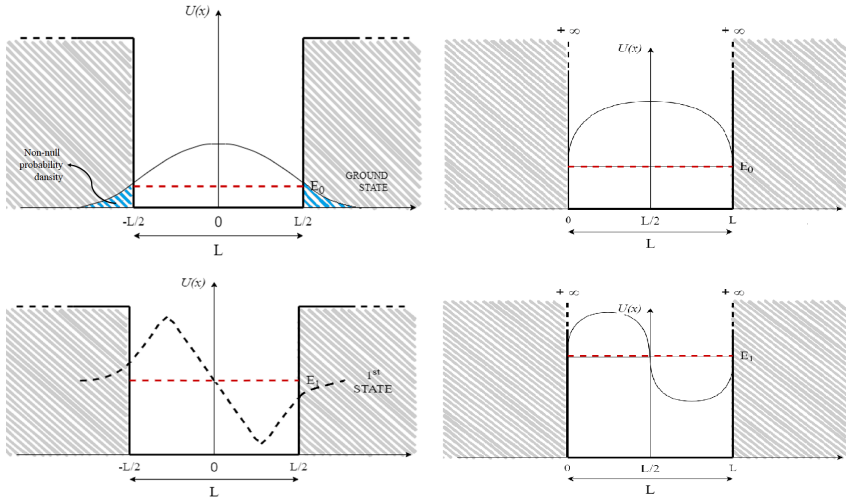


Figure 2.9: Differences in the first two energy levels between a 1D ideal finite square well and an infinite one.

disappears in the infinite one. The same considerations can be remarked for the first energy level E_1. Moreover, the energy states of the finite square well have always a lower energy w.r.t. the corresponding one of the infinite case. That is because, the taller the barrier, the greater are the eigenvalues E_i.

Notice that the infinite square well will be exploited in the next sections as bounding box approximation in order to derive the density of states of nanoscaled systems.

2.2 Modeling conduction at nanoscale

In order to model conduction in any structure, it is necessary firstly to know its Density Of States (DOS) $N(E)$, i.e. the density of the allowed energy states of the system which, if are occupied by carriers, may contribute to the conduction. They are basically determined by the geometry of the system as demonstrated in the following subsection for 3D/2D/1D/0D systems. Once they are known, the density of carriers available for conduction $\rho(E)$ can be obtained as:

$$N(E) \cdot f_{FD}(E) = \rho(E) \tag{2.21}$$

with $f_{FD}(E)$, Fermi-Dirac function which tells the probability distribution of electrons as function of energy. After the determination of $\rho(E)$ then the current

distribution can be easily determined, as done in the following final subsection of this section only for 0D-systems.

The âparticle in a boxâ takes a complex structure like a molecule and approximates it by a homogeneous box. All details, such as atoms, are ignored.

2.2.1 Density of States (DOS)

In order to solve Schrödinger equation and compute DOS, two assumptions are made:

a. in the direction of confinement the "particle in a box" (also called "Bounding Box approximation" (BB)) is exploited. In this approximation the conductor, even if it complex like a molecule is considered as an homogeneous box and so all the details, such as atoms are ignored. Obviously is an approximation, but typically for these types of systems is anyway a good approximation and useful to understand what happens physically.

   So the potential profile in the directions of confinement is approximated with an ideal infinite height square well, in order to get easily some important properties of the system (like for e.g. allowed energy levels).
   Notice that in 3D-system case the potential box is the entire crystal lattice between the S/D electrodes. Indeed, electrons have to overcome a very large surface potential in order to escape from the material (figure 2.1) and so at the end of the story they can be considered confined in the whole material, but inside it they behave like 3D-plane waves since they are free to move in any directions.

b. at the borders of the conductor, along the directions where no confinement occurs, Periodic Boundary Conditions (PBCs) are applied. They consist on impose the following condition: $\varphi(0) = \varphi(n\lambda_{Fermi}) \neq 0$ with n integer period of the wavefunction, and with the length of the quantum well $L_z = n\lambda_{Fermi}$ in such a way that at the interfaces the electron wavefunctions assume all the same values and entire periods of them are enclosed in the bounding box. PBCs are good boundary conditions that make Schrödinger equation simple to solve. Moreover choosing more accurate and complex BC like for example Dirichlet or Neumann will not influence the result from a mathematical point of view, but it may accelerate the computational time required to converge to the solution. Anyway since the aim of this section is simply analytically (with many approximation) derive what physically happens, without correlation with a numerical analyisis, PBCs are the best choice.

3D-systems

In 3D-systems or more commonly bulk systems (figure 2.2) there is *no quantum confinement* in any direction. The carriers are free to move along the conductive channel and a semiclassical approach based on drift-diffusion is accurate enough to get physical insights of the system. In this case the allowed energetic states of the systems are infinite in number, thus they form a different continuous bands, the topmost of which typically contribute to the conduction: the Valence Band (VB) and Conduction Band (CB). It can be shown that the density of the allowed states (i.e. DOS) in CB and VB are:

$$N_{CB}(E) = \gamma(E - E_c)^{1/2} \quad \text{in CB} \tag{2.22}$$

$$N_{VB}(E) = \gamma(E_v - E)^{1/2} \quad \text{in VB} \tag{2.23}$$

with $\gamma = 4\frac{\pi}{h^3}(2m_n^*)^{3/2}$. From this quantity now it is possible, by means of Fermi function distribution, to derive how many states can be occupied by electrons. These electrons that will fill the allowed states at the minimum energies, will contribute to conduction *independently* on their own energy! In other words, all electrons good for conduction are identical, they are not distinguished for their energy: an electron occupying the energy level E_1 is equal to the one occupying the energy level E_2 (figure 2.10). Indeed in the bulk drift-current density formula due to electrons:

$$\vec{J_n} = q\mu_n n \vec{E_{drift}} \tag{2.24}$$

the electron carrier density n is a parameter which takes into account the overall population and there are no differences between them in terms of energy. Carriers are all equal for the current. The only important thing is that they must occupy a state in the CB in order to contribute to that current. This is the key point which differ bulk-systems to the nanoscalede ones, where instead each electron contribute to the conduction in a different manner dependently from the velocity (and thus energy) they have. Indeed, in this case drift-current density due to electrons, becomes:

$$\vec{J_n} = q\mu_n n(E_1)\vec{E_{drift}} + q\mu_n n(E_2)\vec{E_{drift}} + q\mu_n n(E_3)\vec{E_{drift}} + ... \tag{2.25}$$

2D-systems

In 2D-systems (figure 2.3) quantum confinement of carriers is present in one direction, typically along a transversal direction, orthogonal w.r.t. the direction of transport. The confinement is modeled with a bounding box inside which carriers are confined due to two high potential barriers. In the other directions they are free to move and their wavefunctions are modeled as *plane waves*.

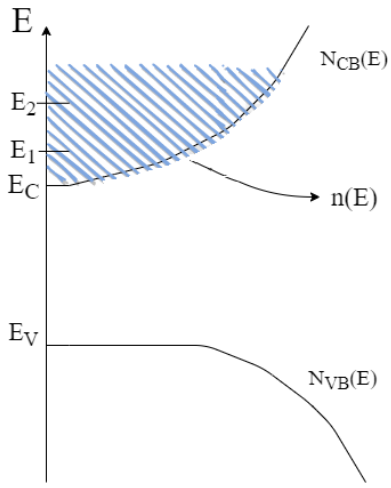


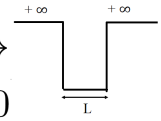
Figure 2.10: 3D-DOS of the Conduction Band (CB)

Therefore the potential energy distribution to which carriers are subjected, is a three dimensional function, but exploiting the *separation of variables* is it possible to split its dependence on the three axis in three independent terms:

$$U(x, y, z) = U_x(x) + U_y(y) + U_z(z) \tag{2.26}$$

If x is the direction of confinement (figure 2.3), then the potential energy along it is approximated with a bounding box profile (infinte square well), whereas along y and z the two components of potential energy are null, since there are no potential barriers for carriers because they are free to move. So:

$$U(x, y, z) = \begin{cases} U_x(x) \longrightarrow \\ U_y(y) = 0 \\ U_z(z) = 0 \end{cases}$$



Now, thanks to linearity of Schrödinger equation and full separability of the Laplace operator ∇^2, it is possible to apply separation of variables also to the wavefunctions (eigenstates) and energy levels (eigenvalues).

Therefore a 3D stationary Schrödinger equation to be solved turns into three 1D uncoupled Schrödinger equations that can be solved independently.

$$\hat{H}\varphi(x,y,z) = E\varphi(x,y,z) \Longrightarrow \begin{cases} \hat{H}_x\varphi_x(x) = E_x\varphi_x(x) \\ \hat{H}_y\varphi_y(y) = E_y\varphi_y(y) \\ \hat{H}_z\varphi_z(z) = E_z\varphi_z(z) \end{cases}$$

Moreover notice that if: the first equation is multiplied by $\varphi_y\varphi_z$, the second by $\varphi_z\varphi_x$, the third by $\varphi_x\varphi_y$, and all the LHS terms and all the RHS terms are summed, again the total Schrödinger equation is obtained:

$$\begin{cases} \hat{H}_x\varphi_x\varphi_y\varphi_z = E_x\varphi_x\varphi_y\varphi_z \\ \hat{H}_y\varphi_x\varphi_y\varphi_z = E_y\varphi_x\varphi_y\varphi_z \\ \hat{H}_z\varphi_x\varphi_y\varphi_z = E_z\varphi_x\varphi_y\varphi_z \end{cases}$$

$$\Longrightarrow \hat{H}_{tot}\varphi_x\varphi_y\varphi_z = (E_x + E_y + E_z)\varphi_x\varphi_y\varphi_z \tag{2.27}$$

with $E(x,y,z) = E_x + E_y + E_z$ and with the 3D wavefunction $\varphi(x,y,z) = \varphi_x\varphi_y\varphi_z$.

Now, for 2D-systems, electrons have two degrees of freedom. If we consider the figure 2.3, they are free to move along y and z. Therefore their wavefunctions can be modeled as plane waves along y and z: $\varphi_y(y) = Ae^{jk_y y}$, $\varphi_z(z) = Be^{jk_z z}$. On the other hand, in the direction of confinement an infinite heigth well is present, so recalling the previous result, the wavefunction along x results to be: $\varphi_x(x) = Csin(k_x x)$, with allowed wavenumbers determined by the geometry of the system, equal to $k_x = \frac{n_x\pi}{L_x}$. Therefore the 3D wavefunction and the total eigenvalues of a 2D-system can be written as:

$$\varphi(x,y,z) = Asin(k_x x)e^{jk_y y}e^{jk_z z} \tag{2.28}$$

$$E(x,y,z) = E_x + E_y + E_z = \frac{\hbar^2}{2m^*}\frac{n_x\pi}{L_x} + \frac{\hbar^2}{2m^*}(k_y^2 + k_z^2) \tag{2.29}$$

where the y and z components are kinetic energies whereas the x-component are the discretized allowed energy levels due to quantum confinement. The only unknowns remain k_y, k_z. Their allowed values are obtained by imposing proper boundary conditions. Periodic boundary conditions (PBCs) are forced at the boundaries of the domain:

$$\varphi_y(y=0) = \varphi_y(y=L_y), \qquad \varphi_z(z=0) = \varphi_z(z=L_z) \tag{2.30}$$

Notice that, considering the reference system of figure 2.2: L_x represent the height of the channel, L_y the width and L_z the length. These conditions mean to fix equal the phases of the wavefunctions at $x = 0$ and $x = L_y$ and similarly for z direction.

And it can be shown that they are satisfied if L_y, L_z are multiple integer of the Fermi wavelength λ_F in the corresponding directions, i.e. if:

$$L_y = n_y \lambda_F(y), \qquad L_z = n_z \lambda_F(z) \tag{2.31}$$

which translate into the allowed values of the wavevectors:

$$k_y = \frac{2\pi n_y}{L_y}, \qquad k_z = \frac{2\pi n_z}{L_z} \tag{2.32}$$

they are substituted in the previous expressions and so the total eigenvalues (E) and the corresponding eigenstates (i.e. φ, that are the propagation modes associated to each energetic state) of the systems are derived. These energetic states and the corresponding wavefunctions with these specified values of wavenumbers are the non-trivial solutions of the Schrödinger equation of the 2D system characterized by the above specified potential profile $U(x, y, z)$.

Now, to determine the 2D-DOS it is necessary to derive the number of states for each allowed bounded energy level E_{x_n}. In order to do it the (k_y, k_z) plane is plotted (figure 2.11) and the allowed $k_y k_z$ are counted. Supposing $L = L_x = L_y = L_z$ then it yields: $k_y = k_z = \frac{2\pi n}{L}$. And so in (k_y, k_z) plane the number of k-states per unit area are $\frac{2}{(2\pi/L)^2}$ since in the unit area $A = base \cdot height = \frac{2\pi}{L} \cdot \frac{2\pi}{L}$ there are 2 allowed states.

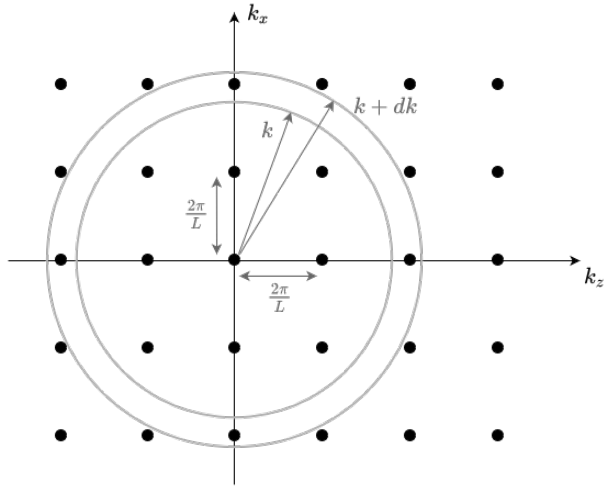


Figure 2.11: The (k_y, k_z) plane

In order to compute the DOS per unit k in this space an infinitesimal width dk is given to the circle $k_x^2 + k_y^2 = k^2$. Therefore now in the ring of radius k and width

$k + dk$ is it possible to count the number of points belonging to the ring, i.e. the number of k-states per unit k per unit area, that results to be:

$$\begin{aligned} N_{2D}(k)dk &= \frac{2}{(2\pi/L)^2} \cdot 2\pi k dk \frac{1}{(L^2} \\ &= \frac{kdk}{\pi} \end{aligned} \tag{2.33}$$

with $2\pi kdk$ the infinitesimal area of the ring and with $1/L^2$ to make $N_{2D}(k)$ independent from the geometry of the device, in such a way that when the number of available states must be computed it is enough to multiply $N(E)$ by $L_y L_z$.

However, the aim is to find the number of states per unit energy E, i.e. $N(E)dE$ with $dE = \frac{\hbar^2}{2m} 2kdk$, thus:

$$N_{2D}(E)dE = \frac{m}{\pi\hbar^2} dE \longrightarrow N_{2D}(E) = \frac{m}{\pi\hbar^2} \tag{2.34}$$

These are the number of allowed states per unit of energy and unit of area for a single E_{n_x} plotted for $x_n = 1$ on the left side of figure 2.12. They are independent from the energy E.

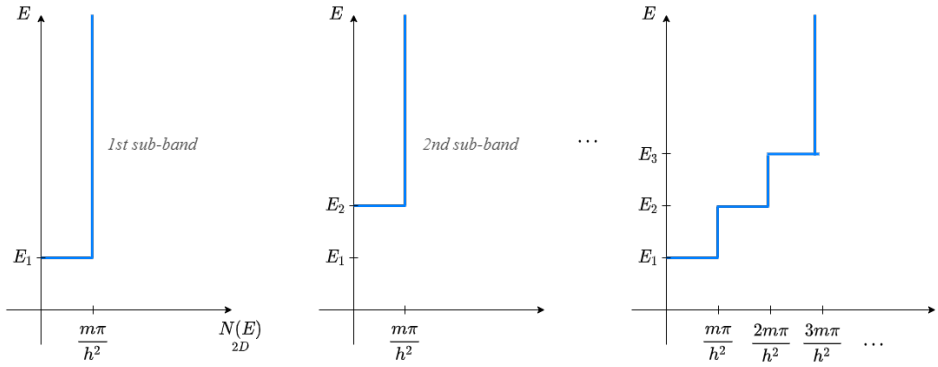


Figure 2.12: DOS of 2D system. On the left side the ones for single sub-bands. On the right side 2D-DOS for the entire system, resulted from the sum of DOS of all the subb-bands.

Summarizing: once the geometry of the 2D-system is given, the allowed discrete energy levels due to confinement in the x direction are obtained solving the Schrödinger equation: $E_{n_x} = \frac{n_x^2 \pi^2 \hbar^2}{2mL_x^2}$. However electrons are free along y and z directions so they can assume any kinetic energy values E_y, E_z ($\in \mathbb{R}$). This means that once value of n_x is fixed, a corresponding value of energy E_{n_x} is obtained and

for this specific energy value there are $\frac{m}{\pi\hbar^2}$ available states. So in conclusion in order to find the complete 2D-DOS, the available states of all the sub-bands E_{n_x} must be summed. As a result:

$$DOS_{2D} = \frac{m}{\pi\hbar^2} \sum_{n_z} \mu(E - E_{x_n}) \tag{2.35}$$

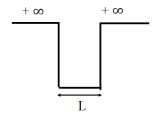
with $\mu(x)$ the step function centered in E_{n_x}

1D-systems

In 1D-systems (figure 2.4) quantum confinement of carriers is present in two directions, typically along the transversal ones, orthogonal w.r.t. the direction of transport. The confinement is modeled with a bounding box inside which carriers are confined due to two high potential barriers. In the other direction they are free to move and their wavefunctions are modeled as *plane waves*.

By following exactly the same method and considerations used for 2D-systems, the potential energy distribution is split in three independent terms and it can be considered piecewisely: with x and y directions of confinement, where the potential

$$U(x,y,z) = \begin{cases} U_x(x) \longrightarrow \\ U_y(y) \longrightarrow \\ U_z(z) = 0 \end{cases}$$



energy is approximated with a bounding box profile (infinte square well), whereas along z the potential energy is null, since there are no potential barriers for carriers because they are free to move.

Separation of variables holds also for the wavefunction, and so the total 3D wavefunction of the system can be written as:

$$\varphi(x,y,z) = Asin(k_x x)sin(k_y y)e^{jk_z z} \tag{2.36}$$

with allowed wavenumbers determined by the geometry of the system, equal to $k_x = \frac{n_x\pi}{L_x}$, $k_y = \frac{n_y\pi}{L_y}$ and $k_z = \frac{n_z 2\pi}{L_z}$ determined by imposing the PBC. Therefore the total eigenvalues of a 1D-system results to be:

$$E(x,y,z) = E_x + E_y + E_z = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2) + \frac{\hbar^2}{2m^*}kz^2 \tag{2.37}$$

where only the z-component is a kinetic energies whereas the x any y -components are the discretized allowed energy levels due to quantum confinement.

Now, to determine the 1D-DOS it is necessary firstly to derive the number of energy states for each couple of bounded states $E_{nx,ny}$. In order to do it the k_z straight line is plotted (figure 2.13) and the allowed kz are counted.

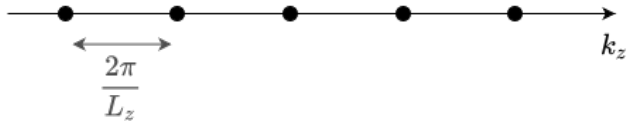


Figure 2.13: The k_z straight line.

So the the number of k-states per unit length are $\frac{2}{2\pi/L}$ and the number of k-states per unit length per unit k is in this case are:

$$\begin{aligned} N_{1D}(k)dk &= \frac{2}{2\pi/L} \cdot dk_z \cdot \frac{1}{L} \\ &= \frac{dk_z}{\pi} \end{aligned} \tag{2.38}$$

with $1/L$ to make $N_{1D}(k)$ independent from the geometry of the device, in such a way that when the number of available states must be computed it is enough to multiply $N(E)$ by L_z.

However, the aim is to find the number of states per unit energy E, i.e. $N(E)dE$ with $dE = \frac{\hbar^2}{2m} 2k_z dk_z$, thus:

$$N_{1D}(E)dE = \frac{m}{\pi\hbar^2 k_z} dE \tag{2.39}$$

that can be rewritten by expliciting k_z:

$$\begin{aligned} N_{1D}(E) &= \frac{m\sqrt{\hbar^2}}{\pi\hbar^2\sqrt{2mE_z}} \\ &= \frac{\sqrt{m}}{\pi\hbar} \frac{1}{\sqrt{2(E - E_{nx,ny})}} \end{aligned} \tag{2.40}$$

These are the number of allowed states per unit of energy and unit length for a single sub-band $E_{nx,ny}$ plotted on the left side of figure 2.14 for one single sub-band. They are not independent from the energy E, differently from the ones of 2D-systems.

Finally, in order to find the complete 1D-DOS, all the available states of the sub-bands $E_{nx,ny}$ allowed by the geometry of the system must be considered. As a result:

$$DOS_{1D} = \frac{\sqrt{m}}{\pi\hbar} \sum_{n_x,n_y} \frac{\mu(E - E_{n_x,n_y})}{\sqrt{2(E - E_{n_x,n_y})}} \tag{2.41}$$

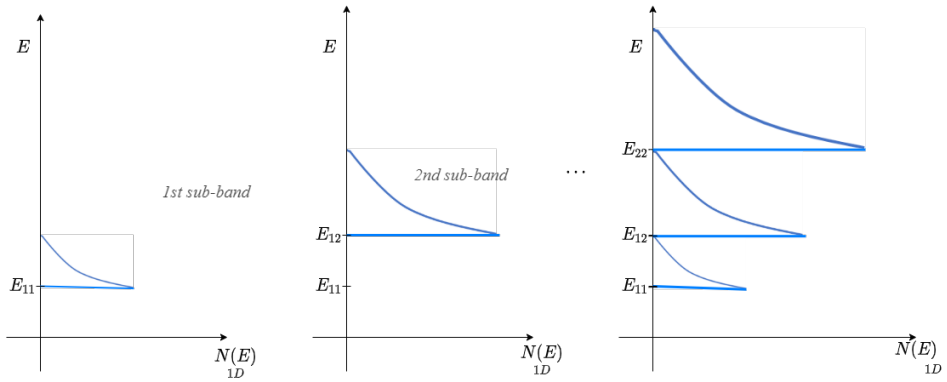


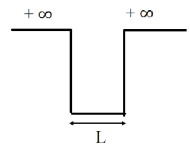
Figure 2.14: DOS of 1D system. On the left side the ones for single sub-bands. On the right side 1D-DOS for the entire system, resulted from the sum of DOS of all the subb-bands.

whose distribution is shown on the right side of the figure 2.14. Notice that the larger is the cross-section of the 1D-system $L_x L_y$, the closer will be $E_{nx,ny}$

0D-systems

In 0D-systems, electrons have no degree of freedom, they are confined in all the three directions x,y,z. It is like having a 3-dimensional bounding box, and so given the geometry of the quantum dot, the allowed energy values E_{n_x,n_y,n_z} which satisfy the Schrödinger equation are directly found. Indeed, in this case the potential profile distribution is the following: And the total wavefunction, solution of the

$$U(x,y,z) = \begin{cases} U_x(x) \longrightarrow \\ U_y(y) \longrightarrow \\ U_z(z) \longrightarrow \end{cases}$$



Schrödinger equation is in this case:

$$\varphi(x,y,z) = Asin(k_x x)sin(k_y y)sin(k_z z) \tag{2.42}$$

since there are no free electrons, no plane waves are present in the expression. The allowed wavenumbers determined by the geometry of the system are equal to $k_x = \frac{n_x \pi}{L_x}$, $k_y = \frac{n_y \pi}{L_y}$ and $k_z = \frac{n_z \pi}{L_z}$. Since there are no degree of freedom for

electrons, no PBC are imposed. Therefore the total eigenvalues of a 0D-system results to be (assuming $L = L_x = L_y = L_z$):

$$E(x, y, z) = E_x + E_y + E_z = \frac{\hbar^2}{2m^*}\frac{\pi^2}{L^2}(n_x^2 + n_y^2 + n_z^2) \tag{2.43}$$

Notice that there are not kinetic energy component and only discrete energy levels are present. As a consequence the 0D-DOS are simply lines. They are the number of states per box at each possible discrete energy level E_{n_x,n_y,n_z} . There-

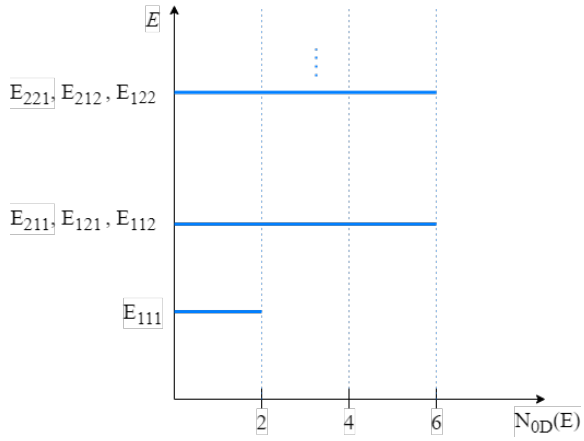


Figure 2.15: An example of 0D-DOS. Notice that is it possible to have more than 2 states per unit energy, since for three different combination of n_x, n_y, n_z the same energy values are obtained, simply because $L = L_x = L_y = L_z$ by hypothesis.

fore, for quantum dots, sub-bands are not obtained, they degenerate in discrete energy states. This situation holds if the quantum dot is ideal, i.e. completely isolated, in the sense that it does not change matter with the outside. Otherwise if an exchange occurs with for e.g. the electrodes at the interfaces, then the discrete energy states are broadened due to the interaction between electrodes and dot.

2.2.2 Conduction in 0D-systems

The derivation of the current as function of bias will be derived only for 0D-systems, since the focus of this thesis is on them. In particular the case of study for 0D-system will be referred to a single molecule, that can be esily modeled as quantum dot.

As already mentioned previously, in 0D-systems all three spatial directions are confined and therefore the solutions of Schrödinger equations support only bounded

states, since the 3D potential profile is approximated with the BB approximation in all three directions. As a consequence, different mechanisms of conduction occurs w.r.t. the other systems (3D, 2D, 1D) which support propagating states at least in one direction.
The 0D-DOS, as shown previously, degenerates in discrete energy levels.
Actually the BB approximation leads to a bad estimation of the energy levels of the dot. Indeed, with this approximation, the 0D-conductor, that for e.g. can be complex like a molecule, is considered as an homogeneous box and so all the details, such as atoms are ignored. So, generally is not so satisfactory as approximation. It is not enough in order to obtain the real energy levels of a quantum dot since a real quantum dot is not a bounding box with infinite height barriers and so energy levels can differ a lot. Therefore, in case like molecules, in order to evaluate the energy levels, ab-initio simulations are needed.

However, quantum dots like molecules, are systems where electrons are confined in any directions. So a reasonable questions arise: how it is possible to transfer charge and thus obtain a conduction in this 3D-bounding box where electrons cannot move? Moreover, the allowed energetic states are bounded states not travelling ones, therefore how it is possible to get a current in this type of systems?
Basically the key-point in answering those questions is the influence of the contact electrodes on the electronic behaviour of the system. Indeed, when the quantum dot (for e.g. a benzene molecule) is contacted with two metal electrodes acting like S/D, it is possible to experience conduction. Depending on how technologically the dot is coupled with the contacts there may be maily two types of conduction:

a) via resonance with specific energy state in weakly coupled systems, thanks to sequential tunneling process. Electron simply tunnels from S toward the dot if there is available state at that particular energy and after a certain arbitrary interval of time it will jump from the dot toward the D make possible transfer of charge, i.e. conduction through the system;

b) thanks to the broadening of energy levels in strongly-coupled systems. In this case the conduction is of ballistic/coherent type like in 1D-systems. The influence of the electrodes, that are huge reservoirs of travelling states, will deeply influence the discrete energy levels of the dot, making them broader until they become like a sort of sub-band.

The conduction model briefly explained in the following holds for strongly coupled quantum dots, in which coherent transport occurs.
Firstly a simple model which considers only one discrete energy level included in the bias window (BW), is considered. The BW is the energy range within which the energy states can contribute to conduction. It is defined by means of the bias applied to the system $BW = -qV_{DS}$.

In the following, this first case of study will treat only one level that makes possible the transfer of charge.

Conduction via a single discrete energy level.

If a single energy level of the quantum dot E_L, is between μ_S and μ_D the Fermi levels of source and drain respectively, i.e. if a single energy level of the dot is included in the BW (figure 2.16), there is a continuous transfer of electrons that gives rise to a flux from S to D and so to a coherent current I_{DS} through E_L. In the following will be derived.

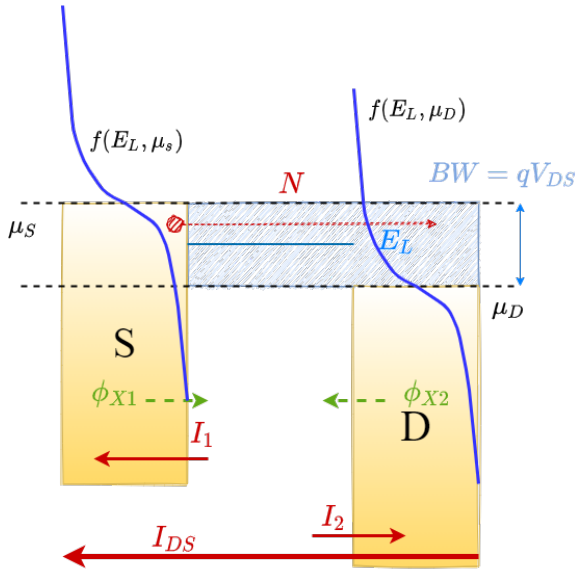


Figure 2.16: Case of a single energy level of the quantum dot E_L falling between μ_S and μ_D.

The quality of interfaces are characterized in terms of coupling by means of transit time already introduced previously. In this treatment:

- τ_1 = time required by 1 electron to move from S to energy level of the dot;
- τ_2 = time required by 1 electron to move from D to energy level of the dot;

Remember that if τ is high, there will be poor quality of interfaces, otherwise if τ is low, there will be good quality (good coupling factor γ). Indeed: $\tau_1 = \frac{\hbar}{\gamma_1}$, $\tau_2 = \frac{\hbar}{\gamma_2}$. Remember also that γ is a measure of how much the electrodes influence

the energy levels of the dot and so it's a measure of the broadening. (If τ decreases, γ increases, and so broadening increases).

There are four independent fluxes (i.e. number of particle/time unit) of electrons, that are defined in the following ways:

1. the flux from the source towards the dot is:

$$\Phi_{x_{1\to dot}} = \frac{2 \cdot f_{FD}(E_L, \mu_S)}{\tau} = \frac{2\gamma_1}{\hbar} \cdot f_{FD}(E_L, \mu_S) \tag{2.44}$$

   the term 2 stands for the DOS, since only at maximum 2 electron can be occupied in the single energy level E_L. The Fermi function tells the probability of occupation of E_L. Hence the term $2 \cdot f_{FD}$ actually tells the actual number of electrons can be transfered towards E_L;

2. the flux from the dot towards the source is:

$$\Phi_{x_{dot\to 1}} = \frac{N}{\tau_1} = \frac{\tau_1 N}{\hbar} \tag{2.45}$$

   with N the number of the electrons in the dot;

3. similarly to (1), the flux from the drain towards the dot is:

$$\Phi_{x_{2\to dot}} = \frac{2 \cdot f_{FD}(E_L, \mu_D)}{\tau_2} = \frac{2\gamma_2}{h} \cdot f_{FD}(E_L, \mu_D) \tag{2.46}$$

4. similarly to (2), the flux from the dot towards the drain is:

$$\Phi_{x_{dot\to 2}} = \frac{N}{\tau_2} = \frac{\tau_2 N}{h} \tag{2.47}$$

The resulting net fluxes from source to drain and vice versa are:

- net flux from source to drain:

$$\Phi_{x1} = \Phi_{x_{1\to dot}} - \Phi_{x_{dot\to 1}} = \frac{2\phi_1}{h}[2f(E_L, \mu_S) - N] \tag{2.48}$$

- net flux from drain to source:

$$\Phi_{x2} = \Phi_{x_{2\to dot}} - \Phi_{x_{dot\to 2}} = \frac{\phi_2}{h}[2f(E_L, \mu_D) - N] \tag{2.49}$$

For the principle of charge conservation, at steady state the flux from source to drain Φ_{x1} is equal in magnitude to the flux from drain towards the source Φ_{x2} and opposite in sign, hence With this condition N, i.e. the number of electrons hosted

in the dot at E_L, can be derived. Indeed, forcing that equality ($\Phi_{x1} = -\Phi_{x2}$), it yields:

$$\frac{\phi_1}{\hbar}\left[2f(E_L,\mu_S) - N\right] = -\frac{\phi_2}{\hbar}\left[2f(E_L,\mu_D) - N\right] \tag{2.50}$$

$$\longrightarrow \frac{2}{\hbar}\left[\gamma_1 f(E_L,\mu_S) + \gamma_2 f(E_L,\mu_D)\right] = \frac{\gamma_1+\gamma_2}{\hbar}N \tag{2.51}$$

and finally the number of electrons hosted in the dot at E_L results to be:

$$\longrightarrow N = \frac{2}{\gamma_1+\gamma_2}\left[\gamma_1 f_{FD}(E_L,\mu_S) + \gamma_2 f_{FD}(E_L,\mu_D)\right] \tag{2.52}$$

that means that, in the hypothesis: $\gamma_1 = \gamma_2 = \gamma$:

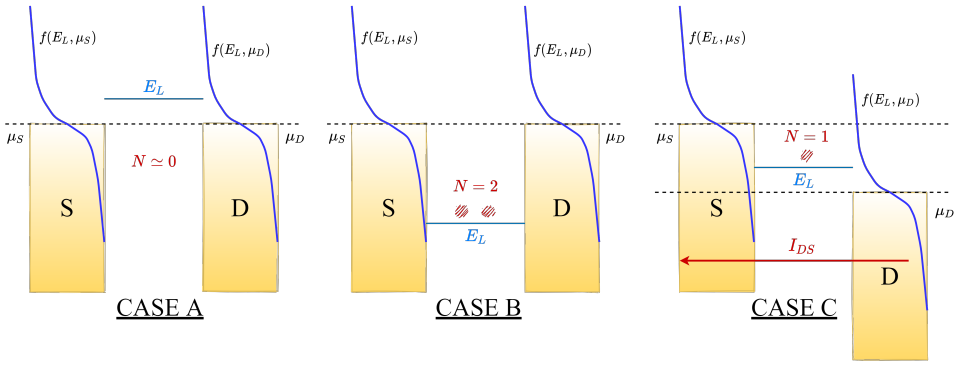


Figure 2.17: The previous cases described.

- if case (A), fig.2.17 : $E_L \gg \mu_S$ and $E_L \gg \mu_D \longrightarrow f(E,\mu_S), f(E,\mu_D) = 0 \longrightarrow N = 0$.
  The number of electrons hosted at E_L tends to zero because the Fermi function distribution at that energy is almost zero. So no states means no electrons, and therefore no conduction.

- if case (B), fig.2.17: $E_L \ll \mu_S$ and $E_L \ll \mu_D \longrightarrow f(E,\mu_S), f(E,\mu_D) \simeq 1 \longrightarrow N = 2$.
  The number of electrons hosted at E_L is 2 because the Fermi function distribution at that energy is equal to 1. However, there are no free states at E_L, and so there is no possibility of electron transfer, since no empty states at drain are present. Hence, $I_{DS} = 0$.

- if case (C), fig.2.17: $V_{DS} \neq 0$ and so $\mu_D < E_L < \mu_S \longrightarrow N \simeq 1$ since $f(E_L, \mu_S) \simeq 1$ and $f(E_L, \mu_D) \simeq 0$.
  The number of electrons hosted at E_L ca n be almost 1, because the Fermi function distribution at that energy is almost equal to 1. Moreover, now there are free states at E_L and also at drain, and so there is the possibility of electron transfer, and therefore conduction is possible. Hence, $I_{DS} \neq 0$.

Now it is possible to derive the expression of the current I_{DS} as the product of the charge of one electron $-q$ with the number of electrons per unit time i.e. the flux of electrons which travel from D to S ($-\Phi_{x1}$). Hence the current from S to D flowing through the dot via a single dicrete level is $I_{DS} = +q\Phi_{x1}$, which can be rewritten by expliciting the flux Φ_{x1}, as:

$$I_{DS} = +q\left[\frac{\gamma_1}{\hbar}(2f(E_L, \mu_S) - N)\right] \tag{2.53}$$

with $N = \frac{2}{\gamma_1+\gamma_2}[\gamma_1 f(E_L, E_F S) + \gamma_2 f(E_L, E_F D)]$ from equation 2.52. So, substituting it into the expression of the current, the following is obtained:

$$\begin{aligned} I_{DS} &= \frac{q}{\hbar}\left[\gamma_1\big(2f(E_L, \mu_S) - \frac{2}{\gamma_1+\gamma_2}\gamma_1 f(E_L, \mu_S) - \frac{2}{\gamma_1+\gamma_2}\gamma_2 f(E_L, \mu_D)\big)\right] \\ &= \frac{q\gamma_1}{\hbar}\left[\frac{2(\gamma_1+\gamma_2)f(E_L, \mu_S) - 2\gamma_1 f(E_L, \mu_S) - 2\gamma_2 f(E_L, \mu_D)}{\gamma_1+\gamma_2}\right] \\ &= \frac{q\gamma_1}{\hbar(\gamma_1+\gamma_2)}\left[2\gamma_2 f(E_L, \mu_S) - 2\gamma_2 f(E_L, \mu_D)\right] \\ &= \frac{2q}{\hbar}\frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}\left[f(E_L, \mu_S) - f(E_L, \mu_D)\right] \end{aligned} \tag{2.54}$$

Notice that this formulation holds for one single discrete energy level in the case of 0D-systems (Qdot) strongly coupled with the reservoirs electrodes. It is not a general expression.
However, to understand better this expression, the following example is considered. Supposing E_L of the dot above of 0.5eV w.r.t. E_F of the dot (i.e. LUMO), the following cases may occur:

- case (a): if $V_{DS} = 0 \longrightarrow \mu_S = EFD = E_F$ since we're at equilibrium. The Fermi functions are almost equal 0 at E_L, therefore the current though it is almost zero, since n transfer of electrons is possible: $f(E_L, \mu_S) = f(E_L, \mu_D) \simeq 0 \longrightarrow I_{DS} = 0$. Anyway, notice that, even if E_L was supposed to be $0.5ev$ below E_F, the same situation would be verified, since at E_L, it would be: $f_{FD} \simeq 1$, and so: $f(E_L, \mu_S) - f(E_L, \mu_D) = 0$ since $\mu_S = \mu_D$, hence the current would be zero anyway also in this case ($I_{DS} = 0$);

- case (b): if $V_{DS} = 0.5v$, then the Fermi level of the source μ_S rises of $+0.25ev$, and the one of the drain μ_D gets lower of $-0.25ev$ since the voltage division factor is supposed to be $\eta = 0.5$ thanks to a supposed symmetric coupling of the dot with S and D ($\gamma = \gamma_1 = \gamma_2$). Notice that actually the exact position of E_F of the dot. The hypothesis that it falls in the middle position between μ_S and μ_D so that:

  1. $\mu_S - E_{F_{dot}} \simeq \frac{qV_{DS}}{2}$
  2. $E_{F_{dot}} - \mu_D \simeq \frac{qV_{DS}}{2}$

  it's a reasonable assumption if the two contacts have both the same quality of the interfaces and so coupling factors.

  So in this case, even if $V_{DS} \neq 0V$, a very small value of current is obtained ($I_{DS} \simeq 0$), since V_{DS} too small leads to $f_{FD} \simeq 0$ at E_L. Instead, as shown in the following case (c), by increasing V_{DS}, I_{DS} will increase until saturation point is reached.

- case (c): if $V_{DS} = 1v$, then:

$$I_{DS} \simeq \frac{2q}{\hbar} \frac{\gamma_1\gamma_2}{\gamma_1 + \gamma_2} \cdot 1 = \frac{2q}{\hbar} \tag{2.55}$$

  that is because:$f_{FD}(E_L, \mu_D) = 0$ and $f_{FD}(E_L, \mu_S) = 1$.
  On the contrary, if $V_{DS} = -1$V, then at E_L: $f_{FD}(E_L, \mu_D) = 1$ and $f_{FD}(E_L, \mu_S) = 0$, hence the current $I_{DS} \simeq -\frac{2q}{\hbar}$.
  A saturation of I_{DS} is reached when at E_L: $f_{FD}(E_L, \mu_S) \simeq 1$, and $f_{FD}(E_L, \mu_D) \simeq 0$. Thus, when there are electrons at S and free states at D, so electrons from E_L can go to D and from S to E_L. The drain tries to align $E_{F_{dot}}$ to μ_D to get a situation of minimum energy and so sinks electrons from dot. And in turns the source tries to fill E_L to get$E_{F_{dot}} = \mu_S$.

So, as a result, the I-V curve in this simple model of one discrete level of LUMO-type (similar to CB conduction for semiconductors since it exploits occupied states) is shown in figure 2.18.

Notice that the current is small below $V_{DS} = 1V$ and saturates since at a certain value of V_{DS} greater than 1V, the Fermi functions are: $f_{FD}(E_L, \mu_S) \simeq 1$, and $f_{FD}(E_L, \mu_D) \simeq 0$.

However, if instead a HOMO-type conduction (similar to VB conduction for semiconductors since it exploits free states) is considered, thus for e.g. E_H is below 0.7 eV w.r.t. the Fermi level of the dot $E_{F_{dot}}$, the previous analyzed cases become:

- case (a): if $V_{DS} = 0V \longrightarrow \mu_S = \mu_D = E_{Fdot} = E_F$ (equilibrium condition).

  There is no conduction ($I_{DS} = 0$) since there's no free states at E_H: $f_{FD,S} = f_{FD,D} \simeq 1 \longrightarrow I_{DS} \simeq 0$;

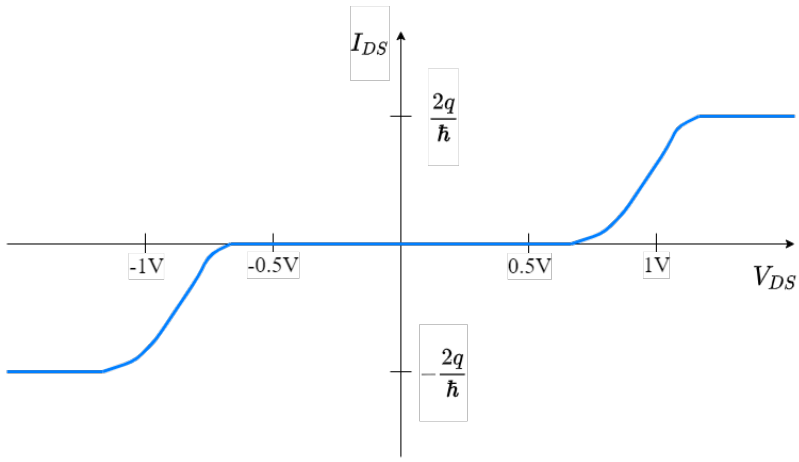


Figure 2.18: The resulting IV curve for LUMO conduction.

- case (b): if $V_{DS} = 0.7eV$, then μ_S is raised up of +0.35 eV and μ_D lowered of -0.35 ev. Therefore $I_{DS} \simeq 0$ since the Fermi functions at E_H are: $f_{FDS} \simeq f_{FD} \simeq 1$.

- case (c): if $V_{DS} = 1.4V$, then the HOMO starts to conduct since it enters in the BW. Indeed, in this case the Fermi functions at E_H are: $f_{FD}(E_H, \mu_S) = 1$ i.e. there are only occupied levels at S, and $f_{FD}(E_H, \mu_D) < 1/2$ i.e. there are some free states at D. By increasing V_{DS} above 1.4V, the the Fermi functions will tend both to zero: $f_{FD} \to 0$ and so I_{DS} will saturate to $\frac{2q}{\hbar}$.

Therefore, similarly for the previous LUMO-type conductio, the final qualitative I-V curve for HOMO conduction will be as shown in figure 2.19.

Notice that it shows two thresholds as LUMO-type conduction. So in conclusion, putting together HOMO and LUMO as in the previous examples, what happens is that LUMO will starts conducting before HOMO because for LUMO the situation $f_{FD_S} \simeq 1$ and $f_{FD_D} \simeq 0$ is verified for smaller V_{DS} while HOMO still sees $f_{FD_D} \simeq f_{FD_S} \simeq 1$.
For larger V_{DS} also HOMO will conduct since $f_{FD_S} \simeq 1$ and $f_{FD_D} \simeq 0$.
Therefore, considering both E_L and E_H what is expected is that each level will independently contribute to conduction when they're within the BW, so the IV curve will be like the one shown in figure 2.20.

However, the bad news is that if the drain current is measured, then the experimental curve is a bit different w.r.t. one shown in the previous figure. That is because the description of conduction by means of discrete levels model is not

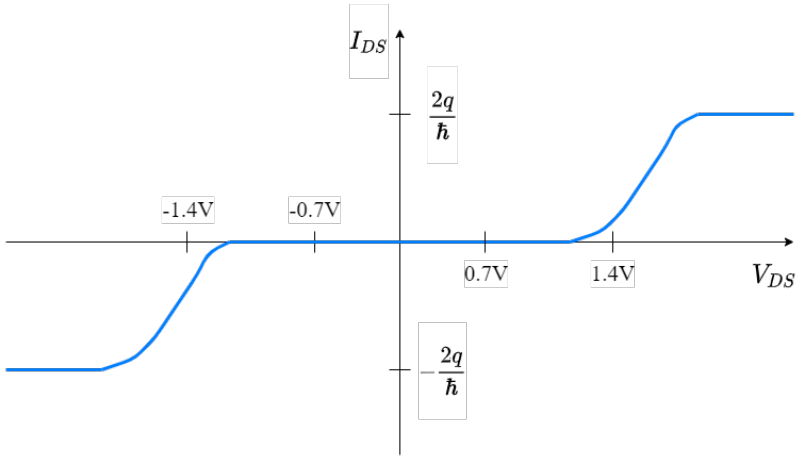


Figure 2.19: The resulting IV curve for HOMO conduction.

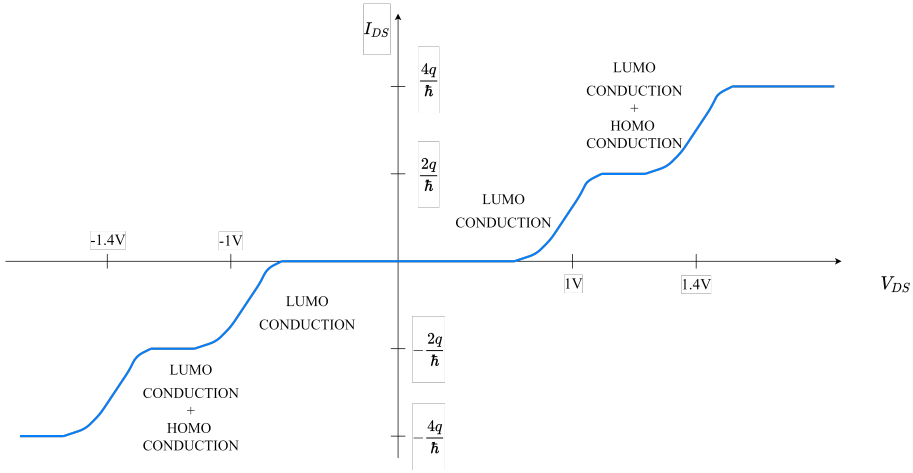


Figure 2.20: The resulting IV curve for LUMO and HOMO conduction.

so accurate. Some corrections need to be introduced to take into account mainly two other effects: (1.) Levels broadening; and (2.) Charging effect. They will be explained in the following.

Levels broadening

It is a phenomenon which consists on making broader the discrete energy levels of the dot. It is due to the influence of the big electrodes (huge reservoir of electron states) on the dot. This effect leads to the delocalization of the wavefunction of the system, which means that the probability density $|\varphi|^2$ is now well spread around the system. It is no more well localized only in the dot, thus now electrons can escape from the dot. Indeed, until now, it has been stated that: for a quantum dot (for e.g. an isolated molecule) the BB approximation can be good enough to obtain qualitative results, since the system is considered closed and the probability to find electrons in the box is 1 and continues to be 1 in time. Considering the example of an isolated molecule, what happens is that the lifetime of a charge in one of its orbital is infinite, and so from the uncertainty principle, infinite lifetimes are associated with perfectly discrete energy states in the isolated molecule.

However, when the molecule is brought in contact with a metal electrode, the electron may actually escape (tunneling is easier) into the metal! The molecule is no more an isolated box, and the electron's lifetime into the molecule is now finite. Now, the associated $|\varphi|^2$ decay exponentially in time and hence the molecular energy levels should also exhibit a finite width. The energy levels are indeed broadened, if a metal-molecule-metal (M-m-M) system is considered. Moreover, the greater is the coupling of the dot with the electrodes, the greater is the broadening of the molecular energy levels (fig 2.21).

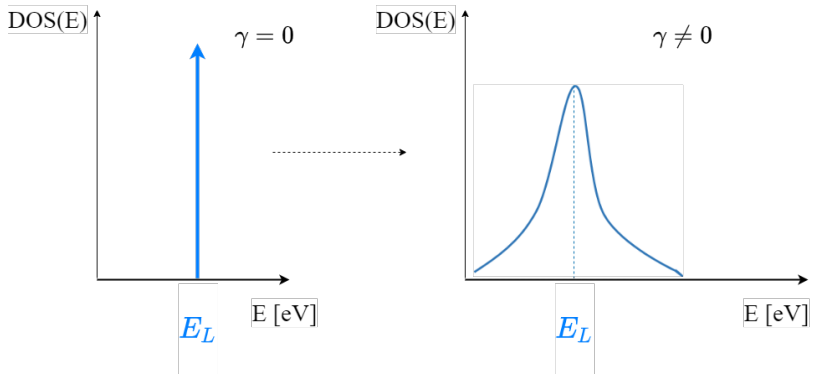


Figure 2.21: The phenomenon of level broadening.

The broadening of the levels are described, in first approximation, with a broadening function having a *Lorentzian* distribution (right side of fig. 2.21) as follows:

$$D_{EL}(E) = \frac{\frac{\gamma}{2\pi}}{(E - E_L)^2 + (\frac{\gamma}{2})^2} \tag{2.56}$$

with $\gamma = \frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}$, to which the width of the *Lorentzian* depends. This expression is the broadening function for the energy level E_L, that now turns to be like a sort of a sub-band, even if the energy level is still one, thus still as maximum 2 electrons (spin up and spin down) can be hosted by it. Indeed, if D_{EL} is integrated, the result turns out to be 2. The only difference is that the energy state is no more well defined in energy but is broadened. Notice that the hypothesis $\gamma \simeq \gamma_1 + \gamma_2$ holds.

Therefore, the previous simple model for conduction can be improved by taking into account this broadening effect, and so the expression of the current 2.54 for a single discrete level becomes for a single broadened level E_L as the following:

$$I_{DS} = \frac{2q}{\hbar}\frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}\int_{-\infty}^{+\infty} D_{EL}(E)\left[f(E,\mu_S) - f(E,\mu_D)\right]\,dE \tag{2.57}$$

sincee now E_L is no more a discrete level, so D_{EL} must be integrated on all energies in order to get the full probability of occupation of the energy state E_L. Notice also that now the Fermi function distribution is applied on generic E, not on E_L. That is because because now it must be applied on the DOS $D_{EL}(E)$ that is not a limited function.

Now, a new quantity which quantifies how much is the probability of conduction of a certain energy state, is defined. This quantity is the transmission spectrum (TS), that for e.g. for E_L is defined as follow:

$$T(E) = 2\pi\frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}D_{EL}(E) \tag{2.58}$$

Thanks to this definition, the equation 2.57 can be rewritten as:

$$I_{DS} = \frac{2q}{h}\int_{-\infty}^{+\infty} T(E)\left[f_{FD}(E,\mu_S) - f_{FD}(E,\mu_D)\right]\,dE \tag{2.59}$$

This equation is the well known *Landauer equation* for a single energy state E_L. However, more in general γ_1, γ_2 are different for each energy level, and thus, as a consequence, each energy level has each own transmission spectrum since broadening can be different for each level. Moreover, not all the transmission peaks are related to a Lorentzian distribution D_{EL}. It is an approximation.

Therefore, in practice, in order to define the Landauer formula for broeadened multilevels strongly coupled 0D-systems, firstly the following quantities must be defined. Since each level has its own broadening, different coupling factors must be defined: $\gamma_{i,j}$, with i label of the level and j label of the contact (1=S source, 2=D drain). With these definitions, the TS becomes:

$$T(E) = 2\pi\sum_i \frac{\gamma_{i,1}\gamma_{i,2}}{\gamma_{i,1}+\gamma_{i,2}}D_{EL,i}(E) \tag{2.60}$$

that is a total TS, as sum of different TS of each level, in the hypothesis that for each level a *Lorentizan* distribution is supposed. In alternative, the NEGF formalism (which does not state a priori any broadening distribution, but it is computed from first principle) can be exploited (see later for a brief overview).
And finally the Landauer formula for broadened multilevels stongly coupled 0D-systems can be written as:

$$I_{DS} = \frac{2q}{h} \int_{-\infty}^{+\infty} T(E)[f_{FD}(E, \mu_S) - f_{FD}(E, \mu_D)]\, dE \tag{2.61}$$

However, is not still enough. Indeed, actually the Lorentzian distribution of the discrete level considered $D_{EL}(E)$ (equation 2.56) depends also on another function: U_{SCF}, that is a potential energy computed by means of a self consistant loop (SCF) that is used estimate the following quantity/effects:

(A). $E_{F_{dot}}$ position when $V_{DS} \neq 0$ i.e. out-of-equilibrium (value that is unknown)

(B). charging effect

(C). gating (i.e. effect on the gate terminal)

The effect of this additional term U_{SCF} is to suitably shift up/down in energy $D_{EL}(E)$ and thus as a consequence also $T(E)$ and so at the end the current $I(V)$.

Therefore, finally, the definitive Landauer formula for broadened multilevels strongly coupled 0D-systems (eq. 2.61), taking into account of this SCF potential energy becomes:

$$I_{DS} = \frac{2q}{h} \int_{-\infty}^{+\infty} T(E - U)[f(E, \mu_S) - f(E, \mu_D)]\, dE \tag{2.62}$$

with

$$T(E - U) = 2\pi \sum_i \frac{\gamma_{i,1}\gamma_{i,2}}{\gamma_{i,1} + \gamma_{i,2}} D_{EL,i}(E - U) \tag{2.63}$$

To estimate this potential U_{SCF}, some words needs to be tell about (A), (B) and (C).

(A). FERMI LEVEL OF THE dot $E_{F_{dot}}$

Only considering a two-probe system, i.e. a quantum dot coupled only with two electrode terminals (not taking into account a third one, the gate terminal), the potential on the dot is the sum of two contributions: $U = U_{V_{DS}} + U_{CE}$, with $CE = ChargingEffect$. The focus now is on the term $U_{V_{DS}}$, the second one will be treated in (B).

The potential term $U_{V_{DS}}$ can be modeled by considering the equivalent circuit of a 0D-system shown in figure 2.22.

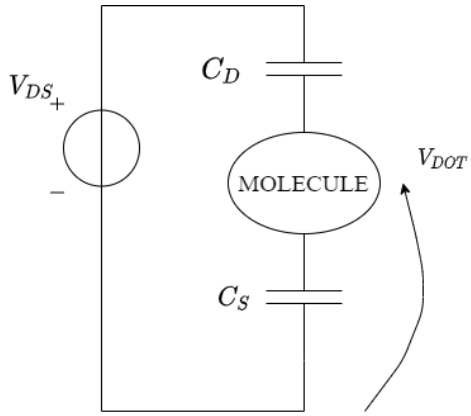


Figure 2.22: Equivalent circuital model of two-probe 0D-system.

The contacts are modeled with $C_{S/D}$ and by performing the capacitive divider, the voltage drop on the dot is the following:

$$\begin{aligned} V_{dot} = V_{DS} &= \frac{\frac{1}{sC_S}}{\frac{1}{sC_S} + \frac{1}{sC_D}} \\ &= V_{DS} = \frac{\frac{1}{sC_S}}{\frac{C_D + C_S}{sC_S C_D}} \\ &= V_{DS} = \frac{C_D}{C_D + C_S} \end{aligned} \tag{2.64}$$

Now, if we have a symmetric structure is considered and the quality of the two contacts are the same, then: $C_S = C_D$ and so $V_{dot} = 1/2V_{DS}$. This means that V_{DS} is equally splitted in two. That is the reason why $E_{F,dot}$ was supposed to fall in the middle of the BW!

Therefore the effect of V_{DS} on $E_{F,dot}$ is:

$$V_{DS} = -qV_{dot} = -q\frac{C_D}{C_{ES}}V_{DS} \tag{2.65}$$

which is also the contribution due to the application of V_{DS} on the 0d-system, that shifts D_E and so the TS $T(E)$. This expression is useful to understand the right position of $T(E - U_{SCF})$ when only U_{DS} effect is considered.

(B). CHARGING EFFECT

It is an effect that can be measured only at nanoscale, not in STD systems, even if it's present also there. It can be explained in the following way. By putting at the same potential the drain and source, i.e. at equilibrium ($V_{DS} = 0$), we want to evaluate the effect of a single electron (actually of 1 more electron w.r.t. equilibrium condition i.e. δn) injected from the electrode into the dot, within the hypothesis of neutral dot. In particular we want to evaluate which is the potential variation in the dot due to this transfer of a single electron w.r.t. equilibrium. We proceed as follows:

- if we model the system as shown in figure 2.23, where the dot is one of the two plates of $C_{S,D}$
- and if the total electrostatic capacitance of the dot is defined as: $C_{ES} = C_S + C_D$ (or $C_{ES} = C_S + C_D + C_G$ in the case of the introduction of the gate terminal)

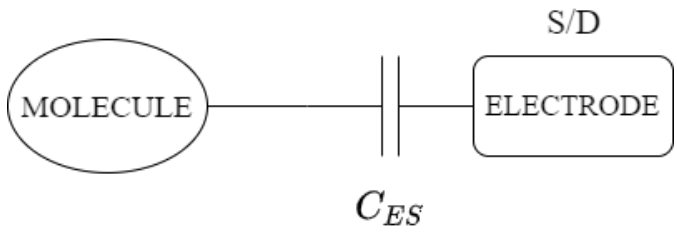


Figure 2.23: Equivalent circuital model of two-probe 0D-system.

what happens is that a quantity of charge ΔQ can be transferred by applying a difference of voltage ΔV: $\Delta Q = C_{ES}\Delta V$, with $\Delta Q = -q$ for one electron. This, in other words, means that by applying a V_{DS}, an electron can flow from the electrodes to the dot, and this flow induces a shift in the energy levels of the dot equal to $U_{CE} = -q\Delta V = \frac{q^2}{C_{ES}}$. This potential shift, in turns leads to a shift in D_{EL}, changing $T(E)$ and therefore at the end I_{DS}.

Therefore the addition of the electron that has jumped into the dot due to the application of drain bias, leads to a shift of all energy levels of the dot up. Therefore it may also happens that also conductive states are pulled up w.r.t. the BW and so they cannot more contribute to the current, blocking in this way the conduction. This is basically the principle of Coulomb blockade (CB). The current will be reduced because when we try to push down E_L with V_{DS} letting fall it in the BW it happens that this level escapes, pulled up by transfer of electron that changes the dot potential.

Actually computing the current considering also this effect is a very complex problem, that needs a self-consistent field (SCF) algorithm in order to be solved. Indeed $U_{CE} = U_0(N - N_0)$ with $U_0 = \frac{q^2}{C_{ES}}$ and $N - N_0$ the difference between the number of electron out-of-equilibrium N and at equilibrium N_0. U_{CE} is the contribution that enters in the TS and must be defined in order to compute the current. But it cannot be find with an analytical solution in closed form, since in turn also N depends on U_{CE} by means of U :

$$N = \frac{2}{\gamma_1 + \gamma_2} \int_{-\infty}^{+\infty} D_{EL}(E - U) \left[\gamma_1 f_{FD}(E, \mu_S) + \gamma_2 f_{FD}(E, \mu_D)\right] dE \qquad (2.66)$$

The only way to solve this problem is by iterating these two equations until a convergence is reached (SCF loop).

(C). GATING EFFECT

If an additional gate terminal is introduced in the system, the equivalent circuital model is the following shown in figure 2.24:

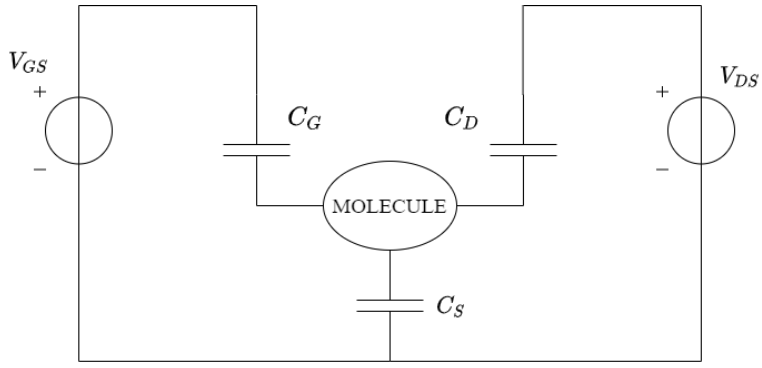


Figure 2.24: Equivalent circuital representation of MolFET.

By exploiting superposition of effects, two independent capacitive divider can be obtained for the voltage V_{GS} and V_{DS}. Therefore the potential drop on the dot due to V_{GS} contribution is:

$$U_{V_{GS}} = -qV_{dot}|_{V_{GS}} = -q\big(V_{GS}\frac{C_G}{C_S//C_D + C_G}\big) = -qV_{GS}\frac{C_G}{C_{ES}} \qquad (2.67)$$

Meaning that if:

- $V_{GS} > 0$, then $U_{V_{GS}}$ is pushed down, shifting towards lower energy values the TS $T(E - U_{SCF})$
- $V_{GS} < 0$, then $U_{V_{GS}}$ is raised up, shifting towards higher energy values the TS $T(E - U_{SCF})$

As a consequence, it may happen that greater/smaller portion of the TS is included in the fixed BW, thus increasing or decreasing the current. Details are provided in section 4.2 for a molecular transistor case study. This term here enters in the total definition of U_{SCF} which depends also on the already discussed contribution $U_{V_{DS}}$ and U_{CE}.

Summarizing, in general:

$$U = U_{VGS} + U_{vDS} + U_{CE} \tag{2.68}$$

with

- $U_{VGS} = -qVGS\frac{C_G}{C_{ES}}$;
- $U_{VDS} = -qVDS\frac{C_D}{C_{ES}}$;
- $U_{CE} = \frac{q^2}{C_{ES}}$.

This means that:

- both V_{DS} and V_{GS} will influence the electron properties of the dot, in particular for positive values they will push down the dot potential U_{dot}, and for negative will pull up it.
- the only contribution that instead will pull up it always is the one due to charging effect.

2.3 NEGF formalism: basics

There are many levels of approximation in modeling conduction through molecules, more or less suitable depending on the applications. Typically, a simple but still physical model, that is also very computationally efficient, is the one based on Landauer formalism described in the previous section, able to describe conduction through multi energy levels broadened with a Lorentzian distribution considering also charging effect. However, this model holds in very specific cases. Indeed real molecules have multiple levels that often broaden and overlap in energy with arbitrary distributions and a proper and rigorous treatment is provided by the Non-Equilibrium Green's function (NEGF) formalism able to do justice also of these behaviours.

In the following, only the **basic set of equations** of the formalism and a ready-to-use explanation of them are provided.

Contact self-energies Σ_1**,** Σ_2 of contact 1 and 2 (drain and source) respectively. They represent the effects of the contacts on the quantum dot (molecule in our case), and they can be though as suitable boundary conditions that take into account the fact that the molecule is not isolated but connected to the contacts:

$$\Sigma_1 = \zeta_1 g_{R_1} \zeta_1^\dagger \qquad , \qquad \Sigma_2 = \zeta_2 g_{R_2} \zeta_2^\dagger \tag{2.69}$$

where:

- ζ_1 and ζ_2 are the contact coupling factors which represent the strength of the coupling between the contacts and the quantum dot, i.e. the molecule. They are intimately linked to the quality of the chemical bonds between molecule (anchoring groups) and the contacts. They are related to the aptitude of moving electrons between contacts and molecule, and of course they are linked with the escaping rate or time, i.e. the average time at which electrons escape from the device towards a contact or vice versa.
  In particular the escaping times are respectively: $\tau_1 = \frac{\hbar}{\zeta_1}$ and $\tau_2 = \frac{\hbar}{\zeta_2}$, where $\hbar$ is the reduced Planck's constant. Notice that $\zeta_{1,2}$ and $\tau_{1,2}$ are matrices, since more energy levels are generally possible, and they can be complex (this is related to finite life-time states).

- g_{R_1} and g_{R_2} are the two surface Green's functions (SGF) of the two contacts. They are essentially the impulse responses of the borders of the contacts (they can be calculated starting from the knowledge on contact structures and only few surface atoms are usually involved in their estimation).

Linearity can be exploited and the effects of both the contacts can be considered together by summing up the two contacts self-energies obtaining:

$$\Sigma = \Sigma_1 + \Sigma_2 \tag{2.70}$$

Broadening functions Γ_1**,** Γ_2 are matrices defined as the anti-Hermitian parts of contact self-energies. They represent the effect of broadening of the isolated molecule energy levels due to the presence of contacts. They are linked to imaginary part of energy (finite life-time states) because the imaginary part of energy is the mathematical tool for describing the broadening:

$$\Gamma_1 = i[\Sigma_1 - \Sigma_1^\dagger] \qquad , \qquad \Gamma_2 = i[\Sigma_2 - \Sigma_2^\dagger] \tag{2.71}$$

$$\Gamma = \Gamma_1 + \Gamma_2 \tag{2.72}$$

where i is the imaginary unit.

Dyson's equation:

$$G^R(E) = [EI - H - \Sigma]^{-1} \tag{2.73}$$

$$G^A = G^{R^\dagger} \tag{2.74}$$

where:

- G^R is the retarded Green's function
- G^A is the advanced Green's function
- I indicates the identity matrix, in the case of non orthogonal basis set it should be substituted with the overlapping matrix S
- H is the device Hamiltonian operator
- E is the considered energy value

Dyson's equation (2.73) can be considered as an operative definition of the retarded Green's function, while equation (2.74) defines the advanced Green's function (it is just the transpose complex conjugate of the retarded one). It is useful to notice that all previously reported quantities are function of energy E, that actually can represents the energy of a «test» electron (or in general quantum particle) injected from contacts into the device, i.e. the energy of an input stimulus to the system. From this remark it should be clear that the NEGF formalism differentiates from conventional quantum mechanics since it corresponds to a forced study of the system (in which the input stimulus, of energy E, is supposed known), instead of a free oscillation study of the system as usually happens in quantum mechanics. The retarded Green's function has the physical meaning of impulse response of the total system (molecule plus contacts), thus it is the system output when the input stimulus is an impulse..

Transmission coefficient T(E):

$$T(E) = Tr[\Gamma_1 G^R \Gamma_2 G^A] \tag{2.75}$$

where:

- T(E) is the transmission coefficient at that energy value E
- « Tr» represents the trace of a matrix

Once the transmission coefficient is known for each energy value E of interest, the current can be evaluated by exploiting the **Landauer's equation**:

$$I = \frac{2q}{h} \int T(E)[f_1(E) - f_2(E)]dE \tag{2.76}$$

where:

- q is the electron charge and h the Planck's constant
- $f_1(E)$ and $f_2(E)$ are the Fermi's distribution functions of the two contacts evaluated at energy value E

Notice that the above reported Landauer's formula includes only ballistic transport, that basically is our case. Indeed considering the nanoscale dimension of a molecule, at first approximation it is reasonable to neglect incoherent scattering [30]. Notice that the power of NEGF formalism lays in the fact that is also possible to model incoherent scattering by means of an additional self-energy Σ_s [29], that is again subtracted to Hamiltonian operator in eq. (2.73).

Part II

Device level

Since the proposal of a single-molecule based device by Aviram and Ratner in 1974, molecular devices have been capable to provide features like conductance switching, rectification and negative differential resistance (NDR).

The aim of the second part of this thesis is to investigate some of these features. Thus some two and three terminals molecular devices (molecular wire, molecular transistor) are presented from a theoretical and simulative points of views, providing also some experimental and fabrication processes related overviews referring to the state-of-art literature.

In particular a full characterization aided by atomistic simulations performed with QuantumWise ATK software [28] of two ambipolar MolFETs, engineered ad-hoc for digital computing applications, is widely discussed in 4.

The main contributions of this second part of this thesis is in the results achieved for both devices w.r.t. previous work found in literature.

- for OPV7-based MolFET an higher $Ion/Ioff$ ratio is obtained thanks to the attention devoted in engineering the device geometry and biasing conditions.
- for PCP-based MolFET a value of around 1400, one order of magnitude higher w.r.t. very few experimental works found in literature, is achieved. This novel MolFET can be considered a promising candidate in computing applications, as will be demonstrated in the next part of this thesis in sections 6.3.1, 7.3.1, 8.4.1.

Chapter 3

Two-terminals molecular device

In general, a two probe system has the following structure (figure 3.1):

- a conductive channel (for e.g. a crystal, a graphene sheet, a molecule or a DNA...), connected on the two side to:
- a donor, able to provide electrons to the channel
- an acceptor, able to accept electrons from the channel

If a perturbation is applied between donor and acceptor, for e.g. an electric field which generates a difference of voltage throughout the system (as in the figure 3.1), and if the channel is conductive and well connected to both donor and acceptor, then a flow of carriers occurs, thus a current is measured.
The donor and the acceptor could be for e.g. typical gold electrodes, graphene sheets, molecules, etc.

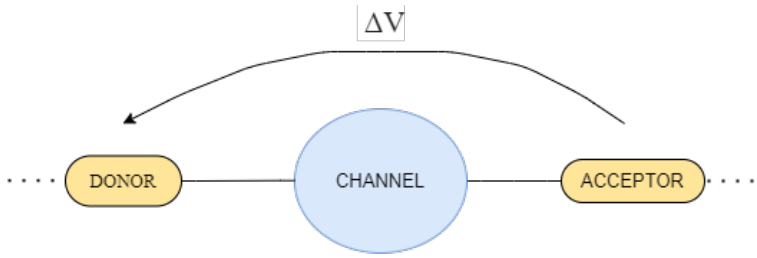


Figure 3.1: A depiction of a two probe system.

The focus of this chapter is on two-terminals single-molecule based device. After a brief review of the state of art, the basic concepts are reported. Then the aim of the next section is to provide a detailed analysis, supported also by ab-initio simulations, on which factors influence the conduction in this structure and qualitatively and quantitatively how.

3.1 State-of-art

How to realize a single-molecule junction, that is the basic unit of a molecular device, has been widely studied since its first discovery in 1974 by Aviram.
However, the challenge of creating this structure, where a single molecule is bonded to two macroscopic electrodes was hard. The situation radically changed with the introduction of scanning probe microscopy technology in the early 1980s. STM and AFM made possible to contact the molecules and measuring the currents flowing through them.
In these three decades, many techniques and methods for fabricate a reproducible and reliable molecular junction or many more in parallel, have been conceived and experimentally demonstrated. The critical problem is the creation of an high quality nanogap of order of few nm into which drop the solution containing the molecules. This can be done by feedback controlled techniques based on electromigration [31], or by the more common break junction techniques [32, 11].
The former consists on breaking a metal wire by applying an electrical signal. Since, basically, thin metallic wire has a threshold current density, beyond this threshold, ions of the metallic wire start to move, until its movement leads to break the wire. However, this technique is not so simple as might be seen, because of many parameters that must be carefully considered for managing the final size and quality of the induced gap. Moreover, is not possible to reach a good surface quality and once the gap is created there is no way to tune it.
The latter, instead, consists on a mechanical break of the metal wire by the inducement of controllable stress. The result is the creation of a very smooth gap at atomic level without bumps. Thanks to the more advanced crack-defined break junction technique presented in [11], also the possibility to create several high-quality nanogaps in parallel has become real.
Once the nanogap is created, the solution containing the molecules is dropped over its surface. If the molecules are chemisorbed by electrodes via anchor groups, a single-molecule junction is formed. Finally, the device is dried leaving the molecule only contacted to the electrodes.
For now, the most astonishing achievement is a fabrication density of 7 million junctions per cm^2, with a fabrication yield of 7% for sub-3nm gaps [11]. However, the way towards very large-scale fabrication break junctions is still long.

Concurrently, in these three decades, the electronic conductance of the single-molecule junctions has been investigated for a huge variety of molecules from simple ones (e.g., hydrogen, oxygen) [33] to complex molecules (e.g., supramolecule, polymer)[34, 35, 36].
Memory, diode, switch, sensors for biomolecules, such as DNA [37], RNA, and protein and other functional properties have been extensively investigated for the single-molecule junctions up to nowadays.

3.2 Molecular wire: basics

A two-terminals molecular device (figure 3.2) consists on a molecular conductive channel contacted with very large electrodes ("reservoirs").
The channel can be a single molecule of few Angstrom in length or a molecular chain (made also of protein, peptide, polymers etc. [34, 35, 36]) with typical diameter less than 3 nm and ranging up to few mm in length. Therefore, according to the channel's extension, the two-terminals molecular device (also called molecular wire) can be classified as a quantum dot (0D-system) or as a quantum wire (1D-system).

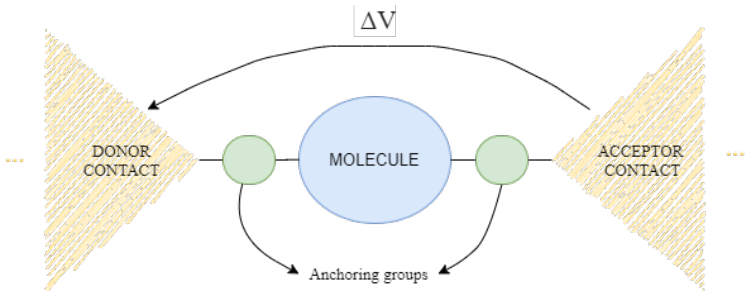


Figure 3.2: A depiction of a two probe molecular wire.

Despite its name, the molecular wire (also referred as *molecular juction* [38]) has a non-linear IV characteristic and does not behave as a simple conventional ohmic conductor. Surprisingly, molecular wires which present an anti-ohmic behaviour (i.e. conductors with reversed conductance/length trend) have been recently discovered, attracting a lot of research interest [39, 40, 41].

Anyway the focus in this thesis is on 0D-molecular wire, in which usually the conductor is a single molecule, consisting on an aromatic ring connected by means of *anchoring groups* (a.k.a. *linkers*) to the electrodes. The anchoring groups have the aim to provide a reliable and mechanically stable chemical bond between the molecule and electrodes. They are not supposed to create any potential barriers in

such a way that any observed behaviour can be directly attributed to the molecule itself.
The most common linkers are thiol groups (-SH). They are exploited since the Sulfur forms a strong covalent bond with gold (and other metals) with which typically electrodes are made of. Thus a good anchoring (i.e. a good coupling factor) is ensured between the molecule and the contacts. Linkers are necessary because naturally molecules does not create bonds with gold.
Other anchoring groups such as amines, nitro, carboxylic-acid and cyano are also employed in literature. However, they have weaker coupling factors with metals if compared with thiols. How different linkers influence the conductance of the molecular wire is briefly analised in 3.3.5.

In single-molecule junctions, the characteristics and the order of magnitude of the conduction, depends on:

- the type and amount of **external perturbation**. Indeed, it is possible to influence the current flowing through the molecular channel with different external sources, such as:
  - an external electric field applied between the electrodes, which generates a difference of voltage mainly dropping across the molecular channel. For very large voltage drop, photoemission may occur (figure 3.3), giving rise to applications like LEDs or lasers. Anyway it migth also induce an unwanted mechanical stress that can lead to deformations, torsions of the molecule or to the break of some chemical bonds.

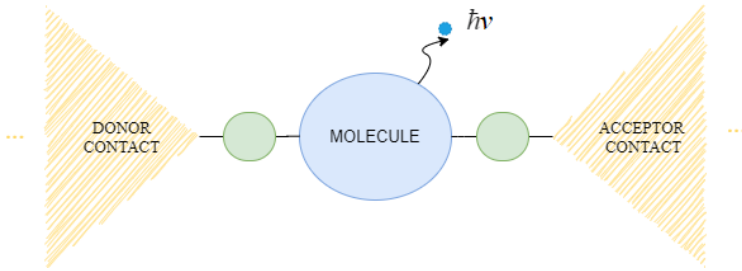


Figure 3.3: A depiction of a molecular system emetting photons.

  - a transversal electric field applied by means of a third terminal (gate electrode) as shown in figure 3.4. Hence, beside the donor and acceptor that work respectively as source and drain, another electrode is added along the length of the molecular channel, in order to obtain an equivalent molecular field-effect transistor. The gate electrode basically

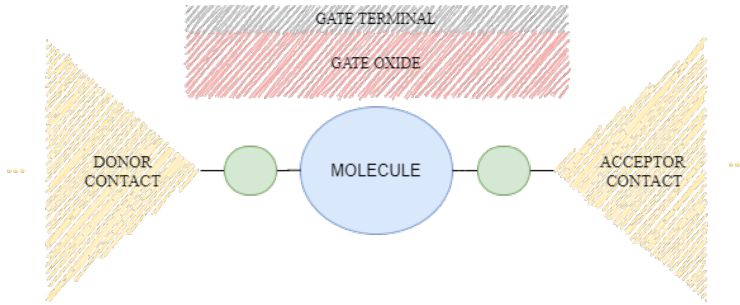


Figure 3.4: A depiction of a molecular FET with a single gate terminal.

modulates the current flowing along the molecule by shifting up and down its energy levels falling within the bias window. The gate can be made of a metal conductor or of another molecule. In the former case it is electrostatically coupled with the channel and a typical Field-Effect transistor (FET) is obtained, whereas in the latter is chemically coupled and a sort of chemical-effect transistor is get (Mol-CET). In any case, what is important is the coupling strength α between this additional terminal and the molecule, mainly determined by the distance, the type of materials/molecule exploited and the bias applied.

– a photon injected towards or in proximity of the molecule (as shown in figure 3.5), that may give rise to photo-voltaic or photo-switch applications. Indeed, the photon injected can eventually turn on the device as a switch by modifying the energy levels of the molecule.

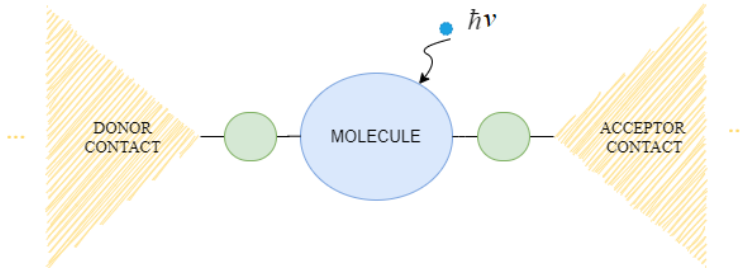


Figure 3.5: A depiction of a molecular photoswitch.

– an external compound (gas, biomolecule, etc,as shown in figure 3.6)) that, by eventually creating chemical bonds with the molecular channel, can drastically change the amount of current, making the system

behaves like a receptor at nanoscale, thus giving rise to sensor/biosensors applications.

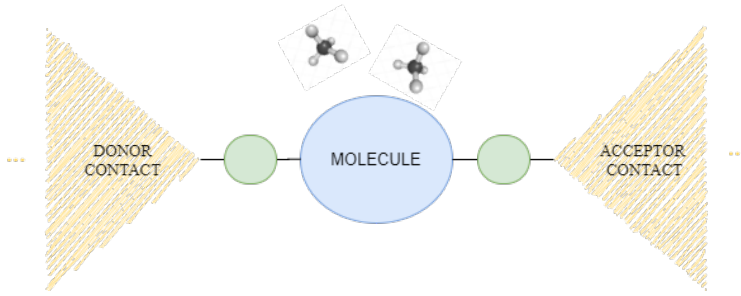


Figure 3.6: A depiction of a molecular biosensor of methane gas.

- the type of compound with which the **molecule** is made of.
- the **size** of molecule.
- the **coupling factor** between the electrodes and the molecule. It is determined by the quality of the interfaces and contacts and the effectiveness of the anchoring groups. Depending on these factors and also on the distance between the molecule (that might be also not anchored by means of linkers) and the electrodes, two different type of different ballistic or quasi-ballistic transport processes can occur:
  - Hopping or Sequential tunneling if the coupling is weak, i.e. if the electron cloud - electron wavefunction - is well localized in the molecule. In this case, physical phenomena like *Coulomb Blockade*, as happens in Single-Electron Transistors (SET), can be present.
  - Tunneling if the coupling is strong. This is the case widely analysed in the previous part of this thesis concerning conduction in 1D/0D - systems.
- **orientation** of the molecule w.r.t. the electrodes.

Notice that the type, size of molecule and also its orientation deeply influence the DOS distribution since different type of molecular orbitals (i.e. electron clouds or wavefunctions) are obtained. The coupling factor, instead, affects the broadening of the DOS distribution.
The fact that all these factors influence the conduction can be seen as a problem in understanding it, but actually is an enormous added value, since it introduces many degrees of freedom, and different physical effects and thus a huge flexibility

and variety of applications can be obtained, if compared with conventional electronic devices. Indeed, by keeping basically the same fabrication process and only by changing for e.g. the type of molecule or the type of linker different devices can be achieved, differently from standard devices, where even changing the type of material for e.g. moving from Si to GaAs, the same IV behaviour is get with only an enhancement in mobility.

Instead, the problem in this scenario characterized by many degrees of freedom is how to handle the process variations of these huge amount of factors. How this variability can impact in the conduction and how to model it? Indeed, there may be defects in the molecule, in the contacts or in the linkers (vacancy, impurities, bumps in the surfaces of the contacts...), leading to a different level of current w.r.t. the expected one. There may be problems in the chemical bonding between the anchoring groups and the molecule or between the anchoring groups and the contacts, leading to a deviation to the expected coupling factor and thus to the current. There may also be problems in the relative orientation of the molecule w.r.t. the electrodes or anchoring group or of a molecular ring w.r.t. previous or the next in the molecular chain, and so on.

In conclusion, in order to make reliable and reproducible these type of molecular devices, as typically happens, a Safe-Operating Area (SOA) that takes also into account process variations must be defined. Moreover, the systematic control of the atomic coordinates of both the molecules and the atoms of the contacting leads is required. Also, molecular junction requires the repetition of thousands to millions of molecules in parallel, with a good reproducibility. However, at research level, molecular electronics is not yet at this point.

3.3 Conduction in molecular wire

The aim of this section is to provide an simplified overview linked to the more rigorous treatment reported in the first part of this thesis, on how conduction through molecules is possible. Finally, a useful analysis of the work of Zahir *et al* [42] which focuses on how some of the previously discussed factors influenced the conduction in molecular wires, is commented. The purpose is to practically understand how actually complex is the understanding of transport in these type of systems. This analysis is based on atomistic simulation results, since these aspects are not well characterized by any theoretical model.
The focus will be on geometrical variations of the molecular wire subjected to an applied bias. In particular on the type of molecule exploited 3.3.2, on the molecular channel length 3.3.3, on the type of linkers or electrodes used 3.3.5, and on torsion3.3.4. This analysis was aided by means of atomistic simulations performed in ATK.

For all type of case study, molecules were firstly built and geometrically relaxed,

using Conjugate gradient algorithm, until all residual forces in each atom are less than 0.05 eV/Å . Then the molecular junction was created, by attaching each molecule to a semi-infinite (3x3) (111) gold electrodes through thiol bonds (-SH), as shown in figure 3.7 for the thiopene molecule. The distance between the sulfur atoms and gold surface was fixed to 1.71 Å which corresponds to S-Au bond length of 2.38 Å, value validated experimentally in [43].

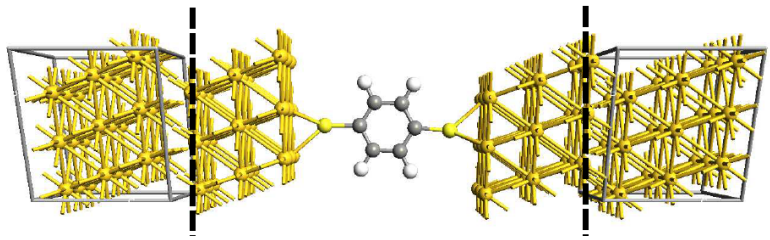


Figure 3.7: A typical structure of molecular junction for one ring of thiopene molecule (1TT). The central part is the active part of the device also called *scattering region*. The portion of electrodes between the two black dashed lines are called *leads*. They influences the broadening of the energy levels of the molecule, thus influences a lot the conduction. The electrodes shown are finite but actually, in order to be realistic, are replicated to become semi-infinite thanks to the imposition of PBC in the transversal directions.

Then, for each case of study, firstly the transmission spectrum at equilibrium and finally the IV characteristic (i.e. the integral within the bias window of the transmission spectra at the different bias voltages) were computed.
NEGF formalism coupled to DFT was exploited for these transport simulations. In particular, the electronic characterization at equilibrium resulting from the DFT were fed into the NEGF model to calculate the drain current self-consistently by means of the Landauer formula 2.76. For all cases, the employed boundary conditions in the Conjugate Gradient Poisson solver were fixed to Dirichlet along the transport direction and Periodic directions perpendicular to the transport.

3.3.1 Simplified modeling based on Landauer formalism

In this subsection some basic concepts on how to model conduction in molecular wire devices are reported. Th conduction model explained in very practical way is the one referring to Landauer formalism in the case of a quantum dot strongly coupled with electrodes, i.e. in the case of a ballistic type of transport mediated via the broadened DOS of the molecular system.
In general the structure of a molecular wire is the one shown in the depiction 3.2 and in figure 3.7. For sake of simplicity, firstly the case of conduction mediated via

a single discrete level E_L without no broadening is considered. The current flowing through the dot, as demonstrated in the first part of this thesis in the subsection 2.2.2 results to be:

$$I_{DS} = \frac{2q}{\hbar}\frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}\cdot 1\big[f(E_L,\mu_S) - f(E_L,\mu_D)\big] \tag{3.1}$$

with γ_1 and γ_2 coupling factors respectively of source (S) and drain (D). Notice that 1 is highlighted since only a single energy level is considered. This equation tells that:

- if the energy level E_L of the molecule (dot) is within the bias window (CASE 1 in the figure 3.8), i.e. between the Fermi levels of the source (μ_S) and the one of the drain (μ_D), it will contribute to the conduction. Indeed, for e.g. for a drain-source voltage $V_{DS} > 0$ V the Fermi function distribution of the source $f(E_L,\mu_S)$ is almost close to 1 and the one of the drain $f(E,\mu_D)$ almost close to 0 (supposing the tempearture at 0 K), hence the drain current becomes:
$$I_{DS} = \frac{2q}{\hbar}\frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2} \tag{3.2}$$
- if E_L is very different from the Fermi levels of source and drain (i.e. E_L» μ_S , μ_D or E_L « μ_S , μ_D) and thus the energy level is far away from the bias window, then it will not contribute to the conduction. (CASE 2,3 in the figure 3.8)

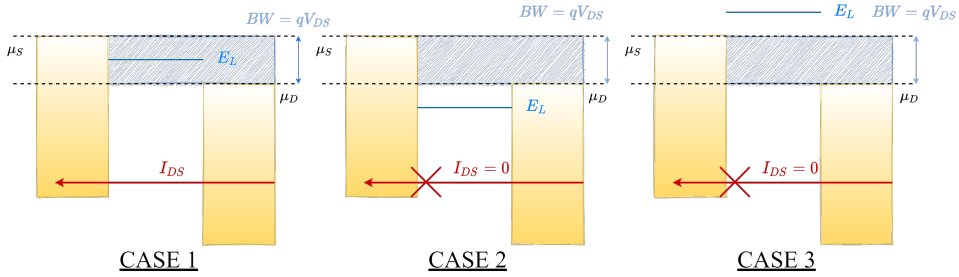


Figure 3.8: Case 1,2,3 desvribed above. The voltage division factor $\eta = 1/2$ shifts the Fermi level of the source $qV_{DS}/2$ above and the Fermi level of the drain below. In case 1: the enegy level of the dotE_L is within the BW therefore $I_{DS} > 0$. In case 2 and 3 instead is not included in the BW threfore the drain current $I_{DS} >= 0$, because there are no available states for the electrons in the molecule, thus there is no possible fluxes from the dot towards the electrodes and viceversa.

However, in more realistic molecular devices, there many energy levels (available energy states) and moreover when a molecule is strongly coupled with electrodes, broadening of these energy levels occurs. Therefore the $DOS_{0D}(E)$ function it is typically approximated with a *Lorentzian* distribution centered in each energy level as the following:

$$D(E) = \frac{\gamma/2\pi}{(E - E_L)^2 + (\gamma/2)^2} \tag{3.3}$$

Therefore, in this case of multiple broadened energy level molecular system, the Landauer formula becomes:

$$I_{DS} = \frac{2q}{h} \int_{-\infty}^{+\infty} T(E)\big[f(E, \mu_S) - f(E, \mu_D)\big] dE \tag{3.4}$$

with $T(E) = 2\pi \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} D(E)$ transmission spectrum. Notice that this is actually a simplified formula of the transmission spectrum. The NEGF formalism, instead, is more accurate since it takes also into account the fact that the transmission spectrum depends both on drain and gate bias, thus typically changing the drain and gate bias, also drastically very different transmission spectra are obtained. This can be also shown by means of the dependence of the transmission spectrum to the SCF potential $U_{SCF} = U_{V_{GS}} + U_{V_{DS}} + U_{CE}$. Indeed by varying V_{GS}, V_{DS}, $T(E - U_{SCF})$ will varies. The three different cases analyzed before, for a single broadened level, now turns to be as shown in the following figure 3.9:

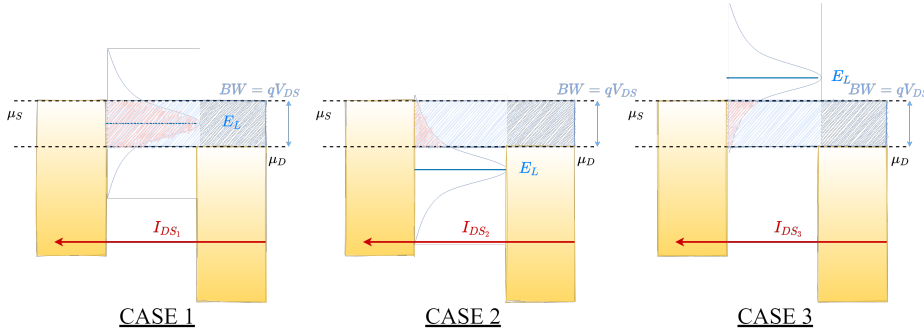


Figure 3.9: Case 1,2,3 described above in the case of a single broadened level. The voltage division factor $\eta = 1/2$ shifts the Fermi level of the source $qV_{DS}/2$ above and the Fermi level of the drain below. In this case $I_{DS} \neq 0$ in all the three cases, since always a portion of the transmission spectrum is included within the bias window. In particular: $I_{DS_1} > I_{DS_2} > I_{DS_3}$ since less and less amount of the $T(E)$ is integrated to obtain the total current.

More in particular what happens for the current in the case of a transmission spectrum shown in the following figure3.10, is that by enlarging more and more the bias window, wider portion of conductive states with transmittivity $T(E)$ are included and integrated to get the drain current flowing through the molecular system. Notice that these considerations are approximations, since the following suppositions not always so realistic hold: the transmission spectrum profile does not change varying the drain source bias V_{DS}, and the temperature of the system is $T = 0K$ so that all the states below the source and drain Fermi levels are occupied. The drain current flowing through the molecular wire obtained by enlarging more

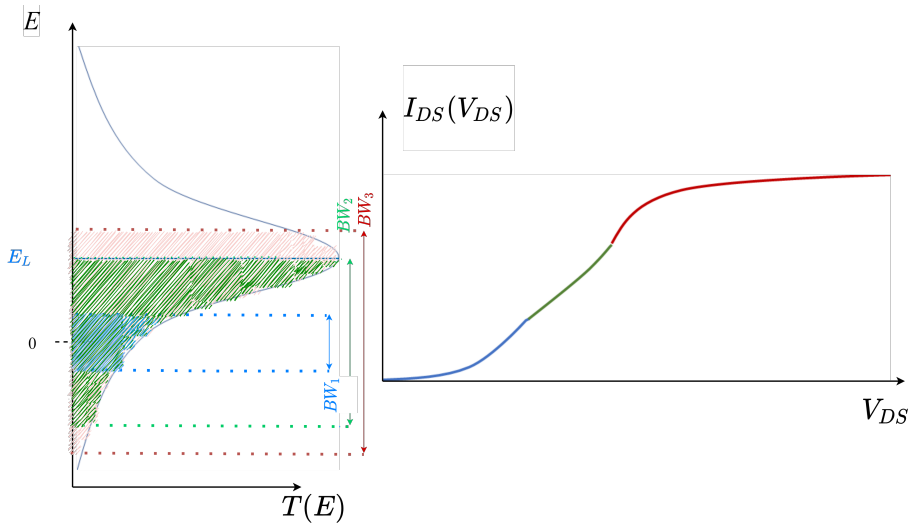


Figure 3.10: Transmission spectrum profile $T(E)$ on the left side and the corresponding drain current on the right side. The Fermi level of the system is placed at the refernce value 0.

and more the bias window. The portion blue of the curve corresponds to the first BW. Small values of current are obtained since the integral catch few conductive states with little bit of transmittivity associated. The green is associated to the second BW: more states with greater transmittivity are integrated so greater values of current are reached. At last the red portion correspond to the third BW. The current saturates after the peak is overcame. A flex point is obtained in the IV curve since the $T(E)$ starts to decrease. The drain current keeps rising since there is still area of $T(E)$ to be integrated, until for a negligible additional amount a saturation is reached. This means that if the BW is enlarged (i.e. V_{DS} is increased) roughly

the same amount of drain current is obtained.

The previous figures are related to a single broadened level, however molecules have multiple energy levels, so actually real examples of $T(E)$ are more like the ones shown in figure 3.11:

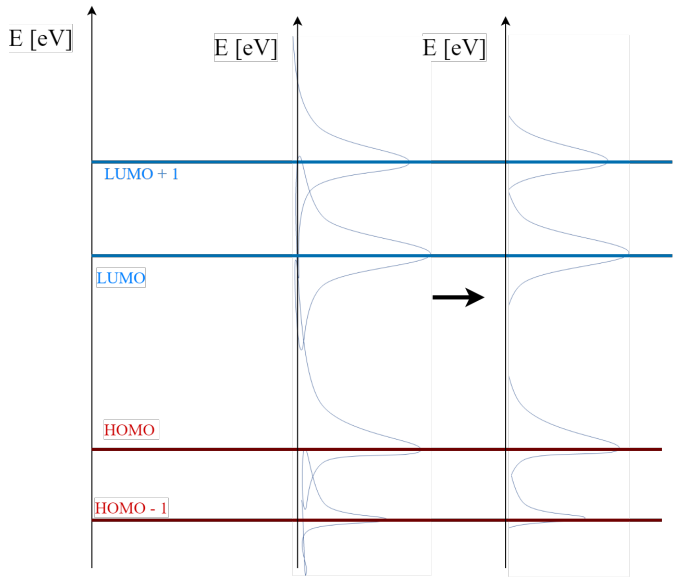


Figure 3.11: Discrete energy levels distribution on the left side, broadened levels on the left side result of the superposition of the single broadened levels.

It is extremely useful to understand what happens in the IV curve if more than one level are considered. For sake of simplicity, firstly the case at $T = 0K$ and no broadening is considered. What results is the following shown in figure 3.12:

- for the BW_1, no energy levels are included therefore the drain current $I_{DS} = 0$
- for some volts above the voltage corresponding to the BW_2, two energy levels (HOMO and LUMO) starts conducting and therefore the drain current $I_{DS} \neq 0$
- by enlarging the bias window until BW_3 no more additional energy levels are included w.r.t. previous case, so the current remains the same
- enlarging again the BW two additional energy levels are included ($HOMO_{-1}$, $LUMO_{+1}$) and so the resulting current is increased since overall four energy levels are conducting.

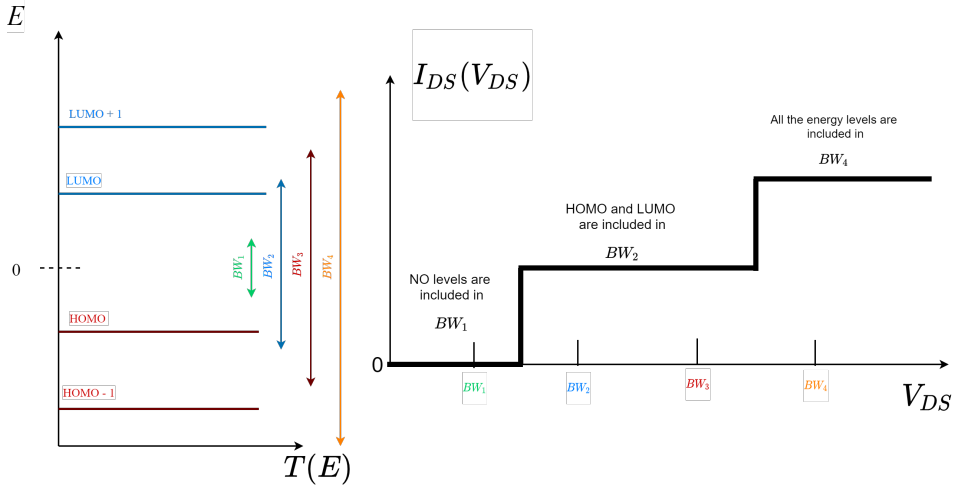


Figure 3.12: Discrete energy levels distribution on the left side, the resulting IV curve on the right side.

If broadening is considered, energy levels transmittivity are no more deltas but have a *Lorentzian* distribution, thus the resulting current has no more a step-like trend but a smooth one. The more is the broadening (i.e. the more is the coupling of the dot with the electrodes) the smoother is the IV curve, since the lower and the wider are the transmission peaks. Moreover, the influence of the temperature makes the IV curve even smoother, since the Fermi function distribution is no more a step like function.

3.3.2 Dependence on type of molecule

One of the most important driver of the development of molecular electronics is the identification and eventually synthesis of suitable molecules for conduction. The main criteria which discriminates the suitability of a molecule for the conduction is the magnitude of the HOMO-LUMO Gap (HLG), that is the energy difference between the highest occupied molecular orbital (MO, that can be seen as a linear combination of atomic orbitals in the LCAO theory) an the lowest unoccupied one. In practice, it can be seen as the equivalent of the energy gap in semiconductor theory, and so the HOMO and LUMO levels are respectively to molecules roughly what the valence band maximum E_V and conduction band minimum E_C are for semiconductors. In the following figure 3.13, a depiction of an example of energy

levels distribution. The HLG depends on the molecule. If the molecule is excited

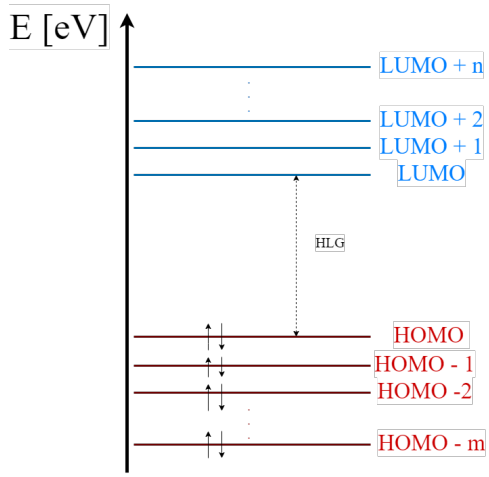


Figure 3.13: An example of energy levels distribution of MOs of a molecule. The energy level $HOMO - m$ is the ground states. All the HOMO levels are occupied at equilibrium, whereas the LUMO levels are unoccupied.

by an external perturbation (i.e. an exchange of energy is present), is it possible for electrons to jump from HOMO to LUMO, depending on the HLG and the intensity of the external perturbation. Obviously, the lower is the HLG, the greater is the probability to promote electrons to UMOs and so the greater is the possibility to have conduction when a bias is applied.

A wide variety of molecules have already been studied and determined to be suitable for molecular electronics. Among these, hydrocarbons have been extensively investigated and selected as appropriate molecules. They are organic compounds made up of carbon and hydrogen atoms only. Carbon atom is unique in the periodic table for its ability to easily concatenate with other carbon atoms forming single/double/triple bonds with another carbon atom with very good stability. As a result, very long and stable chain of hydrocarbons are formed in nature at room temperature.
Molecular wires can be classified into two main types: saturated hydrocarbons chains and unsaturated or conjugated hydrocarbons chains.
In saturated hydrocarbons, atoms are connected only with single bonds and saturated with hydrogen. For e.g. the pentane (figure 3.14) and all alkanes in general are considered to be poor conductors since their HLG) is very large [44], around 8 eV. This is due to their geometry, they are open chains characterized by single C-C bonds, where each C atom is sp^3 hybridized with four σ bonds. So they are

not good for conduction in molecular electronics, since in order to make possible conduction a very high bias voltage would be required. Basically they behave like insulators due to σ bonds.

Instead unsaturated hydrocarbons are characterized at least by a C-C multiple

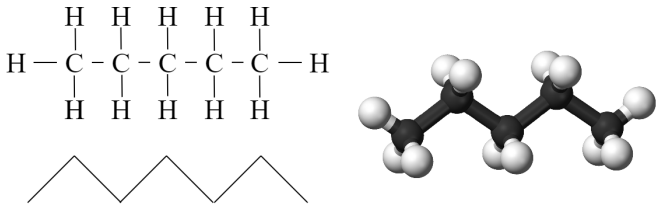


Figure 3.14: Chemical formula of Pentane C_5H_{12}. It belongs to alkanes, a type of acyclic saturated hydrocarbon.

bonds. Among them, there are alkanes, characterized by double bonds (C=C) and alkynes by triple bonds (C≡C). For e.g. the ethene molecule (C_2H_4), thanks to its three orbitals with sp^2 hybridization, presents two π orbitals where electrons are shared and no more linked to specific C atoms (figure 3.15).

These types of molecules have smaller HLG gaps than alkanes and thus are more

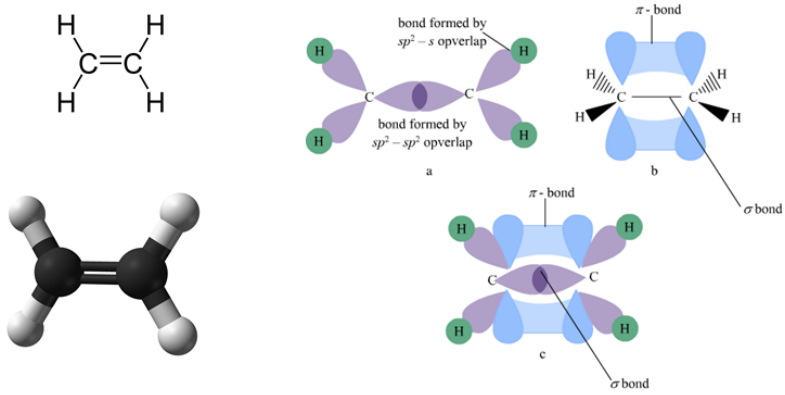


Figure 3.15: Chemical formula of Ethene C_2H_4. Representation of the molecule with balls and stick format, and on the rigth side te depiction of orbitals hybridization and formation of two pi-bonds.

effective in the electron transport [44].

However, among hydrocarbons there are also cyclic structures, and among them a specific class is fundamental for molecular electronics: the so called *aromatic hydrocarbons*. They have at least one ring of saturated carbon atoms with single or double bonds. The most common example is the benzene (C_6H_6) shown in figure 3.16. Also in this case π-bonds are created among the p-orbitals. They are organized in a unique clouds on the top and on the bottom of the molecule, where the electrons are delocalized and free to move because they are shared between carbon atoms. Since, in these delocalized states, electrons can move freely, benzene has very good conduction properties [45].

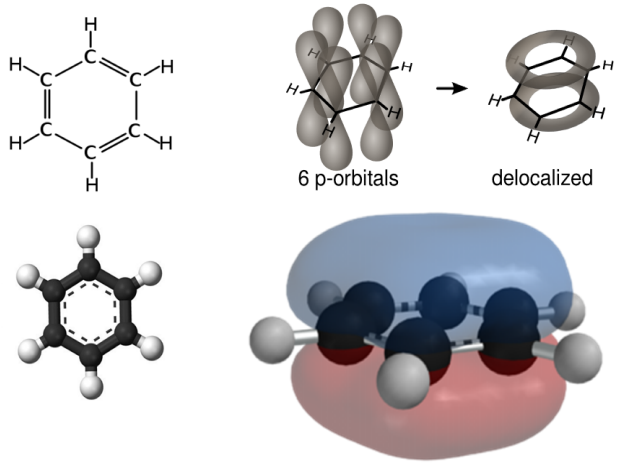


Figure 3.16: Chemical formula of benzene C_6H_6. Representation of the molecule with balls and stick format. On the right top the depiction of orbitals hybridization and formation of two pi-bonds, whereas in the right bottom the electron cloud.

However, the most widely exploited molecular families for molecular devices are based partially on the molecular structure of benzene. They are oligo thiophenes (OT), oligo phenylene ethynylene (OPE), oligo phenylene vinylene (OPV) [46, 47].

OPE derivatives are conjugated molecules with a rod-like shape. Its HLG is around 3.5 eV, much lower than alkanes molecules. Electrons are delocalized over the length of molecule. For these properties, OPE presents a greater efficiency in the charge transport w.r.t a saturated molecule. Its elementary block that can be replicated to form a chain of OPE up to around 5 nm in length [48], is shown in the figure 3.17. OPE blocks are thiolethynylene-terminated and attached on a gold surface by means of -SH thiol groups. Ethynylene groups are attached to protect the molecules from the reactive surroundings. One may think that longer is the molecule, the greater will be the number of electron shared able to contribute to

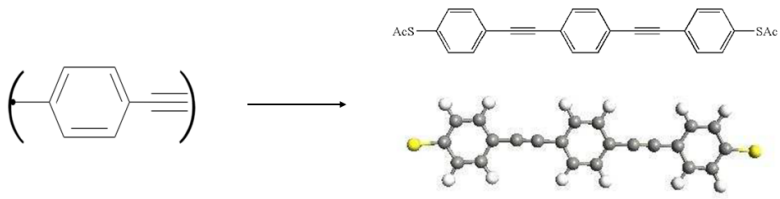


Figure 3.17: Elementary block of OPE, made of a phenyl- and ethynylene (-C≡C-) functional groups. The phenyl one is a benzene ring derivative without either 1 if the point in the figure is the beginning of the chain, or 2 atoms of H if it is in the middle of the chain. On the right side a chain of three Oligo Phenylene Ethynylene ended with thiol (OPE3).

conduction and so the better will be for the conduction. But, unfortunately, this is not true because of the many effects which counterbalance the benefit due to the increase of shared electrons. This effect will be analyzed better in the specific section related to the dependence of the molecular channel length on the current. Anyway, notice that, longer is the molecule, the easier is to synthesize it and to fabricate the nanogap that will host it. On the other hand, unwanted issues like torsion will be more probable, terribly affecting the conduction. Indeed, torsion of the basic blocks of OPEs w.r.t. each others may lead to a molecular wire no more coplanar, which basically means a molecular wire whose orbitals are no more aligned to create π bonds, thus delocalization of electrons. As consequence the current will be lower because the conductive channel path will be no more so well defined, and electrons are no more eased to tunnel from one basic block of OPE from the next.

OPV derivatives can also be used in molecular wires. Also here electrons are delocalized along the length of molecule. Its elementary block is shown in figure 3.18. OPV block unit is made of phenyl functional group like OPE but terminated with a vinylen functional group. The presence of vynilene double bond opposes greater resistance to rotation of the aromatic rings, which makes OPV molecule more planar and more immune to torsion if compared with OPE molecule. In the case of OPV molecule, the bonds between the carbon atoms are shorter. This leads to a decrease in the HLG to around 3.1 eV, thus the conductance is one order of magnitude higher w.r.t OPE. For these reasons, OPV molecules are better suited for conduction than OPE molecules.

OT derivatives are the most used in molecular electronics. They are conjugate molecule whise basic unit is made of 4 carbon atoms and a sulfur one. The delocalization of electrons is present thanks to the π bonds all over the ring of thiopene. Its HLG is the lowest if compared w.r.t. OPE and OPV. It is around 2.9

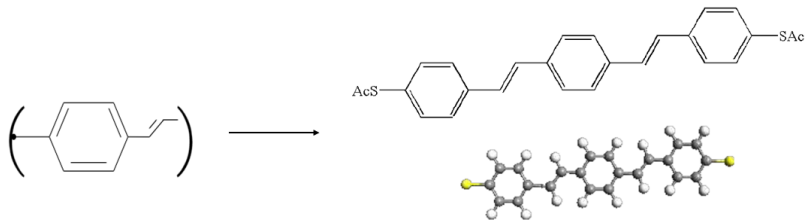


Figure 3.18: Elementary block of OPV, made of a phenyl- and vynilene functional groups. On the right side, a chain of three OPVs ended with thiol (OPV3): dithiol-3oligovynil.

eV, making this type of molecule a better candidate for conducting wires, since a smaller HLG indicates greater conduction. Its elementary block is made of a single thienyl functional group and it is shown in the figure 3.19, where also a chain of thiopene rings is reported. OT blocks are attached on a gold surface by means of -SH thiol groups. They are suitable for long molecular wires.

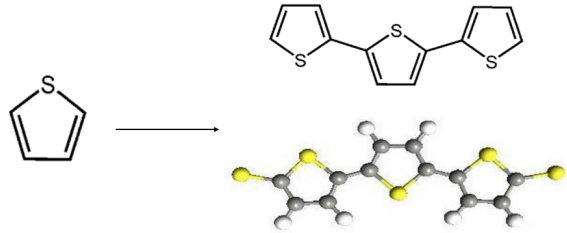


Figure 3.19: Elementary block of OT. On the right side, a chain of three OTs ended with thiol (3TT): dithiol-3thiophene.

Recently, researchers have focused the attention to molecules which exhibit Quantum Interference (QI) effect. Quantum interference can be easily though as destructive interference of electron wavefunctions such that no coherent tunneling can happen in the molecule when small enough voltages are applied. An investigated molecule which shows QI is the pseudo-p-bis((4-(acetylthio)phenyl)ethynyl)-p-[2,2]cyclophane commonly called PCP (figure 3.20). QI effect is basically exploited to turn off the device in a transistor configuration. Indeed, in molecular transistor, the problem is not obtain an high conductance at ON state, but to break down the current at OFF state, in order to obtain an high $Ion/Ioff$ ratio, that barely reach 10^1 with conventional molecules (OPE, OPV, OT).
The PCP molecule presents two separated aromatic rings linked by saturated

methylene bridges (red portion). Intuitively these two methylene links have the role of interrupting the conjugation of the PCP molecular π orbital, and since they are saturated they do not conduct, thus a reduction of transmission occurs. More precisely a current path mediated by π-π overlap between aromatic rings acts in parallel with current paths through the methylene bridges to create destructive quantum interference between the HOMO and the HOMO-1 (figure 3.20), that combined with the interruption of the conjugation leads to a reduction of the transmittivity [49].

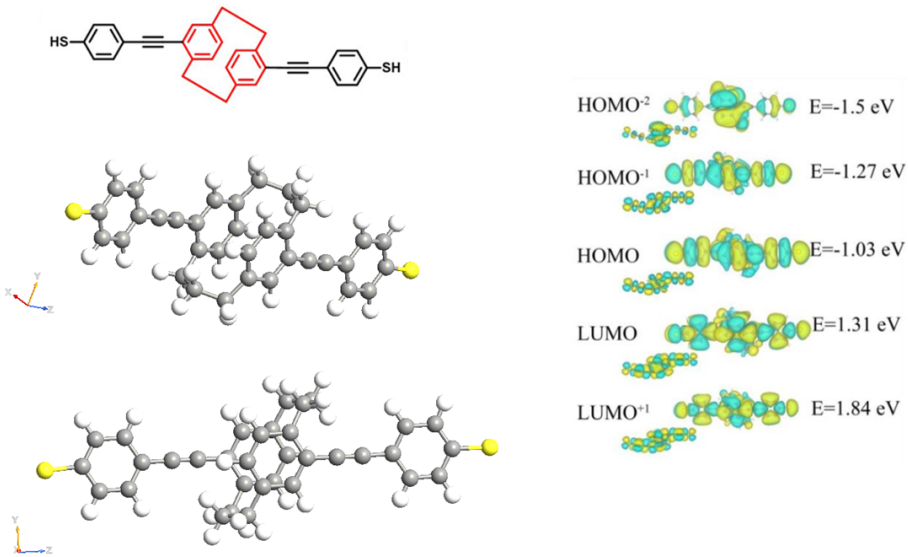


Figure 3.20: PCP molecule chemical structure on the left side. PCP molecule HOMO-2, HOMO-1, HOMO, LUMO, LUMO+1 orbitals representation o the right side.

This molecule will be employed in the next chapter 4 as quantum dot in the design of an ambipolar molecular transistor.

3.3.3 Dependence on the length of molecular channel

In this subsection, how the conductance of the molecular wire is influenced by the length of the molecular channel, is analyzed. The molecule exploited in this analysis, is the thiophene-based molecule. In particular a dithiol-thiopenyl ring (i.e. a phenyl ring terminated with thiols -SH) is firstly considered, then what happens to the drain current when the number of rings increases, is analyzed. So, one ring (1TT), two rings (2TT), and so on, until five rings (5TT) of thiopene, are replicated and the conduction is simulated for each geometry, and finally compared. Notice that in all the five cases nalyzed, the molecular channel length is short enough so that no hopping mechanism takes place.
First of all, is of fundamental importance analyze the equilibrium transmission spectrum of the different molecules: 1TT, 2TT, 3TT, 4TT, 5TT, whose molecular structures are shown in figure 3.21). It is reported for each molecular length together with the DOS in figure 3.22.

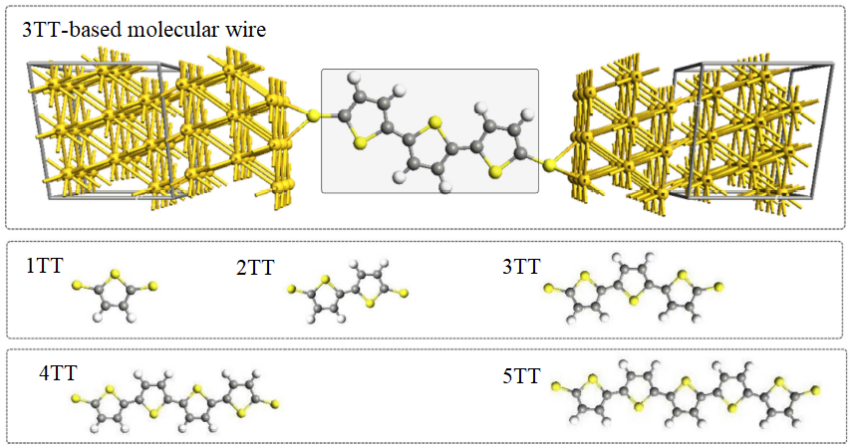


Figure 3.21: On top a typical structure of molecular junction. The molecule rappresented in the scattering region is the dithiol-trithiophenyl (3TT) Thiophene. On the bottom different OT molecules with increasing number of rings.

From figure 3.22, is it possible to understand how molecular length affects the equilibrium transmission spectrum. By increasing the number of rings, the following trends occurs:

a) reduction of the HLG: in 1TT case it is almost 2 eV, whereas in 5TT molecule it is about 1 eV. This effect can be explained in the following way: increasing the molecular channel length means increase the number of rings, hence,

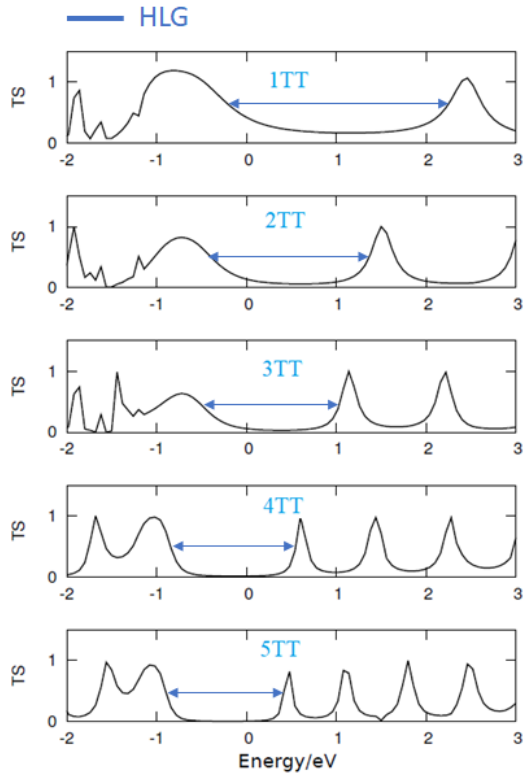


Figure 3.22: Equilibrium Transmission spectra of thiophene molecule with different number of rings. For all molecular systems the average Fermi-level is placed at 0 eV. The peaks on the left side of the Fermi level are the HOMOs and on the right are the LUMOs.

as a result, an increase of MOs/states that makes smaller the HLG. This obviously could be a very good news for conduction, since the channel will starts conducting for lower value of V_{ds}, or more in general, for a given bias window, the current will be greater, since a greater portion of transmittivity will be integrated (remember Landauer equation). However, actually the reduction of HLG does not strictly implies the increase of the current, that is beacause, concurrently the peaks are shifted in position and so w.r.t. the Fermi energy level located at 0 eV.

b) increase of the number of transmission peaks. This effect can be explained in the following way: as already said, increasing the molecular channel length

means increase the number of MOs/states, since for each ring, π orbitals overimposed with others. As a result, an increae on the DOS of the system leads to more electrons that can contribute to conductions, so more conductive states (i.e. more transmisison peaks).
Also this is good for conduction, since with the same bias window, more peaks will be fall into it and so will contribute to conduction, thus increasing the current flowing through the wire.

c) reduction of the broadening: the transmission peaks becomes tighter and tigther. This means that the coupling factor γ between the electrodes and the molecule, becomes smaller and smaller and thus less and less charge is injected into the channel from the reservoir electrodes. This obviously is not a benefit for the conduction since a reduced amount of transmittivity is integrated between the bias window, and so less drain current is obtained.
This result is in agreement with experimental and theoretical findings for which the broadening function exponentially decays with the length of the molecular channel [50, 29]. Indeed, in short molecules, the electrodes contributes to the the DOS of the system, thanks to the strong overlap of molecular and electrodes states. This leads to an increase of the transmission valley near to Fermi-level (Ef = 0 eV) and thus of the current. Instead, in large molecules, the separation between the two electrodes are larger and their contribution to DOS of the system become weaker because of the weak overlap of molecular and electrodes states. As a consequence narrower transmission peaks are obtained and hence the current is reduced.

Therefore, the increase in length contributes to two contrasting phenomena. Indeed (a) and (b) effects leads to an increase of the drain current I_{DS}, whereas on the contrary, (C) effect leads to a reduction. Hence, it is very difficult, only known a priori the information about the molecular channel length, to understand which contribution will be the dominant. And moreover these considered contributions depends on the type of molecule and coupling (linker, electrodes). But above all, it is not sure at all that for e.g. the second effect will leads to an overall increase of the current, that is because out of equilibrium, by increasing the bias window (i.e. increasing the V_{ds}) also the voltage applied will impact deeply in the shape of the transmission spectra, as is evident from NEGF formalism!

In figure 3.23 is shown the resulting drain current I_{DS} for OT molecules as a function of increasing number of rings at $V_{DS} = 1V$. The overall trend is an exponential decrease of the current with the increase of the molecular channel length, meaning that actually the effect (c) is the predominant one.

However, notice that, the trends outlined previously can be different if we consider a different molecule or a different type of electrode. Indeed if for e.g. OPE is consider as molecular channel and the material of the drain electrode is made of platinum (Pt), by repeating the same simulations increasing the number

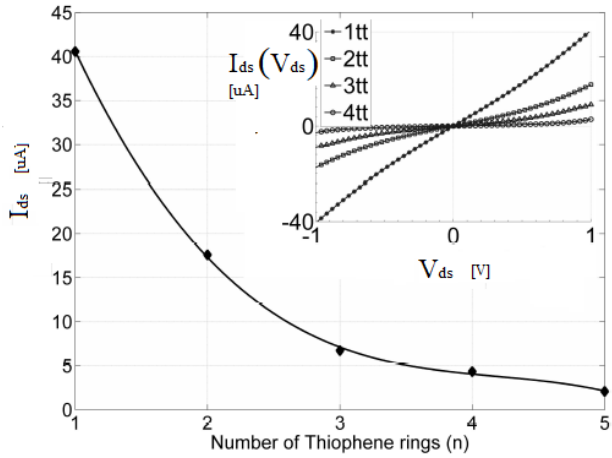


Figure 3.23: The resulting drain current I_{DS} for OT molecules as a function of increasing number of rings at $V_{DS} = 1V$. On the rigth top the IV curve of OT molecules with different lengths in the bias range of [-1V to 1V]

of OPE rings from 1 to 6, it turns out that, even if the transmission spectra have show similar effect, the resultin current trend as a function of number of OPE rings (n) changes w.r.t. the previous case. Indeed, at $V_{DS} = 1V$ the drain current from $n = 1$ to $n = 3$ has a similar trend as before but then increases after $n = 3$. That is because above $n = 3$, the phenomena (a) and (b) become predominant w.r.t. (c). This trend is an advantage from a fabrication standpoint since larger nanogap into which host the molecule, are easier to be created.

Notice that actually in this case, also the coupling factor of the drain γ_2 is changed. This leads to a favourable direction of the current flow from source to drain, and as a result a diode like molecular device is obtained. For details of this application see the work presented in [51].

3.3.4 Dependence on the torsion

In this subsection, how the conductance of the molecular wire is influenced by the torsion of the molecule w.r.t. the electrodes, is analyzed. The molecule exploited in this analysis is the thiophene molecule (2TT). The conduction is simulated for the following degrees of torsion of one thiopene ring w.r.t. other (as shown in figure 3.24): 45°, 75°, 90°, and finally compared. Notice that, typically, torsion is an unwanted feature derived from fabrication process variations. In particular, is a consequence of:

- synthesis of the molecule
- temperature
- anchoring groups
- high bias values

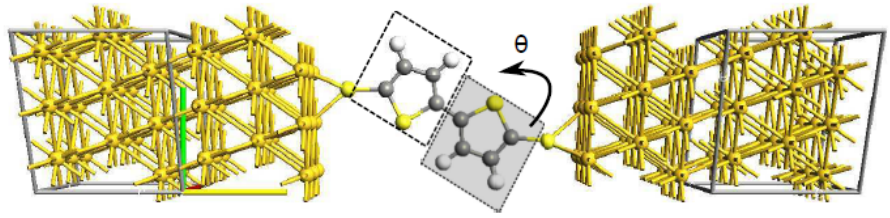


Figure 3.24: Dithiol-dithiophenyl (2TT) molecular wire. The plane of OT ring on the left side is fixed while the shaded plane is rotated with respect to the fixed one of an angle θ.

Again, in order to understand how the conduction in OT molecules changes with inter-ring torsion angles, firstly the simulated equilibrium transmission spectrum and then the resulting current for each θ are analyzed.
From figure 3.25, is it possible to understand how inter-ring torsion angles affects the equilibrium transmission spectrum. By increasing the degrees of torsion, the following trends occurs:

a) substantial increase of the HLG: for 0° case it is almost 1.5 eV, whereas for 90° it is more than 2 eV.

b) substantial decrease of the transmittivity of the peaks. Indeed, for 0° the transmission of the peaks are maximum, while are minimum in case of 90° rotation.

These two phenomena can be explained in the following way. Conjugated molecules like OT derivatives are characterized by a shared structure of electronic cloud thanks to $\pi bonds$. Any change in the inter-ring angle could affect the alignment and so the overlap of the π-orbitals, decreasing the probability of the electron to freely travel from one orbital to another, and thus affects the transport behaviour of the molecular wire.

In figure 3.26 is shown the resulting drain current I_{DS} for 2TT molecule as a function of increasing degree of torsion at $V_{DS} = 1V$. The maximum current is obtained when the inter-ring angle 0° and minimum current is observed at 90°, as

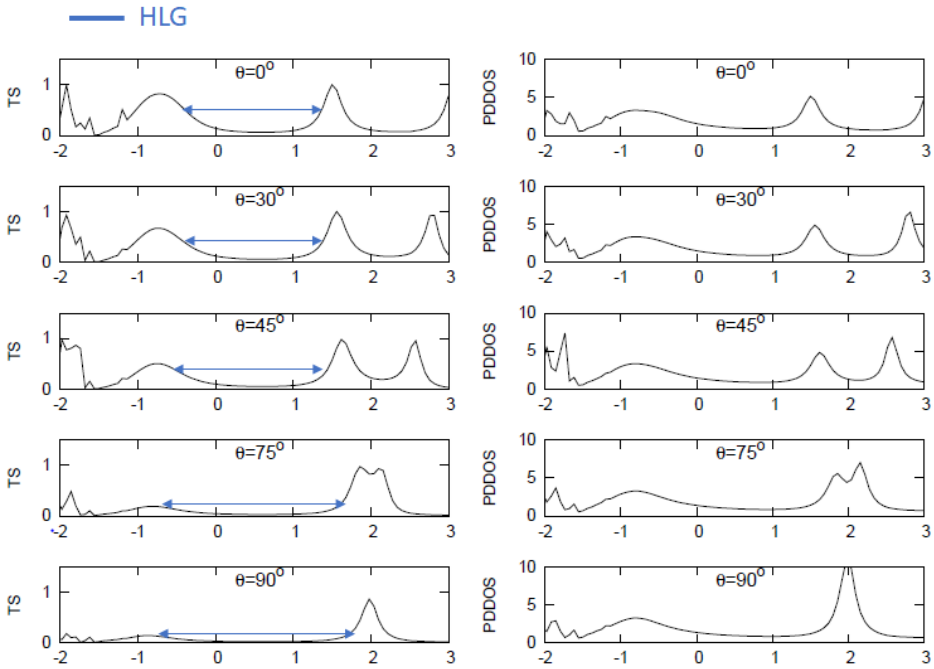


Figure 3.25: Equilibrium Transmission spectra and Projected density of states (PDDOS) of 2TT molecule with different degrees of inter-ring torsion. For all molecular systems the average Fermi-level is placed at 0 eV. The peaks on the left side of the Fermi level are the HOMOs and on the right are the LUMOs. Both transmission spectrum and PDDOS are deeply affected by inter-ring torsion angle. The transmission peaks and PDDOS peaks of HOMOs visibly decrease by increasin the torsion angle.

expected from the analysis of the equilibrium transmission spectra. Notice also that the sulfur atom in the thiophene ring introduces an asymmetry in the structure. Thus, the current in the cases 0° and 180° are not same. For this reason, the current value of the latter case is slightly smaller than the former one. In general, a not negligible reduction of current occurs.

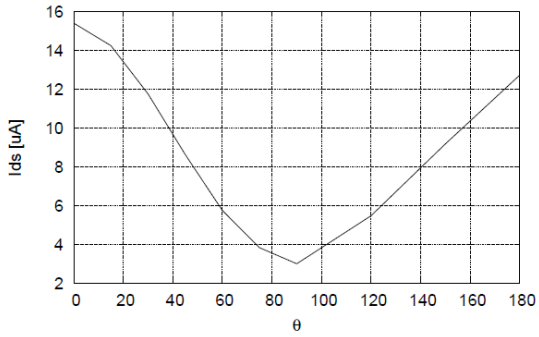


Figure 3.26: The resulting drain current I_{DS} for OT molecules as a function of increasing number of rings at $V_{DS} = 1V$.

3.3.5 Dependence on the type of linker

In this subsection, how the conductance of the molecular wire is influenced by the choice of the anchoring group, is analyzed. A single ring of benzene is considered in this analysis, and terminated with different anchoring groups.
Many researchers have studied this type of dependence [52, 53]. From [52], authors demonstrated that the equilibrium transmission spectra in the case of a benzene terminated with thiol (-SH), cyanide (-CN), oxigen (-O) groups, have the following trends: It is evident how clearly are impacted by the choice of the type of linker.

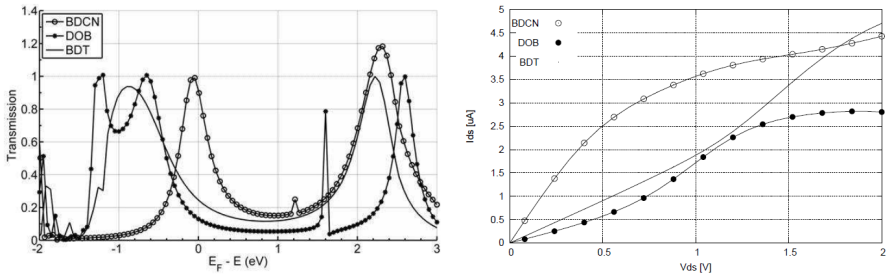


Figure 3.27: The equilibrium transmission spectrum on the left side and the resulting current on the rigth side, in the case of a benzene terminated with thiol (-SH), cyanide (-CN), oxigen (-O) groups.

The resulting current I_{DS} is also shown on the rigth side in figure **??**. It is evident that: for BDO, the current is always the lowest in magnitude independently on the bias window, since the oxygen has the higher electronegativity; for BDT and

BDCN it depends on the bias window. In particular, for low voltages ($V_{DS}<$ 2 V) the current of BDCN is greater, showing a quadratic behaviour and suddendly a saturation trend. Whereas, at high voltage ($V_{DS}>$ 2 V) BDT and BDC currents have similar magnitude, both showing a linear behaviour.
So, in conclusion, a huge difference is present, changing the linkers (also at different bias window), not only on the magnitude but also on the behaviour of the currents. This phenomenon is still not so clear in literature.

Anyway, by considering different elements of the periodic table with very different electronegativity values (Fluorine F with 4, clorine Cl with 3, bromine B with 2.8 and iodine I with 2.5) and by repeating the simulations, an high sensitivity of the conduction on the electronegativity of linkers is shown. Indeed, from the equi-

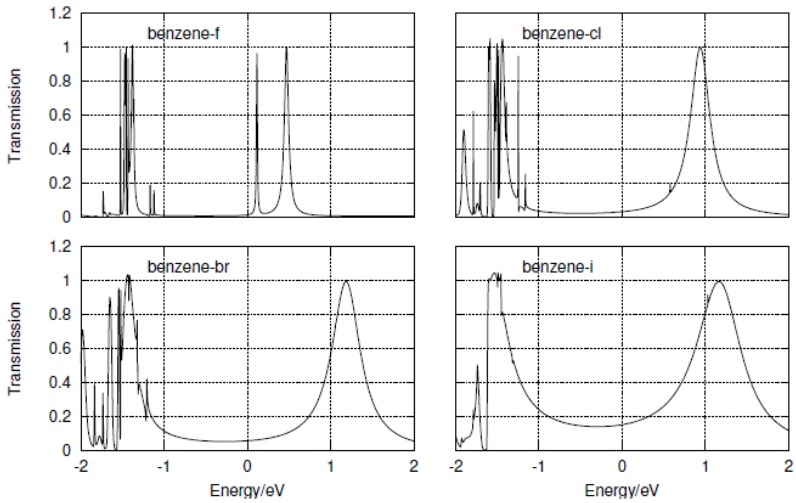


Figure 3.28: The equilibrium transmission spectrum in the case of a benzene terminated with Fluorine F (top left), clorine Cl (top right), bromine Br (bottom left) and iodine I (bottom right).

librium transmission spectra, shown in figure 3.28, is evident that by increasing the electronegativity of the linkers, are present:

- a decrease of the HLG
- a decrease of the broadening of the transmission peaks
- increase of number of peaks

As a result the drain current shown in figure 3.29, increases more than a linear way with the decrease of the electronegativity of the linkers. Indeeed, the highest

current is observed for I linker thanks to the higher transmittivity near Fermi level. This effect can be explained in the following way. First of all, the electronegativity

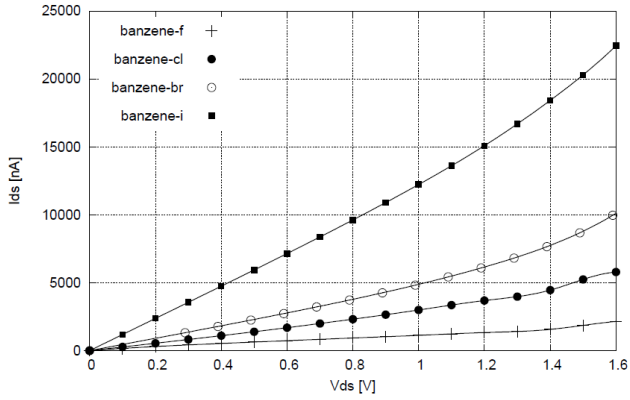


Figure 3.29: The drain current of the molecular wire in the case of a benzene terminated with Fluorine F (crosses line), clorine Cl (full dotted line), bromine Br (empty dotted line) and iodine I (square dotted line).

(E_n) of an atom is a measure of how much it attracts electrons towards itself. The higher is E_n the more attractive is the considered atom. In molecules, the electrons stay for longer time intervals closed to atoms with higher E_n, like trapped, attracted by that atom. Hence, in general a more negative charge is present close to linkers that are more electronegative, since they are traps for electrons. As a consequence, they have also a repulsive force for new electrons coming from the electrodes. Indeed, electrons see an higher barrier in the case of linker with higher E_n, as can be shown in figure 3.30, where the potential energy distribution in 3D-space is shown. Higher potential energy values are labelles with red color, whereas lower with blue color.
In the case of Fl, the potential drops more around the linker, is not well distributed along the system. An high barrier is present between S/D and the molecule, affecting the coupling and as a consequence the broadening of the transmission peaks. Whereas, in the other limit case of I (the element with lowest E_n among the others considered), the potential has a better distribution along the whole system and is less concentrated around the linkers. As a result, lower potential barriers are present, so a better coupling responsible of the broadening of the peaks, is present.
It is evident that these potential energy barriers depends on the type of linkers, especially on their electronegativity. So, the greater is E_n, the higher are the potential barriers between contacts and the molecular channel, and is more difficult

for electrons to tunnel from the electrodes to the molecule, thus the drain current is decreased.

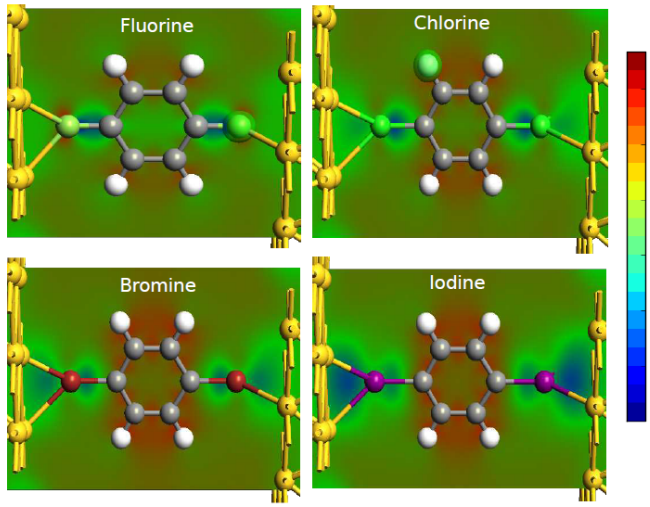


Figure 3.30: The drain current of the molecular wire in the case of a benzene terminated with thiol (-SH), cyanide (-CN), oxigen (-O) groups

Moreover, if an electron from the source electrode goes towards a more electronegative linker, its tendency is to stay more confined in the orbital of the linker, thus without flowing towards the channel and to the drain, thus without giving rise to a current flow.

Chapter 4

Three-terminals molecular device

As already mentioned in the previous chapter, if a two-terminals molecular device is externally perturbed by the introduction of a third terminal, is it possible to speak about a three-terminals molecular device.
The main purpose is to obtain a controlled switch thanks to the transversal electric field applied by means of the additional terminal (gate electrode) as shown in figure 4.1.
Hence, beside the donor and acceptor that work respectively as source and drain, another terminal is added along the length of the molecular channel, in order to obtain an equivalent molecular transistor.

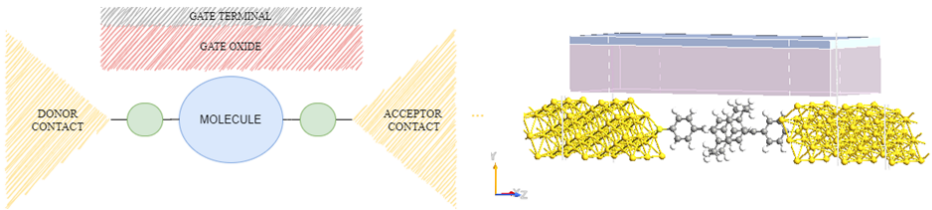


Figure 4.1: A depiction of a molecular FET with a single gate terminal on the right side, and a more realistic representation of the geometry of the device.

The gate terminal has the purpose to modulates the current flowing along the molecule by shifting up and down its energy levels falling within the bias window. Actually, according to the type of gate terminal, different type of transistors can

be obtained. Indeed, the gate can be made of a metal conductor or of another molecule. In the former case it is electrostatically coupled with the channel and a typical Field-Effect transistor (MolFET) is obtained, whereas in the latter is chemically coupled and a sort of chemical-effect transistor is get (MolCET).
Another type of three-terminals molecular device is the one reported in **??**, where an additional atomistic terminal is considered, in order to emulate an intra-molecular logic gate.

However, the focus of this chapter is in particular on molecular field-effect transistors (MolFET). After a very brief review of the state of art, the basic concepts are reported. Then in the next sections a full characterization of two ambipolar MolFETs, engineered ad-hoc for digital computing applications, is widely discussed. The electronic characterization is aided by atomistic simulations performed with QuantumWise ATK software [28]. Two molecules are employed as quantum dot in the MolFET configuration: OPV7 and PCP.
The former has been widely investigated in literature [54, 55, 56], also for MolFET applications [57, 47]. The latter instead, has been only experimentally explored recently [49].

The main contributions of the second part of this work is in the results achieved for both devices. For OPV7-based MolFET an higher $Ion/Ioff$ ratio w.r.t. previous works is obtained thanks to the attention devoted in engineering the device geometry and biasing conditions.
However, an $Ion/Ioff \simeq 50$ is not still enough to effectively employ this device in a digital circuits, especially in static CMOS like logic circuits, as will be demonstrated in the next part of this thesis. Indeed, the typical values of $Ion/Ioff$ ratio in digital circuits ranges around 10^4, value less than one order of magnitude higher w.r.t. the ratio obtained for the PCP-MolFET. A value of around 1400, one order of magnitude higher w.r.t. works found in literature, is achieved by exploiting the Quantum Interference (QI) effect, that only recently has attracted a lot of interest in the molecular electronics field, and by carefully engineering the device. This novel MolFET can be considered a promising candidate in computing applications, as will be demonstrated in sections in the next part of this thesis in sections 6.3.1, 7.3.1,8.4.1.

4.1 State-of-art

A critical challenge for molecular switches is how to effectively achieve modulation control of the charge density in the channel at the single-molecule scale. To this end, in these two decades, many molecular switches with different gating strategies have been demonstrated theoretically and experimentally. Thanks to the fact that molecules in the channel exhibit two or more stable states which can be interchanged by whatever external stimuli, different type of gating stimuli to this

purpose can be used: light [58], mechanical stress[59], heat [60], electrochemical stimuli such as for e.g. Van der Waals forces [61] or more commonly voltage [62, 63, 64, 15, 65, 16]. Voltage-based switches on which this part of the thesis is focused, is the more common type that can be found in molecular electronics literature. Equivalently to the most fundamental component of the complementary metal-oxide-semiconductor (CMOS) technology, the envisioned molecular FET (MolFET) is a three-terminal molecular device in which the drain current flowing through the molecular channel is modulated by the gate voltage. As a molecular counterpart of conventional transistors, molFETs have gained great research interest because of their promising prospect in "Beyond-CMOS" technologies.
The fabrication of such nano-scale three terminal device was experimentally demonstrated for the first time by Song et *al* [15], which also demonstrated experimentally its functionality. Its structure consists on a single molecule anchored by means of suitable anchoring groups to atomistic terminals, typically made of gold, acting as source (S) and drain (D), and electrostatically coupled with a third solid-state metal gate through the gate oxide.
After Song et *al*, even if many other research groups, along the years, have developed different prototypes and investigated several geometries, changing molecules, type of electrodes, adding a back gate etc, the required level of reliability in the fabrication process and the theoretical and experimental confidence is still not reached to envision a future in which MolFET will be able to partially replace the so established CMOS technology.

4.2 Conduction in MolFET: gating effect

The equivalent circuital model of the molecular wire with the addition of the gate terminal electrostatically coupled with the molecular dot is the one shown in the figure 2.24. With this configuration the SCF potential becomes $U_{SCF} = U_{V_{DS}} + U_{V_{GS}} + U_{CE}$, due to the introduction of the gate terminal term:

$$U_{V_{GS}} = -q\frac{C_G}{C_{ES}}V_{GS} \tag{4.1}$$

with $C_{ES} = C_G + C_D + C_S$.

Therefore intuitively, starting from the situation shown in the following figure 4.2 where a generic transmission spectrum of for e.g. the LUMO level has only negligible tails included in the BW, what happens if a gate voltage is applied approximately is the following:

- for a $V_{GS}>0$ the transmission spectrum shifts down in energy, thus the LUMO level contributes a lot in the conduction since a great portion of its $T(E)$ is included in the BW.

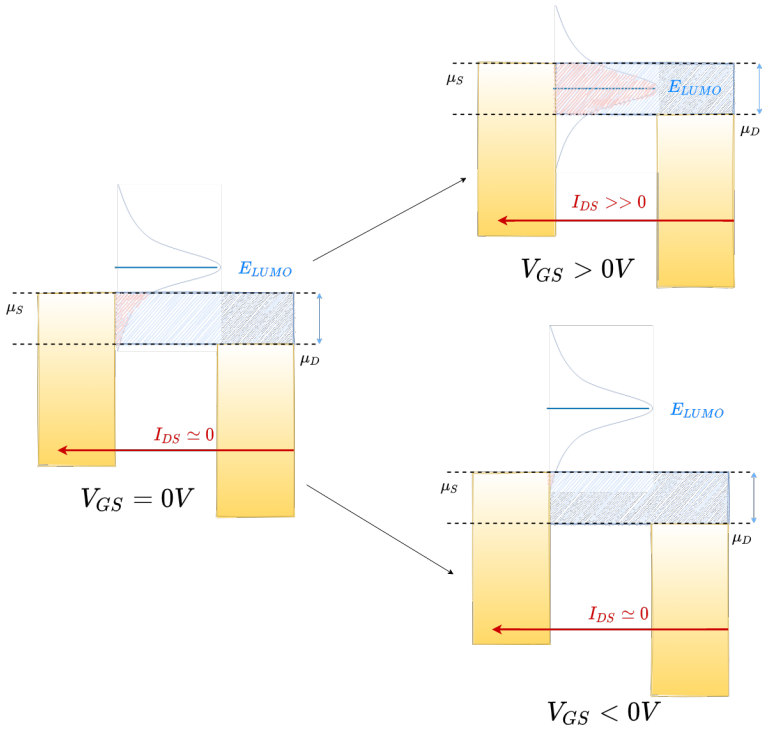


Figure 4.2: Shifting of the transmission spectrum of LUMO level according to the gate voltage applied.

- for a V_{GS}<0 the transmission spectrum shifts up in energy, thus the LUMO level contributes more in a negligible way in the conduction w.r.t. the situation at $V_{GS} = 0V$, beacause a very negligible portion of its $T(E)$ is included in the BW.

The amount of shifting depends on the so called coupling factor of the gate $\alpha = \frac{C_G}{C_{ES}}$, which practically quantifies the electrostatic strength of the gate w.r.t. the channel. The larger is C_G w.r.t. C_{ES}, the more is the electrostatic control of the gate on the modulation of current flowing in the channel, as typically happens for conventional MOSFETs. Therefore the intuitively idea is that: it is possible to modulate the drain current including in the BW the LUMO levels ($V_{GS} > 0$) or the HOMO ones ($V_{GS} < 0$), so obtaining respectively LUMO-type (the analogous of n-type) or HOMO-type (p-type) conduction.

The reason why this phenomenon occurs can be simply explained in a mathematical way. Since the contribution of the gate voltage in the total potential energy is: $U_{V_{GS}} = -q\frac{C_G}{C_{ES}}V_{GS}$, what happens is that:

- for a V_{GS}>0 $\longrightarrow U_{V_{GS}} < 0$ and so $T(E - U_{SCF})$ shifts towards lower energy values
- for a V_{GS}<0 $\longrightarrow U_{V_{GS}} > 0$ and so $T(E - U_{SCF})$ shifts towards higher energy values

4.3 Ambipolar MolFET

The ultimate goal of this thesis is to exploit molecular devices in circuits, in order to understand if some important circuital parameters like functionality, area, power, speed, in a molecular implementation, could be enhanced if compared with MOSFET-based implementation. To this end, since the target circuital application is of static CMOS logic based, a MolFET that works as much as possible like a CMOS devices must be engineered. This requirement is translated in the following specifications:

- ambipolar feature (i.e. p and n-type MolFET with symmetric threshold voltages: $|V_{THn}| = |V_{THp}|$) in such a way to obtain a complementary behaviour of p and n MolFETs. More in particular, in such a way that, for a given fixed value of V_{GS} the n-MolFET is completely ON, and the p-type as much as possible OFF.
- IV characteristics as much as possible similar to MOSFET ones (i.e. resistive, triode, saturation for output characteristics). Indeed molecular devices have the peculiarities also to exhibit oscillations and NDR effects that would impact negatively in the chosen application target.
- reasonable drive on current I_{ON} in order to not impact too much on circuit propagation delay. Indeed the more is the ON current the more responsive will be the device in charging/discharging its load, and so in making the information propagating faster throughout the circuit.
- I_{ON}/I_{OFF} as much as possible close to the minimum requirement for modern digital circuits (10^4) and in order to minimize static power consumption mainly due to leakages.
- $|I_{ONn}| \simeq |I_{ONp}|$ to minimize noise margins and delays of the circuit.

The key element in obtain a molecular device with properties as much as close to these specifications, is the introduction of the back gate, useful to engineer the behaviour of the device, as explained in the following.

4.3.1 Methodology and atomistic simulations

In this subsection the methodology exploited in order to get a functional molecular device with features as much as possible close to the specifications reported above, is discussed. Simulations results are analyzed and commented. It can be mainly outlined with the following successive steps:

- **STEP 1: Choice of the molecules**
  The motivations behind the choice of the molecules exploited to build a molecular transistors are the following. Obviously the key requirement is that the molecules to be chosen must be conductive and so exhibit delocalization of electrons typical of aromatic hydrocarbons. OPV7 was chosen for its popularity in this application, whareas PCP was chosen because, as already mentioned in chapter 4, it exhibit as demonstrated in [16], QI. Moreover was experimentally demonstrated an I_{ON}/I_{OFF} ratio almost equal to 300, that is very promising if compared with molecular transistors already investigated in literature.

- **STEP 2: Building of the geometry of the molecular devices** For both MolFET, molecules were firstly built and geometrically relaxed, using Conjugate gradient algorithm, until all residual forces in each atom are less than 0.05 eV/Å. Then the molecular junction was created, by attaching each molecule to a semi-infinite (3x3) (111) gold electrodes through thiol bonds (-SH). The distance between the sulfur atoms and gold surface was fixed to 1.71 Å which corresponds to S-Au bond length of 2.38 Å, value validated experimentally in [43]. Once the molecular junction was created, two solid-state gate terminals working as front gate and back gate, were properly built, considering also realistic feasibility of the structure. Zirconium dioxide (ZrO2) was employed since it provides a better equivalent oxide thickness ($\epsilon = 25$ with atomic layer thickness equal to 5.7 Å) w.r.t. the common Hafnium dioxide. A single atomic layer of ZrO2, that is physically possible to be fabricated with ALD techniques if the proper precursors exists, was built. The complete geometrical structure for both the molecules (OPV7 and PCP) are shown in the following figures 4.3, 4.4.

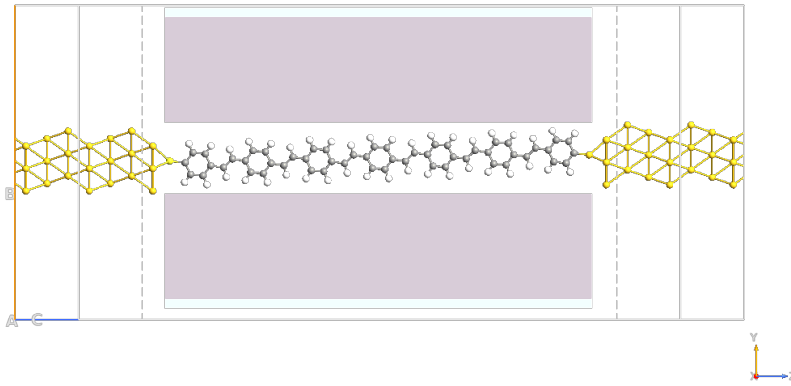


Figure 4.3: OPV7 double gate transistor geometry.

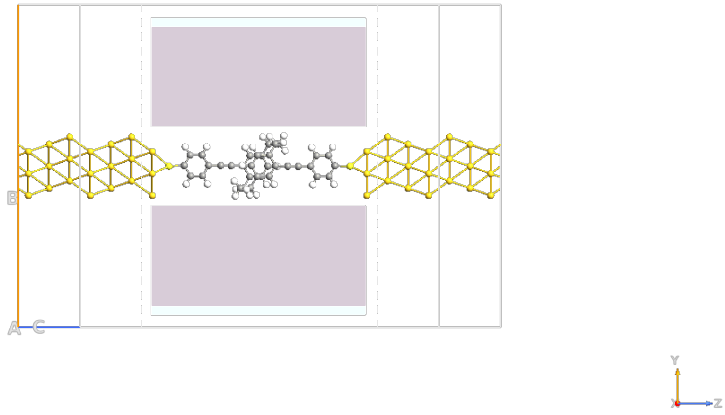


Figure 4.4: PCP double gate transistor geometry.

- **STEP 3: Simulation of the device at equilibrium**
  In order to obtain an ambipolar molecular transistor it is firstly necessary to understand its electronic transport behaviour at equilibrium by means of TS. Therefore, the two devices were simulated at equilibrium condition, i.e. for null bias applied, thus null drain-source, front-gate and back-gate voltages. The set-up of the simulations are detailed in appendix A. In the following figures 4.5, 4.6 are shown the simulation results of the equilibrium TS for OPV7 and PCP MolFET. The Extended Hückel theory (EHT) is used in this first characterization at equilibrium.
  As it is evident OPV7 does not exhibit a well defined HOMO and LUMO

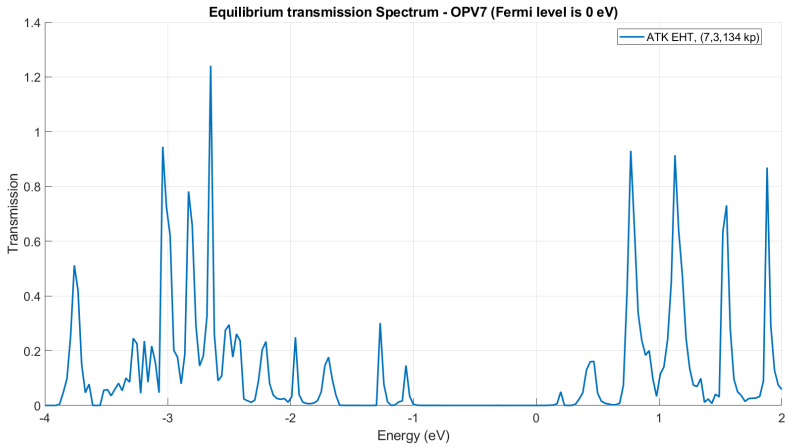


Figure 4.5: Equilibrium TS for OPV7 double gate transistor.

transmission peaks, differently from PCP one. Moreover the peaks are very narrow, this can be likely due to the weak coupling strength because of the length of the molecule (OPV7 case) or to the anchoring or electrodes effectiveness in creating strong chemical bonds.
However, thanks to these first characterization is it possible to roughly have an idea on which range of values can be chosen for the back-gate voltages. Indeed the aim is choose a proper value of V_{BG} that allows at the same time a LUMO (n-type) conduction at $V_{FG} = 1V$ and a HOMO (p-type) conduction at $V_{FG} = -1V$, in such a way to obtain an ambipolar feature.

- **STEP 4: Choice of the ranges of V_{BG} and simulations out-of-equilibrium**
  By looking at the TS results it is possible to have an idea of the sweep range

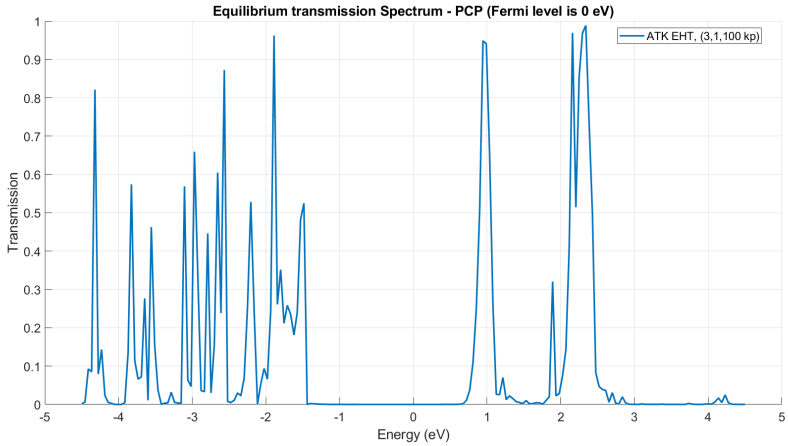


Figure 4.6: Equilibrium TS for PCP double gate transistor.

of the back-gate bias. The simulations out-of equilibrium were performed exploiting NEGF formalism coupled to Extended Hückel theory (EHT). In particular, the electronic characterization at equilibrium resulting from the EHT were fed into the NEGF model to calculate the drain current self-consistently by means of the Landauer formula 2.76. For both cases, the employed boundary conditions in the Conjugate Gradient Poisson solver were fixed to Dirichlet along the transport direction, Neumann in the direction along which the gates terminal are placed and Periodic directions in the remaining one. Details are reported in appendix A. The following bias range for V_{BG} were chosen:

- for n-OPV7 MolFET: sweep in $V_{BG} = [0,3]V$ in order to investigate which is the proper value to get the following conditions: (A.) is it possible to have LUMO conduction (n-type) with a $V_{FG} = +1V$ when a drain-source bias within 0 and 1 V is applied ($V_{DS} = [0,1]V$); (B.) is it possible to have negligible conduction (i.e. negligible portion of TS included in the BW) and so an n-OPV7 MolFET at OFF state, with a $V_{FG} = -1V$ when a drain-source bias within 0 and 1 V is applied ($V_{DS} = [0,1]V$);
- for p-OPV7 MolFET: sweep in $V_{BG} = [-3,0]V$ in order to investigate which is the proper value to get the following conditions: (A.) is it possible to have HOMO conduction (p-type) with a $V_{FG} = -1V$ when a drain-source bias within -1 and 0 V is applied ($V_{DS} = [-1,0]V$); (B.) is it possible to have negligible conduction (i.e. negligible portion of TS

included in the BW) and so an p-OPV7 MolFET at OFF state, with a $V_{FG} = +1V$ when a drain-source bias within -1 and 0 V is applied ($V_{DS} = [-1,0]V$);

- for n-PCP MolFET: sweep in $V_{BG} = [0,3.3]V$ in order to investigate which is the proper value to get the following conditions: (A.) is it possible to have LUMO conduction (n-type) with a $V_{FG} = +1V$ when a drain-source bias within 0 and 1 V is applied ($V_{DS} = [0,1]V$); (B.) is it possible to have negligible conduction (i.e. negligible portion of TS included in the BW) and so an n-OPV7 MolFET at OFF state, with a $V_{FG} = -1V$ when a drain-source bias within 0 and 1 V is applied ($V_{DS} = [0,1]V$);
- for p-PCP MolFET: sweep in $V_{BG} = [-3.3,0]V$ in order to investigate which is the proper value to get the following conditions: (A.) is it possible to have HOMO conduction (p-type) with a $V_{FG} = -1V$ when a drain-source bias within -1 and 0 V is applied ($V_{DS} = [-1,0]V$); (B.) is it possible to have negligible conduction (i.e. negligible portion of TS included in the BW) and so an p-OPV7 MolFET at OFF state, with a $V_{FG} = +1V$ when a drain-source bias within -1 and 0 V is applied ($V_{DS} = [-1,0]V$);

The resulting IV curves of these simulations are reported for sake of completeness in appendix. They are intermediate results useful in order to accurately engineer the behavior of the two ambipolar MolFETs, by properly choosing the biasing conditions, and in particular by choosing the values of back gate voltages in order to have:

- n-type MolFETs ON when $V_{FG} = +1V$ and OFF when $V_{FG} = -1V$;
- p-type MolFETs ON when $V_{FG} = -1V$ and OFF when $V_{FG} = +1V$.

These values were accurately chosen also with the aim to get as much as possible close to the specifications. By accurately inspecting the simulation results, the most suitable vales resulted to be:

- for n-OPV7: $V_{BG} = 1V$; and for p-OPV7 $V_{BG} = -3V$.
- for n-PCP: $V_{BG} = 1.2V$; and for p-OPV7 $V_{BG} = -3.3V$.

- **STEP 5: Devices characterization**
  With the values of back-gate chosen, simulations out-of-equilibrium in order to get output and trans characteristics of the devices, thus in order to characterize them, were performed. In particular they were obtained by sweeping front-gate and drain voltages. The results of these simulations are reported in figures A.8 and comments, for sake of readibility, are directly provided in the captions.

– **n-OPV7: output and trans characteristics**

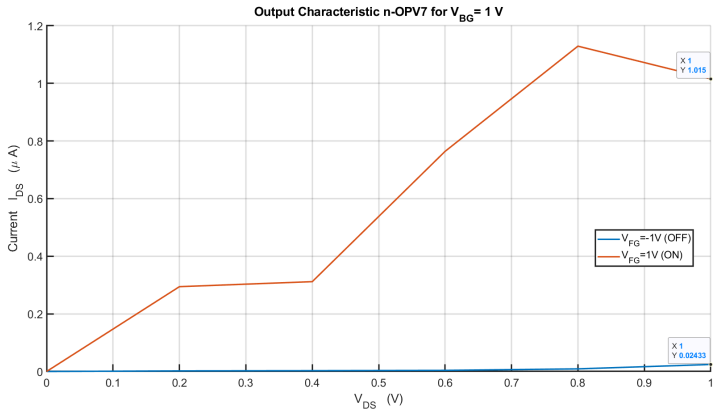


Figure 4.7: Output characteristic of n-OPV7 MolFET. A ratio $Ion/Ioff \simeq 42$ is achieved. There is a little Negative differential resistance (NDR) trend, that poses the question how this trend will impact at circuital level (for e.g. case of CMOS-like inverter in inversion region?).

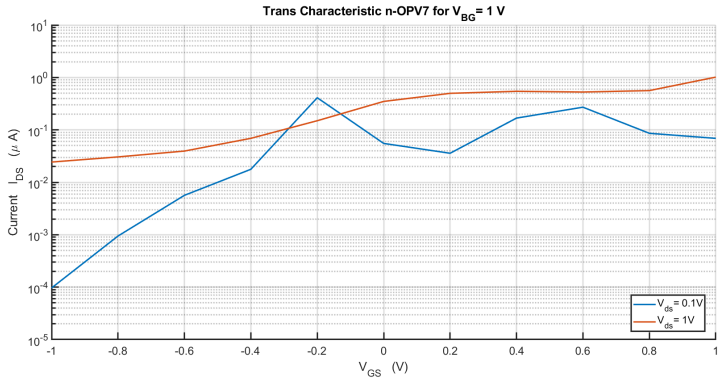


Figure 4.8: Trans-characteristic of n-OPV7 MolFET in log scale. Are evident the deviations wrt conventional MOSFET: there is not a well defined threshold voltage especially for the transchar at $V_{DS} = 0.1V$ and moreover a prominent DIBL is present. The SS that could be defined in a more or less linear portion of bias, is it too low due to the tiny $Ion/Ioff$ ratio. This will obviously impact at architectural level.

– p-OPV7: output and trans characteristics

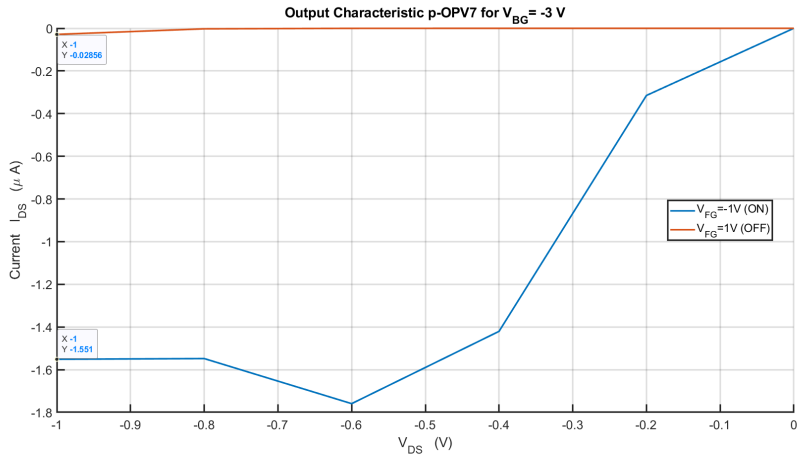


Figure 4.9: Output characteristic of p-OPV7 MolFET. A ratio $Ion/Ioff \simeq 54$ is achieved. Also here is present a little NDR trend.

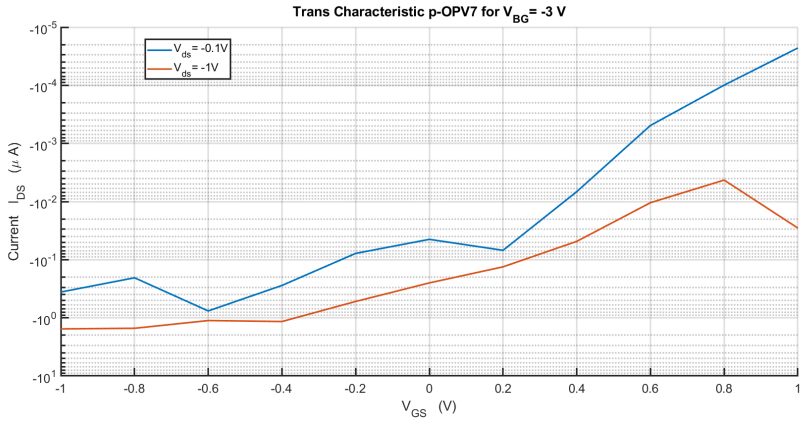


Figure 4.10: Trans-characteristic of p-OPV7 MolFET. Same comments done for n-opv7 transchar.

– **n-PCP: output and trans characteristics**

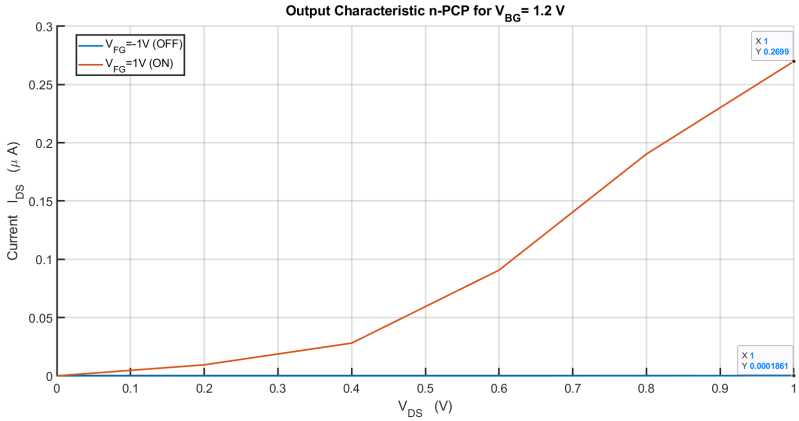


Figure 4.11: Output characteristic of n-PCP MolFET. A ratio $Ion/Ioff \simeq 1452$ is achieved. Here is not present a NDR trend, and the characteristic is almost linear in the range analyzed.

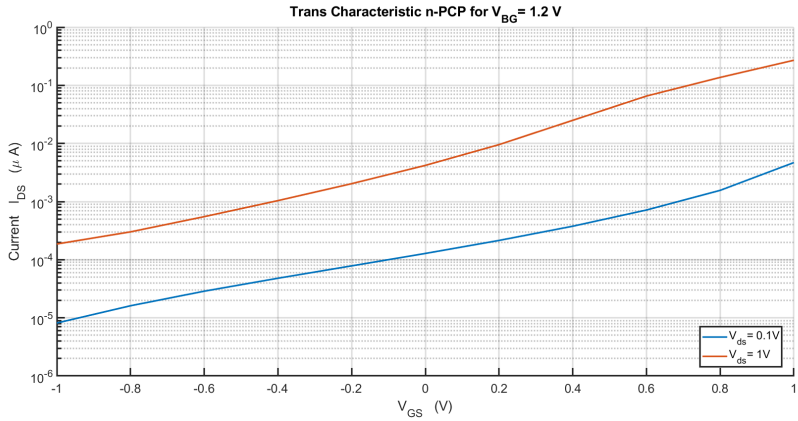


Figure 4.12: Trans-characteristic of n-PCP MolFET. Here the deviations wrt conventional MOSFET are very evident: the DIBL at 1V is almost of the order of 10^2. On the other hand, in this range is it possible to better define the threshold voltages (that are evidently subjected to roll off) and the SS that could be defined in a more or less linear portion of bias.

– **p-PCP: output and trans characteristics**

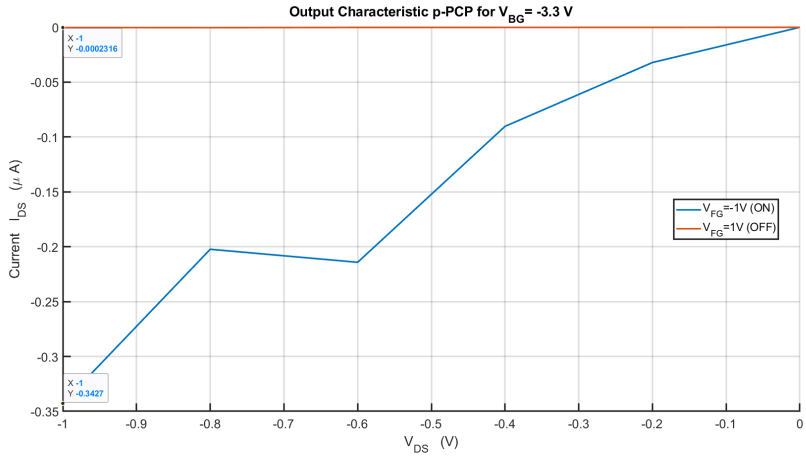


Figure 4.13: Output characteristic of p-PCP MolFET. A ratio $Ion/Ioff \simeq 1478$ is achieved. Here there is a little a NDR trend, and no saturation in the range analyzed.

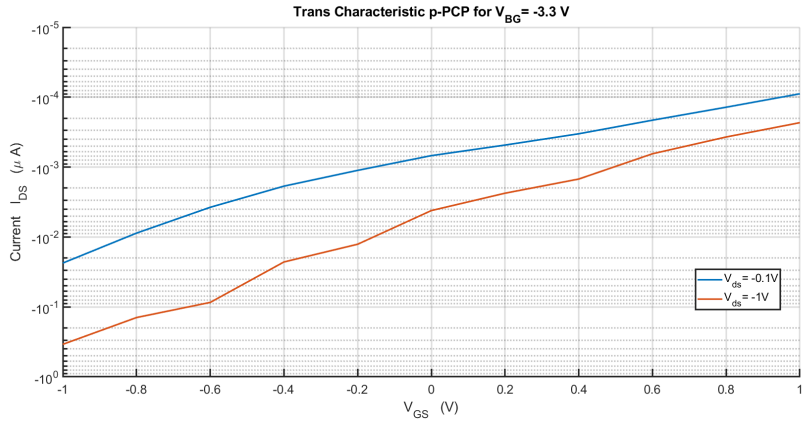


Figure 4.14: Trans-characteristic of p-PCP MolFET. Same comments done for n-PCP transchar.

Due to the fact that all the MolFETs simulated exhibit more or less bad DIBL and SS and roll-off issues, they are not robust transistors like the

established MOSFET and at circuital level (for a static CMOS logic based circuit) are expected malfunctioning due to these properties. However better functioning and performances are expected for PCP MolFET-based circuit thanks to its high $Ion/Ioff$ if compared wrt other molecular transistors already investigated in literature. Anyway is far to distant from the minimum requirement of 10^4 for behaving as a reliable transistor.

Part III

Architectural level

One of the main reason why molecular electronics is still at its infancy level is that the majority of the researchers in this field, expert in different disciplines (chemistry, biology, physics of condensed matter, device physics, process/material engineering, electronic engineering..) still does not effectively interface between them. Typically what happens is that for e.g. the physicist or the chemist does not make its research work readable for an electronic engineer and in turn also the electronic engineer does the same for a biologist or a chemist. Due to the very deep nature of this field, a strong interdisciplinary approach, enriched of constructive feedback, must be adopted in order to make true practical progress.
Therefore, from my electronic engineer perspective, I believe that when studying emerging technologies the research cannot stop at the device level. Indeed to really be aware of the strengths and limitations of a new device it is also necessary to assemble them together in the design of complex and realistic circuits. Only thanks to this further step, it is possible to understand the effectiveness of a new technology, that is simply because parameters like power dissipation and speed can be evaluated accurately only at the system level.

This methodology is also the "mantra" of VLSI lab, in collaboration with which this work thesis is done.

With this perspective in mind, the aim of the third part of this thesis is to exploit the work at physical and device levels, discussed in the two previous parts, in a circuit, more or less complex.

In the first chapter 5 some possible methodologies that could be adopted in order to assemble more than one molecular device in a more complex circuit, are discussed.

Then in the next chapters 6, 7 8, the methodology based on Look-Up-Table is exploited for the molecular implementation of logic (logic gates, Half-Adder, Full-Adder) and memory (SRAM) elements. Finally a target Logic-in-Memory application is discussed and implemented. Design, functional verifications, performances analysis are performed in Cadence Virtuoso® both for molecular ambipolar transistors and FD-SOI MOSFET (28nm) technologies in order to be compared and demonstrate the benefits and problems, for this specific application, in migrating from conventional to a molecular technology.

Chapter 5

Circuital modeling of MolFET

The aim of this chapter is to briefly discuss the problem related in jumping from the device abstraction level to the circuital one, where more than one characterized device is integrated and must be simulated in an effective way, and outline different methodological approaches that can be adopted in performing this jump. Finally in section 5.2, the approach followed in this thesis is widely explained.

5.1 From device to circuits

In general it is possible to engineer a single transistors with formidable properties, but if it is not assembled with many other transistors in the design of whatever circuit, more or less complex, that research effort becomes nullified from a practical standpoint. Indeed only thanks to this further step, it is possible to understand the effectiveness of a transistor or more in general of an electronic device, in a specific application. That is because, some important parameters as power dissipation or speed can be evaluated accurately only at the system level. Therefore the typical guideline to adopt in modern digital circuit design is made of the following steps:

1. a logic family with which implement the digital circuit is chosen and typical device requirements are outlined for that specific choice.
2. the device is optimised for that application in order to satisfy the circuit-level requirements.

3. devices are assembled in the circuit and parameters as power consumption, speed, area are evaluated by simulating the overall circuit.
4. if the power budget, the occupied area or the system delay exceed specifications at system-level, then it is possible go back to the previous steps and work at device or also work at circuital/architectural level providing a feedback with tighter specifications to the technologist/device or electronic engineer and then repeat the procedure in a positive loop until specifications are satisfied.

This approach is essential in circuit design especially nowadays in the low-power era. This approach is also the one followed in this thesis in the design of logic gates, memory cell and at last of the Logic-in-Memory (LiM) cell discussed in the next chapters (6, 7, 8): (1) the static CMOS logic family was chosen to implement those circuits and requirements at device levels were outlined, in particular:

- ambipolar feature (i.e. p and n-type FET with symmetric threshold voltages: $|V_{THn}| = |V_{THp}|$)
- reasonable drive on current I_{ON} in order to not impact too much on circuit propagation delay
- I_{ON}/I_{OFF} as much as possible close to the minimum requirement for modern digital circuits (10^4) and in order to minimize static power consumption mainly due to leakages
- $|I_{ONn}| \simeq |I_{ONp}|$ to minimize noise margins and delays

The more the device deviates from those requirements, the less the circuit in which is exploited will work as expected. And the more complex is the circuit, the more these deviations will impact on its functioning, as practically demonstrated in 6.3.1, 7.3.1.

However, the problem in this approach, especially in nowadays IC design, is the difficulty to get a quite accurate parameters estimation but at design time, before the synthesis phase, in order to make possible almost at runtime to explore the space of different technological/circuital/architectural solutions and how they are reflected in terms of required energy vs. delay. It is difficult or better almost impracticable because getting such parameters estimations means simulate the overall system in question, but present ICs counts billion of transistors to be simulated. As a consequence at the end of the endless simulation, the ICs in question will be already very likely out of market! From here the necessity to get at the same time an effective model of the active device in terms of accuracy and computational time required.

Concerning CMOS devices, compact models (i.e. models that are sufficiently simple to be incorporated in circuit simulators, like for e.g. SPICE, and are sufficiently accurate to make the outcome of the simulators useful to circuit designers) assisted by statistic analysis based on Monte Carlo method make the fortune of modern ICs design systems CAD tools.

Differently, emerging technologies and in particular molecular transistors lack of such accurate and computationally efficient models, even if, for almost five decades by now, molecular electronics gained a great interest from both theoretical and applied electronics point of view. Indeed, a lot of work, mainly based on first-principle or semi-empirical methods, has been carried out in order to understand the physics of molecular devices. However, less effort has been devoted for the circuital modeling of molFET-based system. This lack limits, as a consequence, the analysis of circuits involving the integration of many molecular transistors, thus limiting yet the progress of molecular technology in the computing field. As happened for semiconductors, the availability of solid-circuit modeling can open the way to molecular systems design tools and, therefore, to application. Indeed if, as already said, for a ICs based on standard CMOS technology is almost impracticable simulate the overall system, for MolFET-based circuit, even the simplest one, for e.g. a digital inverter, it will require roughly up to two days in order to get an IV characteristic obtained with reasonable enough bias points! This is not at all acceptable, and smart methodologies and both accurate and computationally efficient model are needed to calculate the electron transport in molecular transistors integrated in a circuit.

Mainly three different approaches are nowadays adopted at research level:

a. Self-Consistent-Field (SCF) modeling based on simple set of equations describing the device in its interaction with the others. These set of equations, dependent one from the others, are iterative solved in a self-consistent loop until a certain desired tolerance is reached. This SCF algorithm, depending on its complexity, can be integrated with hardware descriptive language to make viable circuit simulations and analysis. This approach has the advantage of being able to capture the quantum physics of the device in the system with an accuracy depending on the level of approximations made in the set of equations. The toll to pay is in terms of complexity and computational burden. An example of this approach is presented in [47], where the simulation time is reduced six order of magnitude w.r.t. ab-initio simulations while maintaining the accuracy in a limited operating bias range.

b. Electrical equivalent modeling based on FET SPICE model. This approach roughly models the physical effects that occurs in such devices,

but has the advantage to be less demanding in terms of computational time. An example is presented in [66], where the electrons transport along a molecule is represented with an equivalent circuit containing FET transistors and dependent generators.

c. LUT-based modeling which consists on collect once for all in a tabular structure the output ($I_{DS}(V_{DS})$) and trans characteristics ($I_{DS}(V_{GS})$) of the molecular transistors obtained from ab-initio simulations. The set of gate and drain bias and the drain current values are input to the circuit simulator which interpolate the data. This approach, slightly enriched with intrinsic device parameters, is adopted in this thesis thanks to its extreme simplicity. It is very easy to be implemented, but, as disadvantage, it totally looses any link with the physics. An example of this approach is exploited also in [67].

5.2 Enriched LUT-based model

The approach exploited in this thesis for the design and analysis of molFET-based circuits is a LUT-based one, enriched with intrinsic device parameters. More in particular, the I-V characteristics obtained in QuantumATK (reported and commented in II), together with a suitably modified MOSFET equivalent circuital model are used to create the molFET model.
But before deepen this topic, a reason why exploiting a modified MOSFET equivalent circuital model is reported for sake of clarity.

Since the mathematical and the circuital models of FETs are quite similar in all types of FETs, they are assumed to be same, at least at first approximation, also for molFETs. Indeed the FETs standard approach to the device analysis is based on the gradual channel approximation (GCA). This assumption states that the applied voltages (V_{GS} and V_{DS}) vary gradually along the channel, giving rise to the so-called channel potential. In practical terms this means to have a distributed resistance in the channel, across which the channel potential drops. At the same time, the applied voltages can vary quickly in the direction perpendicular to the channel. Under GCA it is possible to separate the problem in two distinct parts: the longitudinal one that give rise to the drain current by drifting of carriers under the applied V_{DS}, and the transverse one that controls the amount of free charges within the channel by acting on the V_{GS}. The formal expression for the drain current is always the same for all FETs. It is determined by the integral of the total charge present in the channel (φ_{CH} is the channel potential):

$$I_{DS} = \frac{W\mu_n}{L_{CH}} \int_{\varphi_{CH}(source)=0}^{\varphi_{CH}(drain)=V_{DS}} Q_{CH}(\varphi_{CH}) d\varphi_{CH} \tag{5.1}$$

What changes, depending on the specific kind of considered FET or technology, is the explicit expression for the channel charge Q_{CH} as function of the applied gate voltage and the channel potential, what that is called *charge control law*.
Once is known for a specific FET, the drain current is derived by means of eq. (5.1). Notice that if Q_{CH} is linearized w.r.t. channel potential, the integral in eq. (5.1) would always lead exactly to the same explicit expression for I_{DS}:

$$\begin{aligned} I_{DS} &\approx \frac{W\mu_n}{L_{CH}} \int_0^{V_{DS}} C_{CH} \cdot (V_{GS} - V_{TH} - \varphi_{CH}) d\varphi_{CH} \\ &= \frac{W\mu_n C_{CH}}{L_{CH}} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \end{aligned} \tag{5.2}$$

with C_{CH} the channel capacitance.

So, in this sense each FET (e.g. MOSFET, MESFET, JFET, HEMT, etc...) is described by a similar mathematical model. Moreover starting from mathematical equations, an equivalent (large or small signal) circuital model is usually found and used in analog or mixed signal circuit design and analysis. As a consequence all FET equivalent circuit models are quite similar. Of course differences are present, and especially values for stray capacitances, resistances are different depending on the specific technology and device.

In this context it is possible to think to the molFET device as a general FET, and use as approximate model for design and analysis of circuits an equivalent circuital model somehow similar from the MOSFET one. This is of course an approximation, and it must be treated as it is. Indeed there is of course a fundamental difference in the transport modeling approach: with molFET we are in a purely quantum regime, and semi-classical models (usually exploited in convential FETs) are out of their regime of validity and cannot be applied. Moreover scattering and incoherent transport are neglected, considering the most probable mechanism of transport through a molecule to be coherent tunneling [30]. Anyway this model can be considered as a good starting point, especially if the following points are highlighted. First, if it is possible to provide a linearized expression for the number of electron states that allow for having coherent tunneling through the molecule, the case is similar to the linearized charge control law of more conventional FETs, thus a similar behaviour for the (generally large signal) $i_{DS}(v_{GS}, v_{DS})$ is expected. This correspond to a first order Taylor's approximation, i.e. to have an enough small variation of available states for conduction in function of a voltage variation, such that a linear approximation is not so bad. This could be a good approximation in the case of a molecule strongly coupled

to contacts, in which the discrete energy levels of the molecule are so widely broadened in energy that a smooth and almost constant density of states is obtained, leading to a gradual transition of states into the bias window. Instead a weakly coupled molecule would generally lead to an abrupt variation of number of conducting states as the bias window is enlarged. This linearized approach is essentially the one widely used in [30] (where only small signal models are provided) and also presented (in a slight different manner) in [29].

Second, there are strong similarities in conventional standard MOSFET device structure and process technology w.r.t. many basic molFET devices and processes as shown in chapter 4. The molFET structure is essentially the same of the conventional MOSFET, with metal drain, source and gate contacts, and with a gate oxide that isolates the gate contact from the channel, allowing (theoretically) only for electrostatic control of the channel charge Q_{CH}. From these similarities, in a molFET we expect to have a gate capacitance C_g, a very large gate resistance R_g (ideally an open circuit), and also electrostatic capacitive effects at source and drain: C_s and C_d (that represent the amount of charge moved toward/from the channel when applied voltages are varied).

The main MOSFET vs molFET difference is the channel, that in the case of the molFET is not a piece of silicon but instead a single molecule, chemically bond to source and drain contacts and positioned («suspended») above the gate insulator (or oxide). As a result it is expected to have no access resistances in molFET, thus having only source and drain contact resistance. In summary, the expected molFET capacitance contributions are: C_s, C_d and C_g (quite analogous to conventional MOSFETs), while the only expected resistance contribution is R_g, since, no source or drain resistances are expected to be found.
In conclusion the circuit model derived from the MOSFET one, with some adjustments and changes can be considered a good starting point for molFET based circuit design and analysis, and for this reason is exploited in this thesis. Modifications are explained and used for performances analysis in the following. Indeed, the static molFET I-V characteristics, collected in the LUTs, are used to find the drain current corresponding to an input couple (V_{GS}, V_{DS}), and to estimate the static power consumption (simply estimated by means of the product $I_{DS} \cdot V_{DS}$). Instead, in order to estimate the transient behaviour and duration, and the dynamic power, the gate, source and drain capacitances and the gate resistance should be suitably exploited.

In the following the employed molFET model to be used in circuit behavioural and performance analysis, is derived and reported. The problem is splitted into two parts and two different modeling approaches will be used, depending on the aim of the performed circuital analysis.

5.2.1 Functional and static analysis

In order to evaluate the functionality and the static power performances of a molFET circuit the only information to be known is the current-voltage characteristics of the considered molFET. In particular, starting from previously simulated trans-characteristics and output characteristics, it is possible to obtain a sort of Look-Up-Table (LUT) in which the current values for a given input couple (V_{GS}, V_{DS}) are stored. Once the output current I_{DS} is known for a set of (V_{GS}, V_{DS}) values, it can be interpolated (in more or less refined way) to get a realistic quantitative previsions of the behaviour of the the considered molFET device, placed within a given circuit under given operating conditions (i.e. specified V_{GS} and V_{DS} values).

This is of course true at steady state, when all dynamic phenomena are extinguished, and the circuit DC operating point is stable. However, this model can also be used in performing some kinds of dynamic analyses. In particular if we are not interested in dynamic power nor in delay performances of the network, but we would just like to quantitatively estimate the response of the circuit to a time-varying stimulus, this LUT-based simple model is effective as well. Indeed it is able to provide a good estimation of the output quantity values for a whatever time-varying input quantity. The obtained results are reliable if time-varying phenomena are slow enough to address a quasi-static regime of operation, in which all transients extinguish before a new variation of input voltages occur.

Two main limitations are present in this approach: first, there is no delay in the response of the circuit (the simulator will instantaneously update the output when the input is changed); second, the outputs shapes are ideal, in the sense that no stray capacitive or other stray effects are taken into consideration with this model. As a result the response will neglect all waveform variations due to these effects.

In summary the overall methodological flow of this approach is depicted in figure 5.1.

In practice, starting from a given geometry for molecular transistors and a given bias range, a IV report containing the IV-characteristics of the device, is generated by the ab-initio simulator *QuantumWise ATK*®. From ab-initio simulation results a LUT is created and recalled by the *VerilogA* file which models the molecular transistor. A Cadence schematic of a given circuit implemented with molFET technology is created. The schematic is simulated with given stimuli by means of *Spectre* simulator embedded in Cadence Virtuoso environment, and output waveform and data are generated. Thanks to this flow it is possible to characterize the given circuit in terms of functional behaviour and dissipated static power.

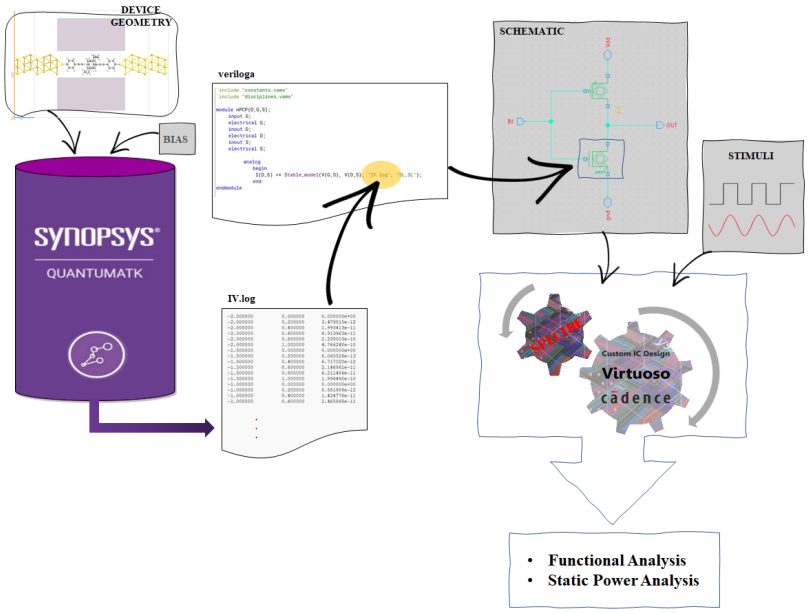


Figure 5.1: Methodological flow adopted for functional and static analysis of molFET-based circuit.

More in detail, the link between the instance of molFET in the circuit and its model was performed directly in Cadence Virtuoso as follows. For every considered molFET type (i.e. for every considered molecule: OPV7, PCP) a new SYMBOL was created. This means creating a new CELLVIEW of type symbol. The symbols used are reported in figure 5.2. Then another CELLVIEW of type *verilogA* (with the same name) for each molFET was created, in which a *verilogA* description of the device was provided. It consisted in an interpolation (3^{rd} order spline polynomial functions) of the LUT, that was built starting from simulated output and trans-characteristics. More details and the *verilogA* code are provided in appendix, section B.

Before design, test and analyze any molFET-based circuit, the different molecular transistor were tested in *Spectre*. In figures 5.3, 5.4, 5.5, 5.6, the output characteristics of PCP, OPV7 molecules that will be exploited in the following chapters. A parametric simulation was performed by sweeping V_{DS} in order to obtain an output characteristic and with V_{GS} as parameter ranging from -1V to 1V. They are the interpolated version of the output characteristics provided by ab-initio simulations, discussed in 4.

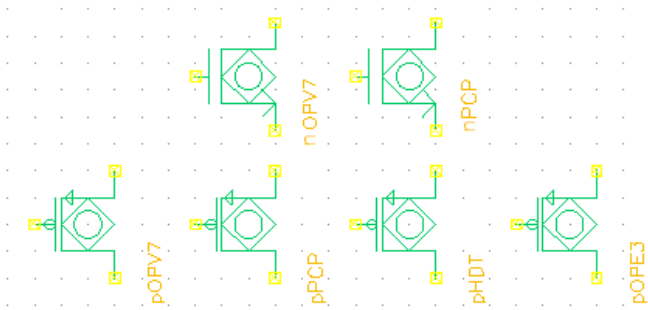


Figure 5.2: Circuit symbols for molFETs: on the top the nOPV7 and nPCP molFETs; on the bottom pOPV7, pPCP, pHDT, pOPE3 molFETs in order from left to right.

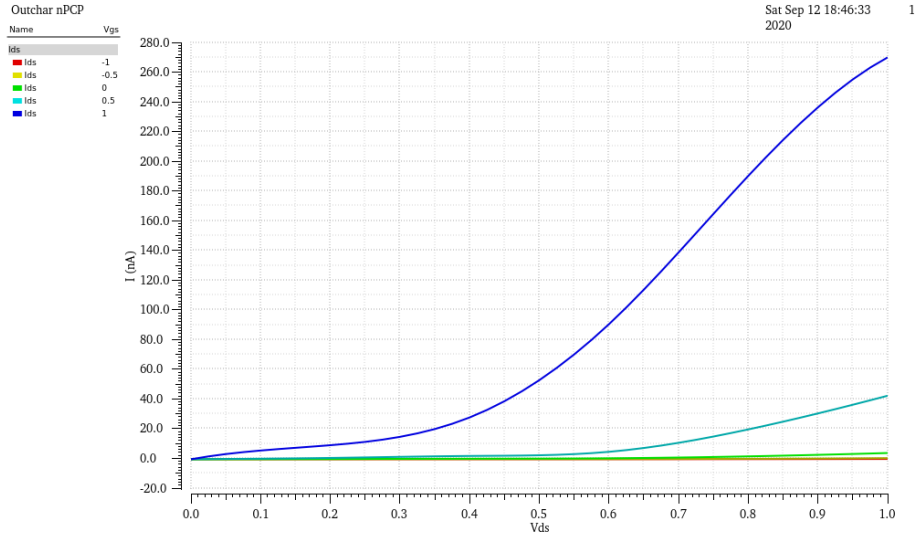


Figure 5.3: Output characteristic of nPCP-molFET.

In conclusion the simple LUT-based modeling approach can be used, in an effective way, for three main purposes:

- DC operating point or static analyses: steady state analyses require the knowledge of the $I_{DS}(V_{GS}, V_{DS})$ only.
- Static power analyses: the static power for single device can be simply

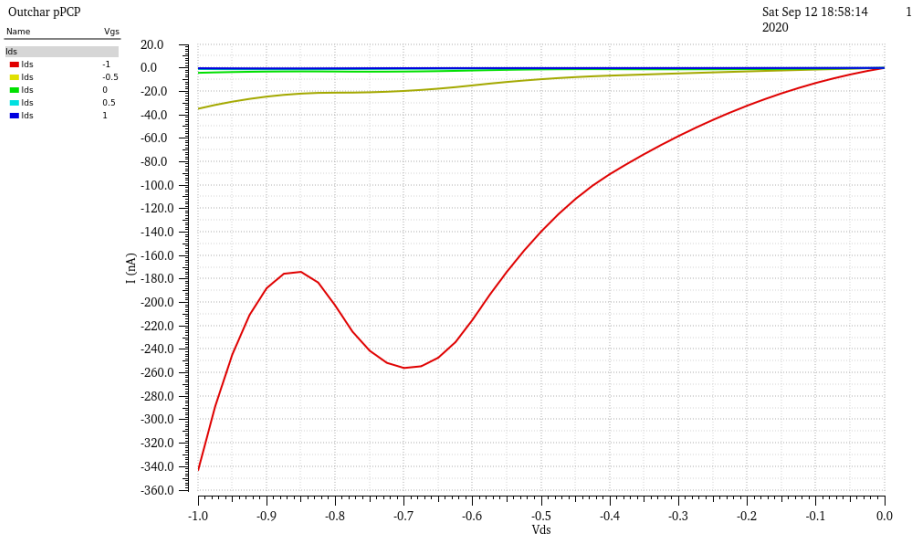


Figure 5.4: Output characteristic of pPCP-molFET.

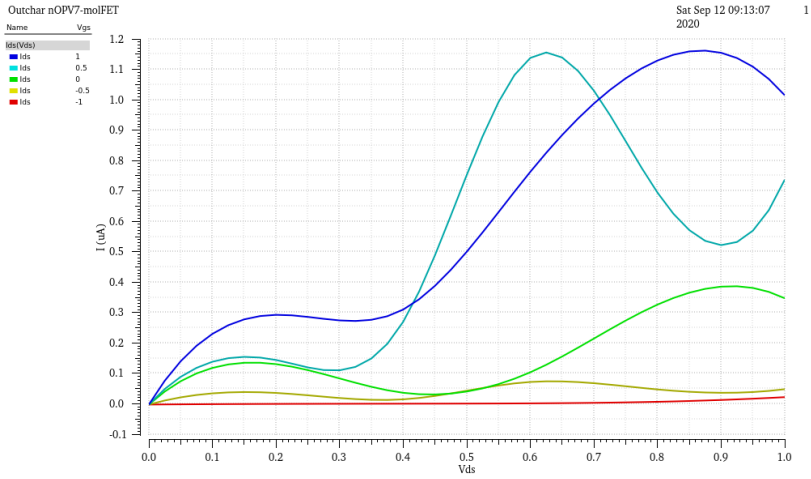


Figure 5.5: Output characteristic of nOPV7-molFET.

estimated as $I_{DS} \cdot V_{DS}$ (gate current is assumed null).

– Dynamic functional analyses (quasi-static regime): dynamic analyses can be performed for slow variations of the inputs.

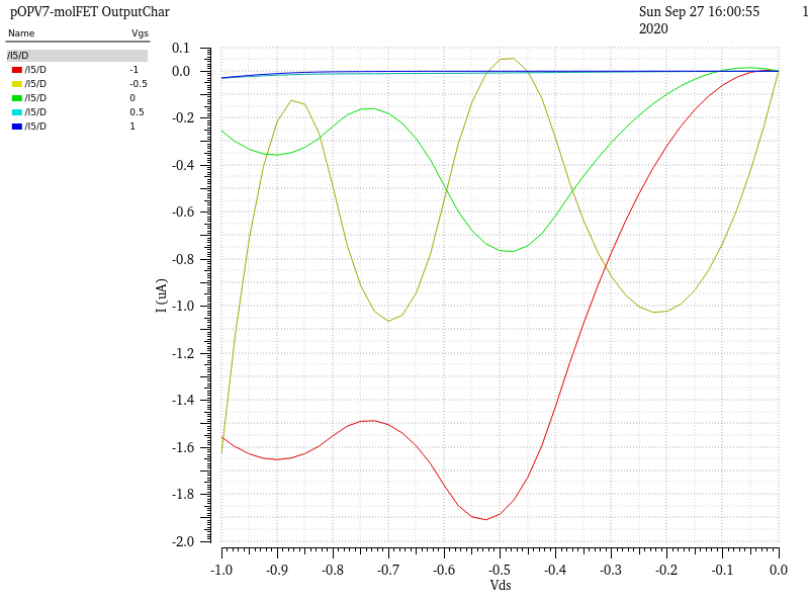


Figure 5.6: Output characteristic of pOPV7-molFET.

The above reported analyses were performed in different circuits in the following chapters. Anyway, in addition to them it is also important to perform dynamic power and transient duration analyses. In order to do that the LUT-based model needs to be improved as described in the following.

5.2.2 Dynamic performance analysis

In order to perform dynamic power analyses it is necessary to estimate the electrostatic capacitances of the molFET device and also the so called quantum capacitance, related to molecule state filling. Indeed the dynamic power is essentially the amount of energy (per unit time) that is dissipated for charging/discharging those capacitances, when a commutation occurs. Moreover, in order to estimate the time duration of transient phenomena related to the maximum operating frequency of the network, the gate resistance R_g should be estimated as well. Indeed the time constants, at least within the rough single-pole approximations, are essentially proportional to the RC products. In addition to R_g it is also necessary to introduce two additional resistances: the source dynamic resistance R_s, and the drain dynamic resistance R_d. These two resistances do not model access resistances of the device, that are

not present in a molFET. They represent the average amount of charge that, due to finite value of escaping rates, can be exchanged between the molecular channel and the source and drain contacts in the unit time. They are useful in modeling the dynamics of the molFET, and they are identical null in steady state case. Notice that source and drain resistances themselves are of course already included in the $I_{DS}(V_{DS})$ slope, and statically the channel conductance is known for each V_{DS} value as the tangent of the $I_{DS}(V_{DS})$ curve in that point. But here the R_s and R_d values are not linked to the slope of $I_{DS}(V_{DS})$ curve; their physical meaning is not the one of «conventional» resistances. They are just resistive contributions, to be considered in correctly modeling of source and drain electrostatic capacitances charging and discharging phenomena. In particular they are just a way of modeling the amount of charge that (on average) can be exchanged between the S/D contacts and the molecule in order to charge/discharge C_s and C_d. They are thus useful in transient duration analysis only.

In the following how to compute the different capacitive and resistive quantities above mentioned is reported. All the calculations referred in this section were performed in *MatLab*, and the employed scripts are reported in appendix, section B.1 (further details on methods and algorithms are explained directly in comments within the codes).

- Gate capacitance C_g:

The molFET gate capacitance arises from the device structure, and it is a physical electrostatic capacitance. It is due to the fact that a gate oxide is in between the metal gate electrode and the channel. In this work the interest is not on get extremely accurate estimation of the capacitances, but on their order of magnitude, therefore the gate capacitance is evaluated simply exploiting the parallel plate capacitor approximation:

$$Cg = \frac{\epsilon_{ox}\epsilon_0}{t_{ox}} \cdot Area \tag{5.3}$$

where $\epsilon_{ox}\epsilon_0$ is the gate oxide dielectric constant, t_{ox} the gate oxide thickness, and the gate contact area is evaluated as the molecule length in longitudinal direction L, multiplied by the total width W_{tot}.

A final remark on the importance of the gate capacitance is now highlighted. In every FET the gate capacitance expresses the proportionality factor between the applied gate voltage and the amount of charge within the channel. It is the physical quantity that above all determines the channel electrostatic control by means of the gate voltage. The greater is C_g the better is the channel electrostatic control. The relative weight of the gate capacitance in the whole capacitive balance of the device is of fundamental importance.

What counts is indeed the C_g value compared to C_s and C_d ones. A «good» FET needs an high C_g w.r.t. C_s, C_d to ensure good channel control by the gate voltage. This holds also in the case of the molFET, even if it is often difficult, due to the small dimensions of the molecular channel (especially if compared to conctact dimensions), to have a molFET with $C_g >> C_s, C_d$.

- Quantum capacitance C_q:

The quantum capacitance represents the molecule states filling. In the moment in which the molecule-contact junction is created, a re-distribution of charges occur between them and the thermodynamic equilibrium is achieved, with an alignment in the Fermi levels. A good quality contact should have an enormous amount of states around the Fermi level, if compared to the channel ones, and thus it should be an unlimited reservoir of electron states for the channel. Under this approximation the contact Fermi level is essentially pinned. Instead the molecule Fermi level is moved up or down in order to achieve the equilibrium. A small amount of charge can be already enough to produce a huge molecular Fermi level shift. Notice that even a fractional electron charge can be exchanged between the molecule and the contacts. Indeed if an electron is involved in a chemical bond between the contact and the molecule (molecular anchoring group), the electron cloud is partially in the contact and partially in the molecule, thus the electron is for a certain time localized inside the contact and for another time interval inside the molecule. On average, therefore, a charge less than q (electron elementary charge) can be localized inside the molecule and considered as transferred charge in order to achieve the equilibrium. The amount of shift in the Fermi level depends on the amount of states in the molecule around the Fermi level. In particular if a molecule has a large density of states (per unit energy) around the Fermi level, the amount of Fermi level shift due to a given amount of charge transfer between the contact and the molecule would be lesser than if the density of states would be small. The total number of electrons in the molecule is in general given by:

$$n = \int_{-\infty}^{+\infty} DOS(E) f_{FD}(E, E_F) dE \tag{5.4}$$

where $DOS(E)$ is the density of states per unit energy of the molecule and $f_{FD}(E, E_F)$ the Fermi-Dirac distribution. At zero kelvin the amount of occupied states (i.e. the number of electrons) is:

$$n = \int_{-\infty}^{E_F} DOS(E) dE \tag{5.5}$$

remember that $n \in \mathbb{R}^+$ since a fractional charge can be transferred. From fundamental theorem of integral calculus:

$$\frac{dn}{dE} = DOS(E) \quad \Rightarrow \quad \left.\frac{dn}{dE}\right|_{E_F} = DOS(E_F) \tag{5.6}$$

and by supposing only small variations of Fermi energy:

$$\delta E_F = \frac{\delta n}{DOS(E_F)} = \frac{q^2}{C_q}\delta n = \frac{q}{C_q}Q_n \tag{5.7}$$

where it is defined the quantum capacitance as: $C_q = q^2 DOS(E_F)$, and where the total channel charge is $Q_n = q\delta n$. «δ» indicates a small variation.

Equation (5.7) holds only for small variations and it is an linearized expression, useful for determining the amount of charge transferred between the contacts and the molecule, for reaching the thermodynamic equilibrium, once the Fermi level shift is known.. From eq. (5.7) it is evident that the greater is the quantum capacitance, the smaller is δE_F for a given δn.

The concept of quantum capacitance can be generalized for any (small) variation of the total channel potential due to whatever effect, such as the application of bias. It strictly holds for small deviations from equilibrium, i.e. for linearized models. In conclusion it is an additional contribution of capacitance that represents the state filling in a nanodevice. More precisely, it is in general a measure of the amount of charge (electrons) that can be transferred into (or from) the molecule in a given bias condition, indeed it is intimately linked to the density of states within the molecule. It must be considered in molFETs capacitive balance, and it is to be used in the determination of the source and drain electrostatic capacitances, as explained in the following.
In practice the calculation of the quantum capacitance becomes too much complicated for large variations. Thus a rough linearized approximated estimation of it, based on the following considerations, was done. The equilibrium quantum capacitance value was determined exploiting the linearized definition:

$$C_q = q^2 DOS(E_F) \tag{5.8}$$

Then out-of-equilibrium values were determined exploiting again eq. (5.8), but with an average density of states DOS_{AV} value instead of $DOS(E_F)$. This average DOS_{AV} was evaluated as the average (arithmetic mean) of the molecule DOS, considering only the values within the bias window for the considered bias point. Moreover an additional average quantum capacitance C_{qAV} was then estimated as the average of all C_q values found in all bias points.

- Source and drain electrostatic capacitances C_s, C_d:

In addition to state filling (i.e. C_q), conventional electrostatic capacitances should be considered. Indeed the movement of charges in response to the applied voltages would lead to capacitive effects. In total there are two effects on the channel potential that are linked to electrostatic capacitances:

- the electrostatic effect: it follows from the capacitance definition, i.e. from the fact that an applied bias would produce a variation in the channel charge
- the charging effect: it is an effect that in general occurs also in conventional electronic devices, but is so small that it is in general negligible. Instead in a nanodevice it is comparable to the electrostatic effect and must be considered in the system modeling. In practice the amount of charge that is moved toward/from the channel alters the energy levels in the channel itself. This effect is very important in all nanodevices, and also in molFETs. Indeed the amount of changing in the channel potential due to a single electron that is moved toward/from it can be of the order of a fraction of eV or even of the order of an eV, thus comparable to the applied bias effect. Notice that this effect occurs also in macroscopic devices, indeed the presence of a charged particle (e.g. an electron) would alter somehow the band diagram, but negligibly in this case.

  In response to a variation in the applied voltage, a given amount of charge $Q_n = q\delta n$ is moved toward/from the channel, accordingly with the total electrostatic capacitance C_{ES}. This process would alter the channel potential energy as expressed in the following equation (see [30] part5 or [29] chapter 7 for a complete treatment):

$$\delta U_{charging} = \frac{q^2}{C_{ES}}\delta n \tag{5.9}$$

  The contribution $\delta U_{charging}$ is then to be summed to the electrostatic effect to get the total effect.

Notice that in general the electrostatic effect and the charging effect are linked, and a self-consistent loop is necessary to ensure convergence.

In order to correctly estimate the source and drain electrostatic capacitances in an easy way, the simple capacitive model of figure 2.24 can be considered. This is again a linearized model in which superposition of effects is used to find the total channel potential variation due to the different contributions, i.e. due to the applied voltages V_{GS} and V_{DS}, and the charging effect.

By exploiting the capacitive model of figure 2.24, by knowing the applied voltages V_{GS} and V_{DS} and the resulting channel potential, it is simple to derive the source and drain capacitances as follows. Exploiting the superposition of effects, the total average channel potential (this is the average potential inside the quantum dot, that in this model is considered without physical dimensions) is given by:

$$U_{totAV} = U_{Vgs} + U_{Vds} + U_{charging} \tag{5.10}$$

$$\Rightarrow U_{totAV} = -q\frac{Cg}{C_{ES}}V_{GS} - q\frac{Cd}{C_{ES}}V_{DS} + q^2\frac{\delta n}{C_{ES}} \tag{5.11}$$

where $C_{ES} = C_s + C_d + C_g$ is the total electrostatic capacitance. Remembering the definition of quantum capacitance, and equations (5.8) and (5.9), it is possible to rewrite eq. (5.10) as follows:

$$U_{totAV} = -q\frac{Cg}{C_{ES}}V_{GS} - q\frac{Cd}{C_{ES}}V_{DS} - \frac{C_q}{C_{ES}}U_{totAV} \tag{5.12}$$

from which:

$$U_{totAV} = -q\frac{Cg}{C_{ES}+C_q}V_{GS} - q\frac{Cd}{C_{ES}+C_q}V_{DS} \tag{5.13}$$

that finally can be rewritten in terms of small (linearized) variations as:

$$\delta U_{totAV} = -q\frac{Cg}{C_{ES}+C_q}\delta V_{GS} - q\frac{Cd}{C_{ES}+C_q}\delta V_{DS} \tag{5.14}$$

At this point considering that the chosen bias points (i.e. V_{GS} and V_{DS}) are of course known, the total channel potential U_{tot} is known from *QuantumATK* simulations, the gate capacitance C_g is known (evaluated starting only from the molFET geometry as explained previously) and the quantum capacitance (in each bias point) was already estimated, the remaining only two unknowns are C_s and C_d. They can be easily estimated if it is assumed that $C_s = C_d$. This hypothesis is good in the ideal case in which the molecule is geometrically symmetric along the transport direction (as usually happens), and it is well (and symmetrically) anchored to the source and drain contacts. In particular this is often true if no dangling chemical bonds are present, and the source and drain contacts are made of same material (and same orientation). Under these hypotheses C_s and C_d were estimated starting from the simple capacitive dividers of equation (5.16). In particular starting from the *QUantumATK* outputs, it was possible to recover the space average (along the transport y-direction) channel potential U_{totAV} for each

bias value. Then by fixing either V_{GS} or V_{DS} (i.e. considering an output characteristics only or a trans-characteristics only), it possible to consider only the relative term in eq. (5.16), thanks to the superposition of effects. Considering only small variations, that means considering the succession of simulated bias points starting from equilibrium and going up step-by-step up to the last one, and then starting again from equilibrium and going down step-by-step till the last one, it is possible to get the total capacitive ratios:

$$\left.\frac{\delta U_{totAV}}{\delta V_{GS}}\right|_{V_{DS}=0} = -q\frac{Cg}{C_{ES}+C_q} \quad \Rightarrow \quad \frac{Cg}{C_{ES}+C_q} \quad known \tag{5.15}$$

$$\left.\frac{\delta U_{totAV}}{\delta V_{DS}}\right|_{V_{GS}=0} = -q\frac{Cd}{C_{ES}+C_q} \quad \Rightarrow \quad \frac{Cd}{C_{ES}+C_q} \quad known \tag{5.16}$$

From which it is possible to find C_s and C_d values in function of the bias. In order to minimize the numerical error, source and drain capacitances were evaluated in the middle bias point between two consecutive simulated bias points, from which the variations δU_{totAV}, δV_{GS} and δV_{DS} were evaluated. Notice that a different script was used to recover the equilibrium C_s and C_d. Indeed, equation (5.7) can be rewritten, by taking into account also the electrostatic effect, as follows ([30]):

$$\delta E_F = \frac{q^2}{C_q}\delta n + \frac{q^2}{C_{ES}}\delta n \tag{5.17}$$

where the first right-side term considers the quantum capacitance and the second one the relative charging electrostatic effect. This equation represents the charge transfer phenomenon to reach the equilibrium condition, and C_q and C_{ES} to be considered are the equilibrium values. The equilibrium quantum capacitance was already estimated (see above). The δn term is the difference in the number of electrons in the isolated molecule and the number of electrons in the molecule with contacts when the thermodynamic equilibrium is achieved. The only unknown is then C_{ES}, that represents the total electrostatic capacitance at equilibrium. We assumed that C_{ES} in equation (5.17) was given only by the sum of the source and the drain capacitances, since the charge transfer (electron transfer) to reach the thermodynamic equilibrium involves essentially only source and drain contacts and not the gate contact. Again once C_{ES} was recovered (from eq. (5.17)), under the previous hypothesis and under the assumption of having $C_s = C_d$ it is trivial to recover them at equilibrium.

Moreover the average Cs_AV and Cd_AV capacitances (simply obtained as the arithmetic mean over all bias point including equilibrium as well) were also estimated, in order to have single scalar values that immediately

provide a fast estimation of the order of magnitude of the source and drain electrostatic capacitances for the considered molFET.

Finally another script was implemented in order to have a very fast, but also very rough, estimation of C_s and C_d. In this case they were evaluated using the parallel plate approximation, thus starting only from the molFET geometry, considering as source contribution half of the molecule in the transport direction, and for the drain contribution the other half. The dielectric constant of the molecule was simply approximated as the vacuum one. Even if this last approach is essentially wrong, it surprisingly led to estimated values that were not so different from the averaged ones (for the considered molecules the Cs_AV and Cd_AV values were only 1.3 times greater than the parallel plate ones).

- Gate resistance R_g:

The molFET gate resistance R_g arises from the device structure, and it is essentially the resistance of the gate oxide. Again in order to keep things easy, we decided to simply estimate R_g as the resistance of the gate piece of insulator. Thus, starting from the knowledge of the insulating material resistivity (in our cases it was always zirconium dioxide ZrO_2) and the geometry of the gate structure (see also above the discussion concerning the gate capacitance), we estimated R_g as:

$$R_g = \frac{\rho_{ZrO2} \cdot t_{OX}}{Area} \tag{5.18}$$

where ρ_{ZrO2} is the ZrO_2 resistivity, t_{OX} is the gate insulator length, and the area is evaluated from the knowledge on the system geometry, similarly to what explained for the gate capacitance C_g.

- Source and drain dynamic resistances R_s, R_d:

As described above, in response to bias voltages variations, a variation in the channel free charges and thus in the device current occur too. This is modeled accordingly with eq. (5.16), by considering the electrostatic capacitances as proportionality factors between voltages and channel charge variations. However this is possible considering also the additional quantum capacitance contribution, i.e. the amount of available electron states (within the bias window), for which transitions between the contacts and the molecule are possible. For example it is possible to have an electron transition from the source contact to the molecular channel only if in the molecule there is at least a free electron state at the considered energy (eigen)value. The amount of the exchanged charged is thus taken into consideration by means of the previously introduced capacitances.

But how much time is required to accomplish a charge movement between the contacts and the molecule, in response to a voltage (small) variation is linked not only to the quantum capacitance concept, that models the state availability in the channel, but also to the concept of escaping time, introduced in equations (2.69)). In particular the total amount of charge that can be transferred between the contacts and the molecule is determined by C_q (i.e. by the DOS), while the time required to move that amount of charge is intimately linked with the escaping times of source and drain contacts τ_s, τ_d. Under the hypotheses of having only small variations, ideal source and drain Ohmic contacts, and to be at zero kelvin, it is possible to show that the current injected by the source contact into the molecule can be approximated as [30]:

$$\begin{aligned} I_S &= \frac{q}{\tau_S} \int_{-\infty}^{+\infty} DOS(E - U_{totAV}) \left[f_{FD}(E, E_{FS}) - f_{FD}(E, E_{Fdot}) \right] dE \\ &\approx \frac{C_q}{\tau_S} \frac{E_{FS} - E_{Fdot}}{q} \end{aligned} \tag{5.19}$$

where E_{FS} is the source Fermi level, E_{Fdot} is the molecule Fermi level, τ_S the source to molecule escaping time and C_q the quantum capacitance. It follows that (even under the hypothesis of having ideal Ohmic contacts) there is a source resistive contribution that, together with C_s, determines the amount of time required in transferring charges toward/from the molecule. This is the physical meaning of R_s. Moreover this is of course only a dynamic quantity that plays a role only in the moment in which input voltages are varied. In particular from eq. (5.19) it is possible to define:

$$R_s = \frac{\tau_S}{C_q} \tag{5.20}$$

Proceeding analogously for the drain contact:

$$R_d = \frac{\tau_D}{C_q} \tag{5.21}$$

In practice C_q was already known for each bias point. From ATK output data it was possible to recover the coupling matrices (for source and drain contacts) in each bias point:

$$\zeta_S = ES_{SD} - H_{SD} \quad , \quad \zeta_D = ES_{DD} - H_{DD} \tag{5.22}$$

where E is the energy, S_{SD} is the source-to-device overlap matrix, H_{SD} is the source-to-device coupling Hamiltonian, and analogous notation is assumed

for drain case. At this point by exploiting energy-time uncertainty relation it was possible to recover the escaping time matrices:

$$\tau_S = \frac{\hbar}{\zeta_S} \quad , \quad \tau_D = \frac{\hbar}{\zeta_D} \tag{5.23}$$

Then the R_s and R_d values for each considered bias point were evaluated by means of equations (5.20) and (5.21). Finally overall average scalar values of R_s and R_d (namely *Rs_AV* and *Rs_AV)* were recovered by averaging over all bias points.

- Cadence circuital dynamic model:

At this point all the dynamic contributions of interest are known, and it is possible to proceed in making more realistic dynamic previsions concerning molFET circuits. In particular from the knowledge of C_g, C_s, C_d it is possible to derive and estimate the dynamic power of the circuits analyzed in the next chapter (6.4). Instead considering also the R_g, R_s and R_d it is possible to estimate the transients duration. In particular in order to do that an equivalent molFET circuit, that includes all these effects, can be derived, simply by placing each contribution next to the relative ideal S, D, G terminal. The values used for C_g, C_s, C_d, R_g, R_s and R_d were the average scalar ones, estimated as described previously. A circuit schematic is reported in figure 5.7 for the nOPV7-molFET case (the others are of course analogous).

Therefore, the methodological flow reported in figure 5.1 is modified as depicted in figure 5.8, by the addition of intrinsic parameters of the device computed in *MatLab*. This enriched LUT-based approach make possible also to perform transient and and dynamic power analysis.

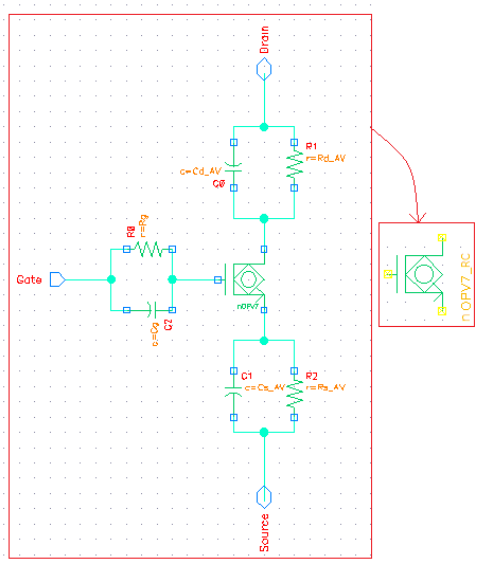


Figure 5.7: MolFET circuital model including C_g, C_s, C_d, R_g, R_s and R_d and corresponding circuit symbol. In this case the nOPV7-based molFET is considered. It is analogous for the other cases.

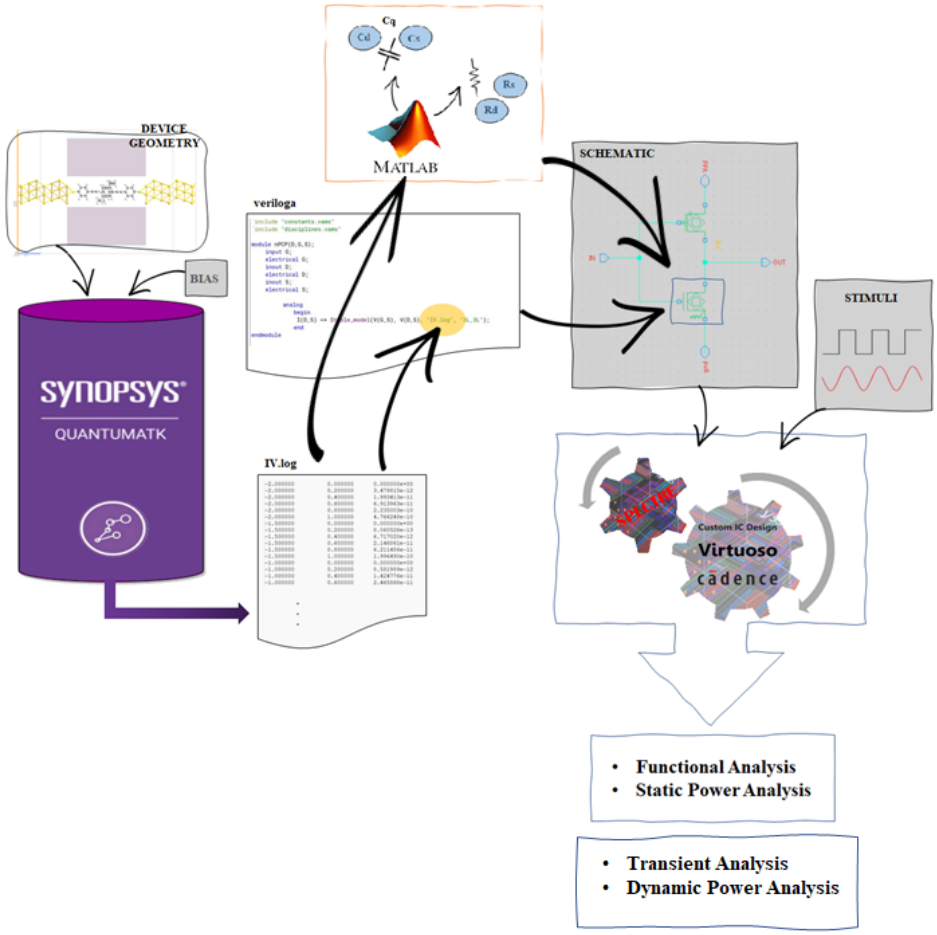


Figure 5.8: The definitive methodological flow adopted for functional, static and dynamic analysis of molFET-based circuits

Chapter 6

Logic implementation: logic gates, HA, FA

The aim of this chapter is to implement molFET-based logic circuit, by exploiting the methodology based on Look-Up-Table discussed in the previous chapter 5. The employed molFETs are molecular ambipolar transistors ad-hoc engineered and characterized in chapter 4 of part II. To achieve this goal, a methodology based on the following steps is adopted:

A. firstly a review on working principles of static CMOS logic is done in section 6.1, mainly in order to outline the key points for an effective functioning of circuits implemented with this logic family. In this step some design hints related to transistor sizing, are theoretically clarified;

B. a design procedure of the main logic elements is performed in Cadence Virtuoso® exploiting a quite recent technology, the FD-SOI MOSFET (28 nm) one. This step actually is not so strictly necessary to achieve the goal of the chapter but is performed in order to acquire a more practical understanding on what are the effects on some important figures of merit when the design constraints outlined in the previous step are satisfied or partially are not. This would be very helpful in the next step for interpret the simulation results in the case of the molFET implementation of the logic circuits. Indeed, in this case, due to molFET behavioural deviation from standard MOSFET an due to the difficulties in satisfying the design constraints, simulation results are expected to be a bit different from the ones obtained for MOSFET implementation. So, in subsections 6.2.1, 6.4, the MOSFET-based logic circuits are respectively designed, verified and characterized in terms of power, speed

and area;

C. in subsections 6.3.1 and 6.4 the design, testing and analysis phases of the previous step are repeated for molFET technology;

D. finally the implementations based on the two different technologies are compared in terms of functional behaviours, power, speed and area in 6.4.

6.1 CMOS Static Logic family: a review

There are many circuit styles to implement a given logic function. The common design metrics by which a circuit is evaluated are area, delay and power. Depending on the application, the focus is on different metrics. The most widely exploited logic style since 1980s, is the static CMOS, thanks to its best performance per watt. It is in practice, an extension of the static CMOS inverter to multiple inputs. The main advantage of the CMOS structure is robustness (i.e, high noise immunity), good performance, and low static power consumption [68].

The CMOS circuit style belongs to a broad class of logic circuits called static circuits in which for each time instant (except during the switching transients), each gate output is connected to either to the supply voltage (V_{DD}) or to the reference voltage (V_{SS}) via a low-resistance path. Moreover, at each time instant (again except transients) the outputs of the gates are always defined and always assume the values of the Boolean function implemented by the circuit. This is different for dynamic circuit class, which consists on temporary storage of signal values with the advantage of simpler and faster gate. However the design and operation of dynamic logic-based circuits are more involved and prone to failure due also to an increased sensitivity to noise.

Conceptually, a static CMOS gate is a combination of two networks connected together, the pull-up network (PUN) and the pull-down network (PDN) (figure 6.1). Within the former network there are combinations of p-MOSFETs in series or parallel or both, connected to a supply voltage. Indeed the function of the PUN is to provide, according to the inputs, a connection between the output and V_{DD} anytime the output of the logic gate is meant to be 1. Instead, within the latter network, there are combinations of n-MOSFETs in series or parallel or both, connected to a reference voltage. Similarly to PUN, the function of the PDN is to connect the output to V_{SS} when the output of the logic gate is meant to be 0. The PUN and PDN networks are ,for construction, mutually exclusive in such a way that one and only one of the networks is conducting at steady state. In this way, once the transients are

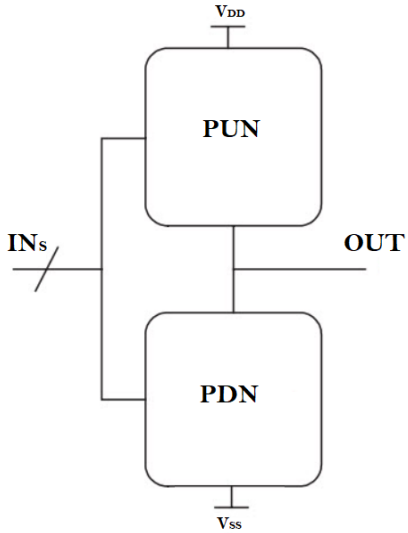


Figure 6.1: Conceptual schematic of a static CMOS gate.

extinguished, exists always a path between V_{DD} and the output, realizing a high logic value output (1), or, alternatively, between V_{SS} and output for a low logic value output (0).

In designing the PDN and PUN networks, the following considerations should be kept in mind:

- An nMOS is ON when the controlling gate signal is high and is OFF when it is low. For a pMOS the opposite happens, ON when is low and OFF when is high.
- All Booelean logic functions can be obtained with this structure, according to a set of construction rules. In particular nMOS devices connected in series corresponds to an AND function, and connected in parallel represent an OR one. Whereas a series connection of pMOS conducts if both inputs are low, representing a NOR function, while pMOS transistors in parallel implement a NAND.
- This structure is naturally inverting, implementing only functions such as NAND, NOR, and XNOR. The realization of a non-inverting Boolean function (such as AND OR, or XOR) in a single stage is not possible: it necessarly requires an extra inverter stage.
- A careful transistor sizing procedure must be performed in order to equalize the response time in charcging/discharging the load capacitance

of PDN and PUN, and thus optimize the gate propagation delay and avoid glitches in more complex circuits. This procedure nowadays is directly performed by CAD tools.

6.2 MOSFET-based Logic circuits

In this section some logic gates (NOT, NAND, NOR, XOR) are implemented with a quite recent technology, exploiting ST FD-SOI 28nm technological library in Cadence Virtuoso. The focus is on them since they are the base of much more complex modules, such as multipliers and processors. Then the same step is repeated for some basic adders (Half-Adder, Full-Adder). In subsections 6.2.1, **??**, they are respectively designed, verified and characterized in terms of power, speed and area.

6.2.1 Design and testing

Inverter. The inverter is truly the nucleus of all digital circuits. Indeed, the electrical behavior of more complex circuits can be almost completely derived by extending the results obtained for inverters.

Its operation is readily understood with the aid of the simple switch model of the MOS transistor: in sub-threshold region (i.e. $|V_{GS}| < |V_{th}|$) the transistor, at first approximation, can be considered as a switch with an infinite OFF resistance, and a finite ON-resistance above the threshold (i.e. $|V_{GS}| \geq |V_{th}|$).

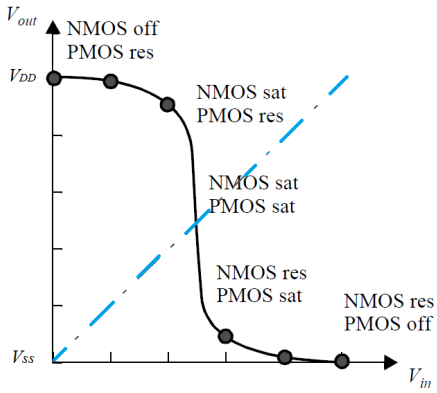


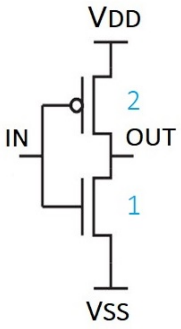
Figure 6.2: Voltage-Transfer characteristic (VTC) of the inverter.

Thus, for the inverter gate, that is made of a pMOS as pull-up transistor connected to the voltage supply V_{DD} and of a nMOS as pull-down connected to the reference voltage V_{SS}, the following functioning yields. For low input voltage V_{in} the nMOS is OFF and the pMOS is in resistive (or triode) region, thus the output voltage goes to V_{DD} because a direct path exists between

V_{DD} and OUT. For increasing values of V_{in} , the nMOS starts conducting until for a specific point, known as the *inversion point*, both nMOS and pMOS are in saturation region and behave as current generators in series. Up to V_{THn} the nMOS conducts in resistive region and the pMOS turns OFF, thus the output goes to the reference voltage V_{SS}, since a direct path exists between OUT and the reference node, resulting in a steady-state low output value (V_{SS}, typically 0V). The gate clearly behaves like an inverter, whose truth table is shown in 6.2. All these considerations translate in the Voltage-Transfer characteristic (VTC) of the inverter, shown in figure 6.2.

During the transient region the gate dissipate power since a direct path between supply and reference is present. Therefore, in order to minimize this waste, the transient region must be reduced as much as possible and this is done by exploiting transistors with a fast switch response, able to pass from saturation to resistive region as fast as possible.

Moreover, in order to optimize the noise margins, the inversion point is placed in the middle of the voltage swing ($V_{DD}/2$ if the reference voltage is 0V). This choice, it can be shown, implies a similar drive ON current both for pull-up transistor and pull-down, which leads to the optimization of the propagation delay of the gate. For this reason the pMOS gate width should be sized 2-3 times larger than the nMOS one in such a way to compensate the differences in the carrier's mobility and thus get the same driving strength. The conceptual and Cadence schematics are reported in figure 6.3.



NOT	
IN	OUT
0	1
1	0

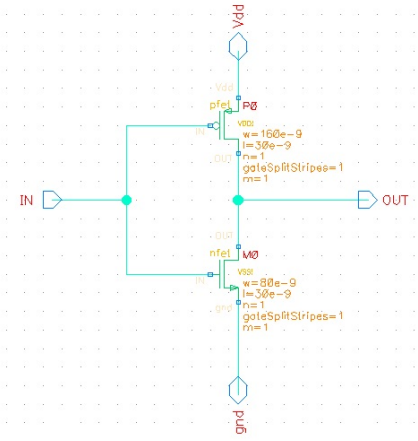


Figure 6.3: Conceptual and Cadence schematic of the inverter with transistors properly sized. Also the truth table of the inverter is shown.

Then the schematic was tested. Two simulations were performed: (A.) a sweep was performed in the input signal of the gate and the output was probed, in order to get the VTC of the inverter; (B.) a DC functional verification consisting on applying the two possible different signal configurations at the input was finally performed to verify the correct behavior of the gate. As shown in the following figures 6.4, 6.5, everything correctly works as expected. Comments are reported directly in the figures captions.

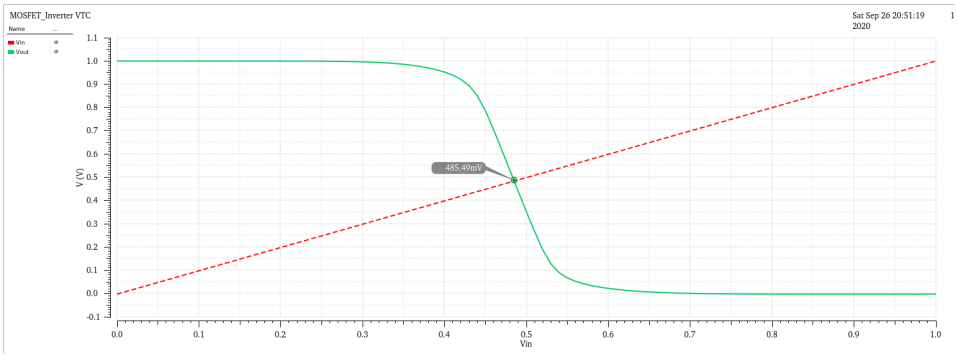


Figure 6.4: VTC of the inverter simulated in Cadence. The input signal V_{in} was swept from 0 to 1 V. The obtained VTC is not perfectly symmetric due to the fact that the perfect matching of driving strenghts of pMOS and nMOS is not completely achieved by means of the transistor sizing procedure. In particular the inversion point V_M is around 485 mV, slightly above $V_{DD}/2 = 0.5V$ resulting in a shift towards 0V of V_M. This imply that the electrical strength of nMOS is slightly bigger w.r.t. pMOS transistor, thus a reduced noise margin.

The other basic logic gates were designed accordingly to the following rules of thumb. In general transistors must be sized according to the ratio of reference inverter. More in particular:

- for N transistors in series: N must be multiplied by the size of the reference inverter.
- for M transistors in parallel: the size of each transistor must follow the same size of the reference inverter.

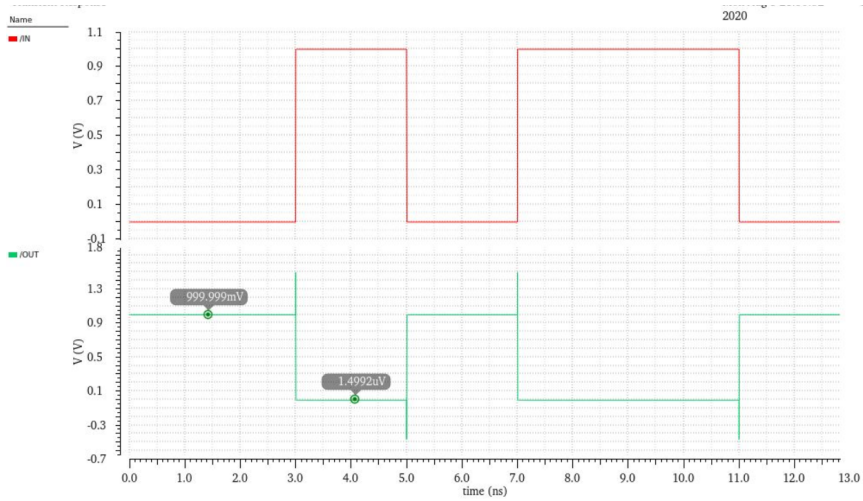
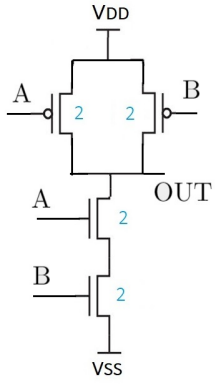


Figure 6.5: DC functional verification of the inverter gate. The output is correctly complemented. Obviously the output dynamic is reduced w.r.t. the input one, due to very negligible unwanted voltage drops on the R_{ON} of the ON transistors.

NAND gate. Transistors in NAND gate are sized in the following way shown in figure 6.6. Since there are two nMOS in series, the ratio of nMOS of the reference inverter (i.e. 1) is multiplied by two. Concerning the size of pMOS it is the same of the reference inverter (i.e. 2).



NAND		
A	B	OUT
0	0	1
0	1	1
1	0	1
1	1	0

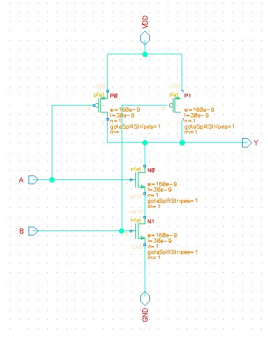


Figure 6.6: Conceptual and Cadence schematic of the NAND with transistors properly sized. Also the truth table of the NAND is shown.

The schematic in figure was tested with a DC functional verification consisting on applying the possible different input signal configurations. In figure 6.7 is shown the simulation result where the expected behaviour is verified.

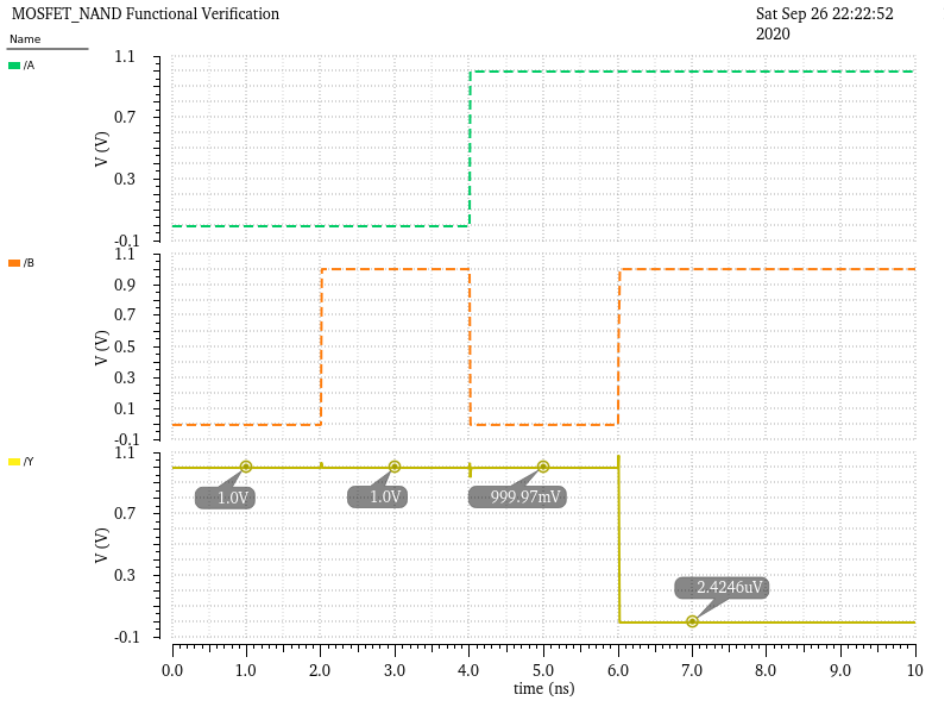
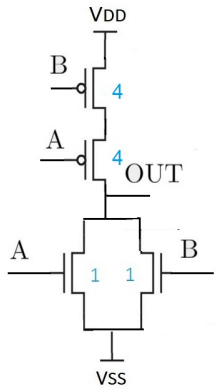


Figure 6.7: DC functional verification of the NAND gate. The output correctly follows the NAND truth table. A (green dash line) and B (orange dash line) are the inputs. Y (yellow solid line) is the output.

NOR gate. Transistors in NOR gate are sized in a complementary way w.r.t. NAND gate, as shown in figure 6.8. Since there are two pMOS in series, the ratio of pMOS of the reference inverter (i.e. 2) is multiplied by two. Concerning the size of nMOS it is the same of the reference inverter (i.e. 1).



NOR		
A	B	OUT
0	0	1
0	1	0
1	0	0
1	1	0

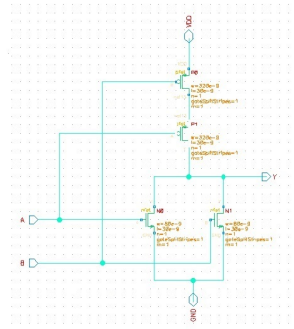


Figure 6.8: Conceptual and Cadence schematic of the NOR with transistors properly sized. Also the truth table of the NOR is shown.

The schematic in figure was tested with a DC functional verification consisting on applying the possible different input signal configurations. In figure 6.9 is shown the simulation result where the expected behaviour is verified.

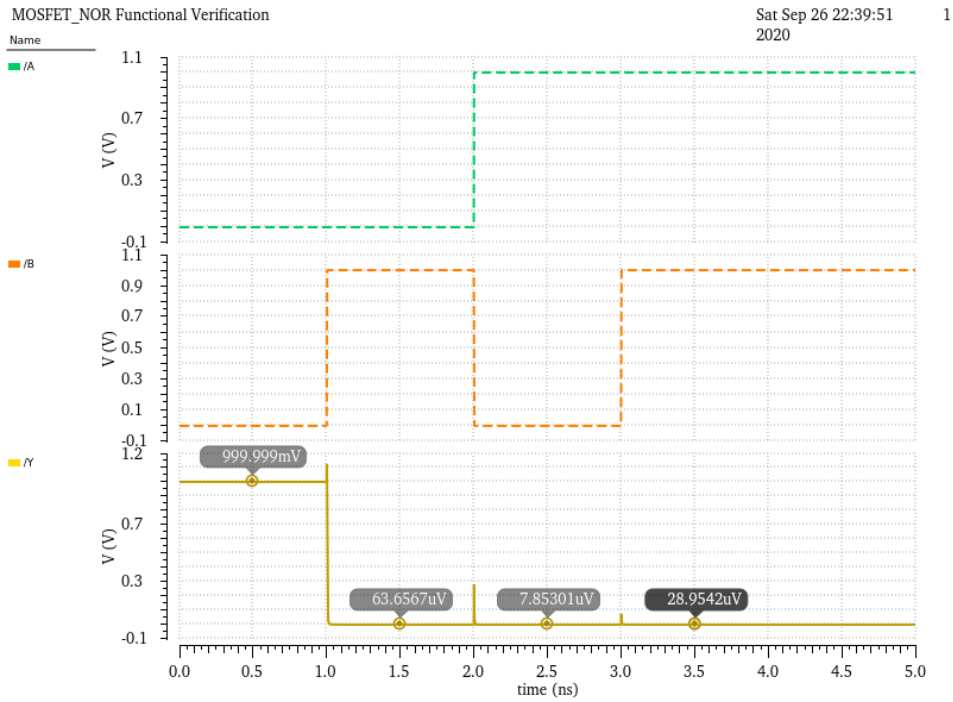
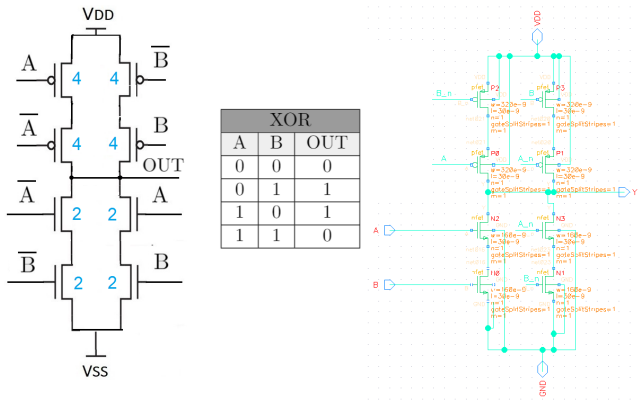


Figure 6.9: DC functional verification of the NOR gate. The output correctly follows the NOR truth table. A (green dash line) and B (orange dash line) are the inputs. Y (yellow solid line) is the output.

XOR gate. The transistors in XOR gate are sized in the following way shown in figure 6.10. Since there are two pMOS in series for each side, the ratio of pMOS of the reference inverter (i.e. 2) is multiplied by two for both sides. Similarly for nMOS: there are two in series for each side, thus the ratio of nMOS of the reference inverter (i.e. 1) is multiplied by two for both sides.

The schematic in figure was tested with a DC functional verification consisting on applying the possible different input signal configurations. In figure 6.11 is shown the simulation result where the expected behaviour is verified.



XOR		
A	B	OUT
0	0	0
0	1	1
1	0	1
1	1	0

Figure 6.10: Conceptual and Cadence schematic of the XOR with transistors properly sized. Also the truth table of the XOR is shown.

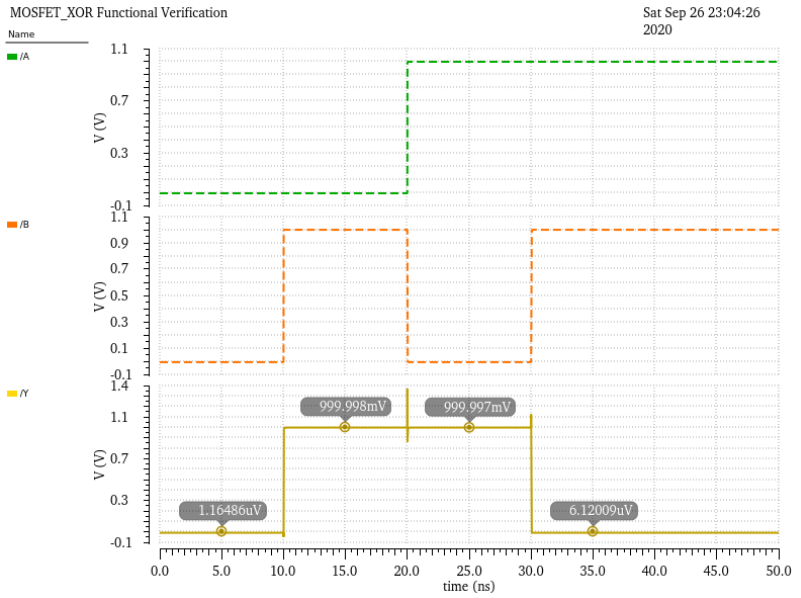
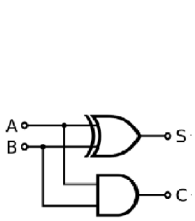


Figure 6.11: DC functional verification of the XOR gate. The output correctly follows the XOR truth table. A (green dash line) and B (orange dash line) are the inputs. Y (yellow solid line) is the output.

Half Adder. The NAND, NOT, XOR gates as previously designed are assembled together in the following configuration shown in 6.12. Actually the transistor sizing procedure must be repeated. The schematic in figure was



Half Adder			
A	B	S	Cout
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

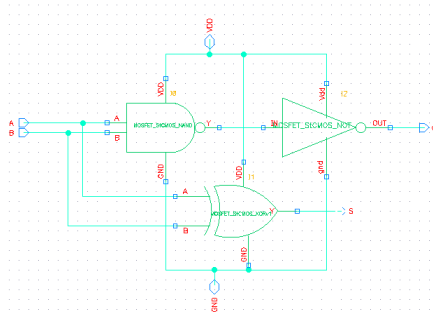


Figure 6.12: Conceptual and Cadence schematic of the HA. Also the truth table of the HA is shown.

tested with a DC functional verification consisting on applying the possible different input signal configurations. In figure 6.13 is shown the simulation result where the expected behaviour is verified.

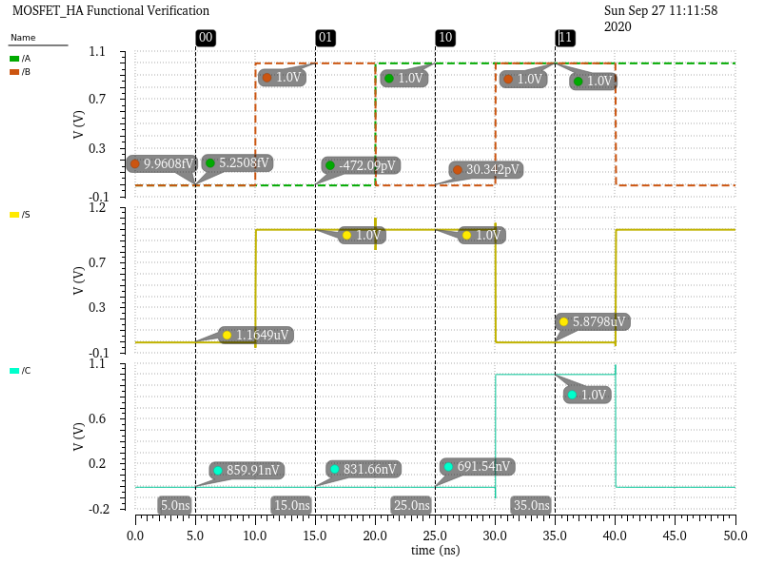
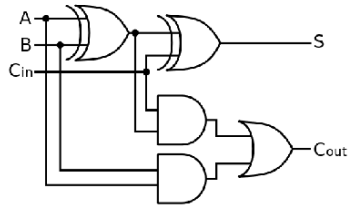


Figure 6.13: DC functional verification of the HA.

Full Adder. Two HAs are connected in cascade to create a FA as shown in figure 6.14. The Cadence schematic in figure 6.15 was designed and tested with a DC functional verification consisting on applying the possible different input signal configurations. In figure 6.16 is shown the simulation result where the expected behaviour is verified.



Full Adder				
A	B	Cin	S	Cout
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Figure 6.14: Conceptual schematic and truth table of the FA.

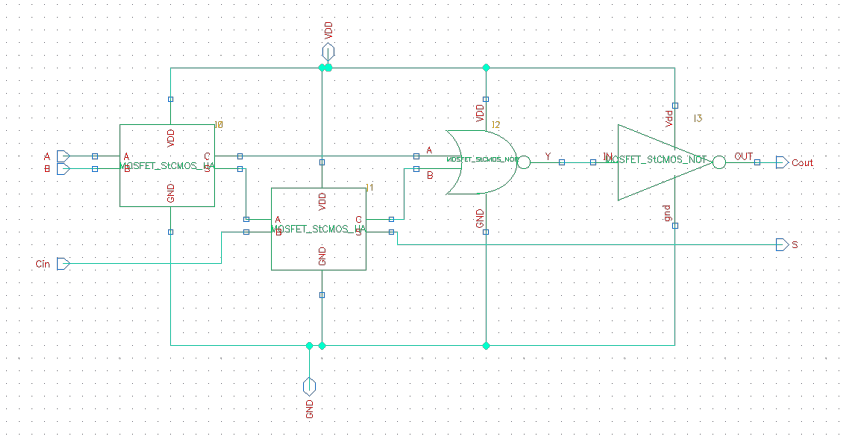


Figure 6.15: Cadence schematic of the FA.

Notice that as for HA, actually the transistor sizing procedure must be repeated in order to avoid unwanted spurious transitions (glitches) and thus waste of power. Indeed glitches in this circuit arise from the fact that the two different logical paths are not balanced in terms of propagation delay of the gates. Typically this balancing operation by tuning the channel width of

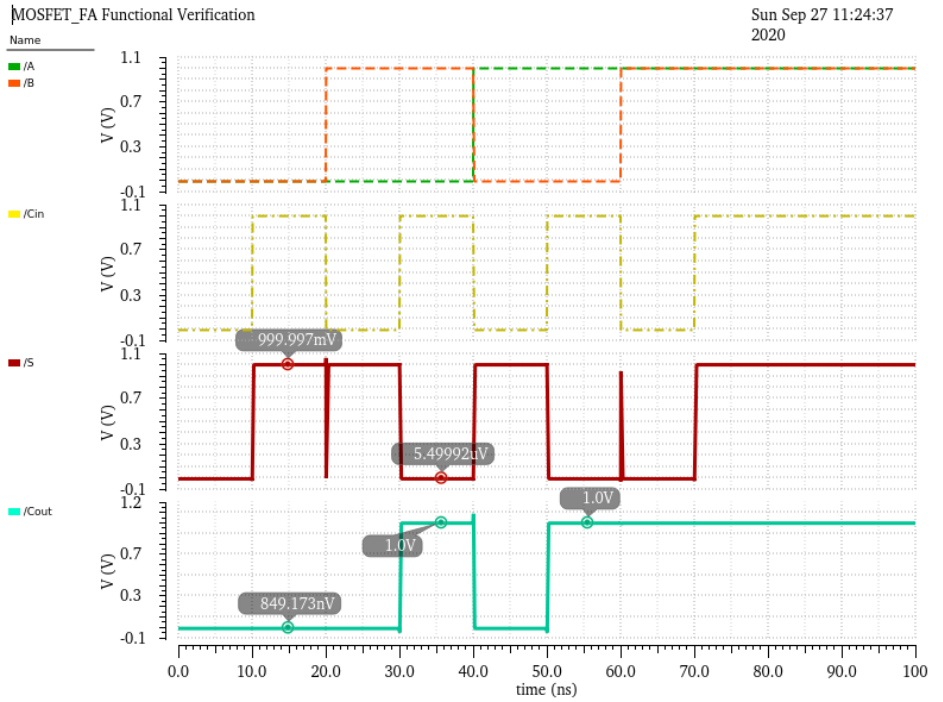


Figure 6.16: DC functional verification of the FA. Glitches are present.

transistors is done automatically by synthesis CAD-tools.

6.3 MolFET-based Logic circuits

6.3.1 Design and testing

The logic elements implemented in the previous section in MOSFET technology are now implemented in molecular technology, by simply replacing in each schematic MOSFETs with molecular transistors, and properly adjusting the supply and reference voltages. Indeed concerning MOSFET-based circuits, V_{DD} was set equal to 1V and the reference voltage V_{SS} to 0V. The high logic value was set equal to 1V and the low one to 0V. On the other hand, for molFET-based circuits n-molFET was engineered, in 4, in such a way that is ON when its gate voltage signal V_{GS} is equal to 1V, and OFF when its gate voltage signal V_{GS} is equal to -1V. Similarly, p-molFET was engineered in such a way that happens the opposite cases: ON for $V_{GS} = -1V$, OFF for $V_{GS} = 1V$. In order to make compatible these voltage values with the ones of the next stages:

- the high logic value was set equal to 0V, whereas the low logic value to -1V.
- V_{DD} was set equal to 0V, whereas V_{SS} equal to -1V.

Indeed if the same set up of MOSFET-based circuits is kept, considering for e.g. a CMOS inverter: when the n-molFET is ON, it will ideally pull down the output to 0V, letting the pull-up network of the next stage turning completely ON but also not allowing the pull-down one to turn completely OFF. Indeed at OV the n-molFET is conducting, differently from nMOSFET.

Notice also that even if the exploited molFETs are engineered to works in static CMOS logic-based circuits, it is not possible to perform the transistor sizing procedure as done for MOSFETs-based circuits, due to the limitation in the adopted LUT model for circuits simulations. Moreover in molecular transistors, tune the channel width is not possible: is not a solid crystalline layer of doped silicon that can be enlarged or reduced by means of photolithography. The molecule is that one, cannot be enlarged or reduced in size, and even if it is possible (for e.g. by means of mechanical strain or high electric field) this would lead to completely different transport response beacause of the modification of molecular orbitals. A possibility could be to stack more than one molecule between the source and drain electrodes.

Anyway these more exotic device configurations are not considered in this work, therefore is not possible, by changing a parameter in the device geometry, to increment/decrease the electrical strength of the molFET, thus perform the transistor sizing procedure. As a consequence noise margins and

gate delay would not be optimized, and in some cases gate malfunctioning are expected from simulation results, expecially for circuits made of more than one stage in cascade (e.g. XOR, HA, FA..).

OPV7 Inverter. In the following figures are shown the *Spectre* simulation results of OPV7-based inverter. In 6.17 the simulated VTC. An input voltage sweep is performed within the whole useful dynamics (i.e. from -1V to 0V). Anyway the response at the output is characterized by an unacceptable noise margin for the '0' logic value. Indeed when at the input is applied a low logic value (-1V), then at the output a voltage quite close to the high logic value (i.e. 0V) is generated (around $-4.1\,\text{mV}$), thus a quite good noise margin for the high logic value is obtained. On the other hand, when at the input is applied an high logic value (0V), then at the output a voltage too much different from the low logic value (i.e. -1V) is generated (around $-202.9\,\text{mV}$), thus unacceptable. The reason of this behaviour can be traced back to the fact that the nOPV7 molFET has not enough electrical strenght to force the output to the reference voltage. That is also the reason why the inversion point of this inverter is around -0.1V, quite different from the value at the half of the dynamic (-0.5V).

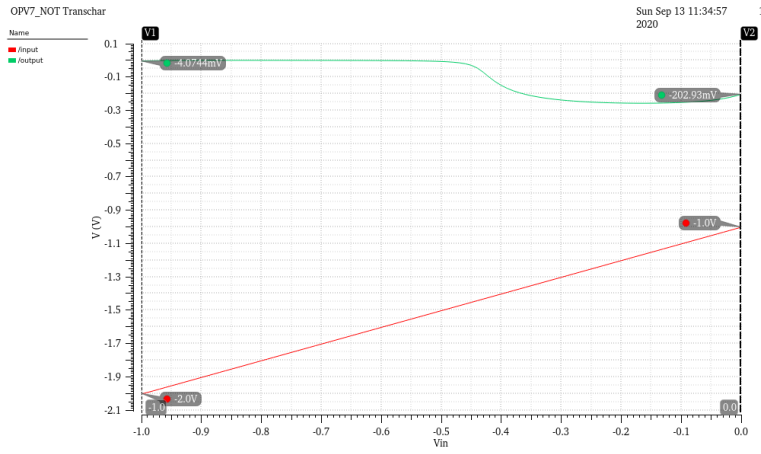


Figure 6.17: Simulated VTC of OPV7 molFET-based inverter. The red line is the input voltage V_{in} and the green one the output V_{out}.

In figure 6.18 the behaviour of OPV7 molFET-based inverter is tested. It correctly behaves but with unacceptable noise margins since a stage put in cascade will see an input swing almost halved, thus will hardly work as an inverter. In particular the noise margin for the low logic is reduced of around 20.1%, whereas the one for the high logic value is reduced of around 50.8%.

Moreover the input voltage transition from -1V to 0V creates problems in the functioning, an oscillating behaviour is present due to numerical convergence

issues

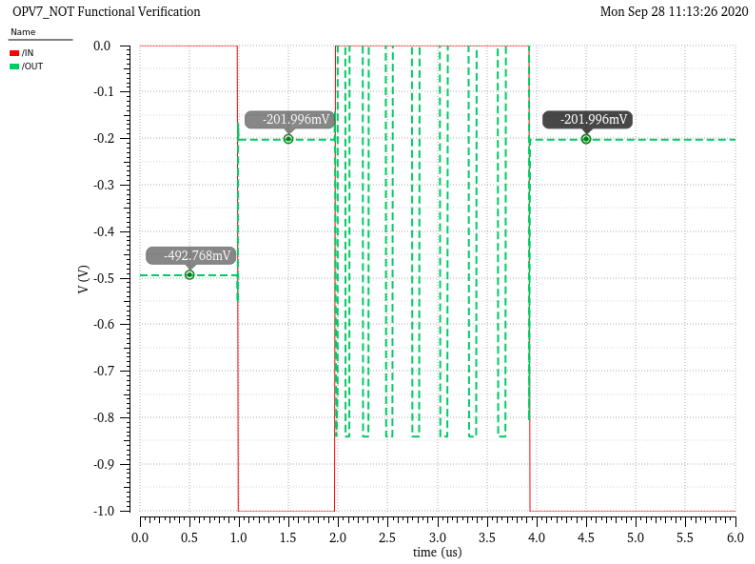


Figure 6.18: Functional verification simulation of OPV7 molFET-based inverter.

PCP Inverter. In the following figures are shown the *Spectre* simulation results of PCP-based inverter. In 6.19 the simulated VTC. An extended input voltage sweep w.r.t. the previous case is performed from -1V to 1V in order to show also the inversion point, i.e. the input voltage over which the p-molFET starts turning OFF.

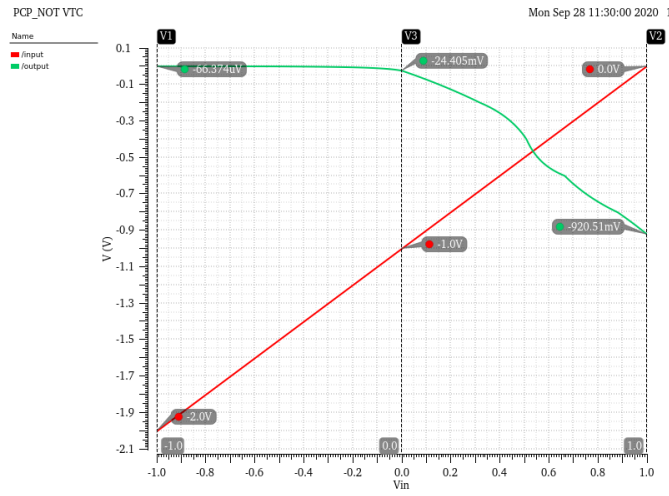


Figure 6.19: Simulated VTC of PCP molFET-based inverter. The red line is the input voltage V_{in} and the green one the output V_{out}.

In figure 6.20 the behaviour of PCP molFET-based inverter is tested. It correctly behaves with acceptable noise margins. Obviously by putting in cascade a stage, the noise margin will gradually deteriorate until getting no more acceptable values. In particular the noise margin for the high logic is reduced of around 2.4%, whereas the one for the low logic value is reduced of around 8%.

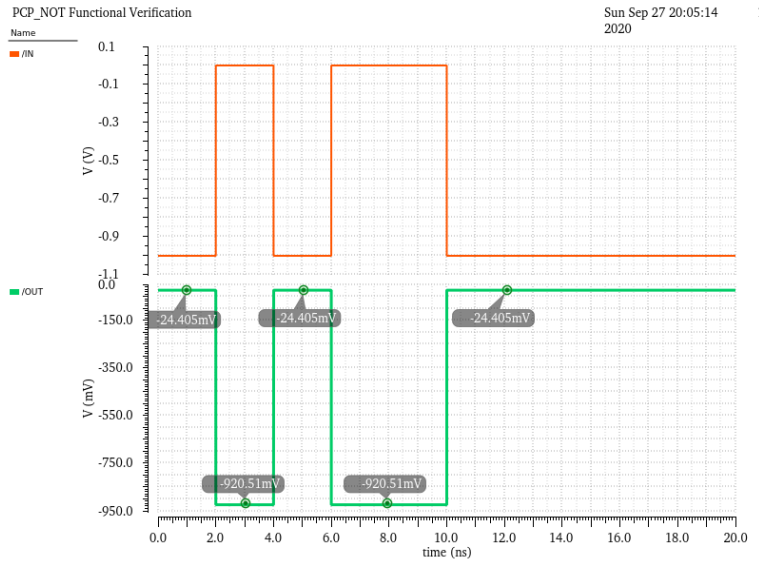


Figure 6.20: Functional verification simulation of PCP molFET-based inverter.

OPV7 NAND. In figure 6.21 is shown the *Spectre* simulation results of OPV7-based NAND. In this simulation its behaviour is tested. It not correctly behaves at all. Only the input configuration with both low logic value generates at the output a voltage value close to the correct one. The circuit cannot clearly work as NAND gate due to the electrical weakness of nOPV7 molFETs that do not manage to pull down the output line.

PCP NAND. In figure 6.22 the behaviour of PCP molFET-based NAND is tested. It correctly behaves with acceptable noise margins. Obviously by putting in cascade a stage, the noise margin will gradually deteriorate until getting no more acceptable values. In particular the noise margin for the high logic can be deteriorated of a percentage ranging from 0.1% to 1.7%, whereas the one for the low logic value is reduced of around 20.7%.

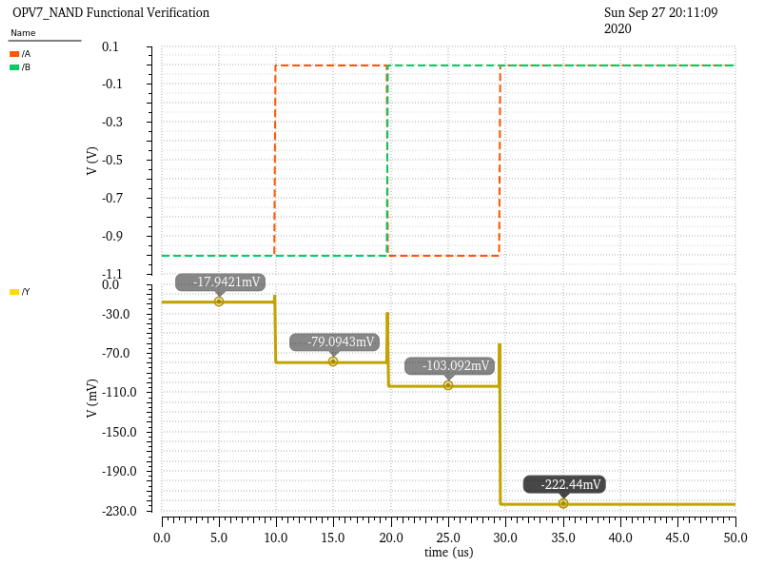


Figure 6.21: Functional verification simulation of OPV7 molFET-based NAND.

OPV7 NOR. In figure 6.23 is shown the *Spectre* simulation results of OPV7-based NOR. In this simulation its behaviour is tested. It not correctly behaves at all for any configuration. Also steady state values really out of the dynamic of the circuit are genereated, thus clearly behaves as a NOR gate due to the inability of the pull-up and -down network to work correctly.

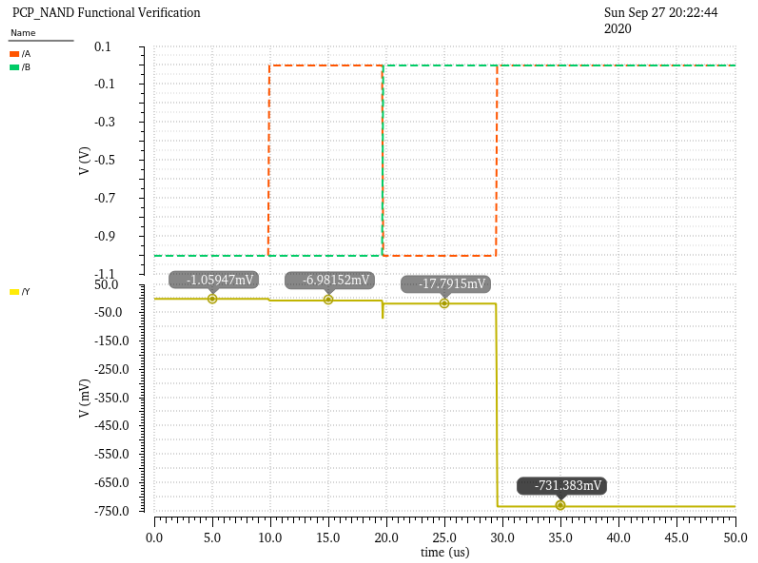


Figure 6.22: Functional verification simulation of PCP molFET-based NAND.

PCP NOR. In figure 6.24 the behaviour of PCP molFET-based NOR is tested. It correctly behaves with acceptable noise margins. Obviously by putting in cascade a stage, the noise margin will gradually deteriorate until getting no more acceptable values. In particular the noise margin for the high logic is reduced of around 8.4%, whereas the one for the low logic value can be deteriorated of a percentage ranging from 1.2% to 7.4%.

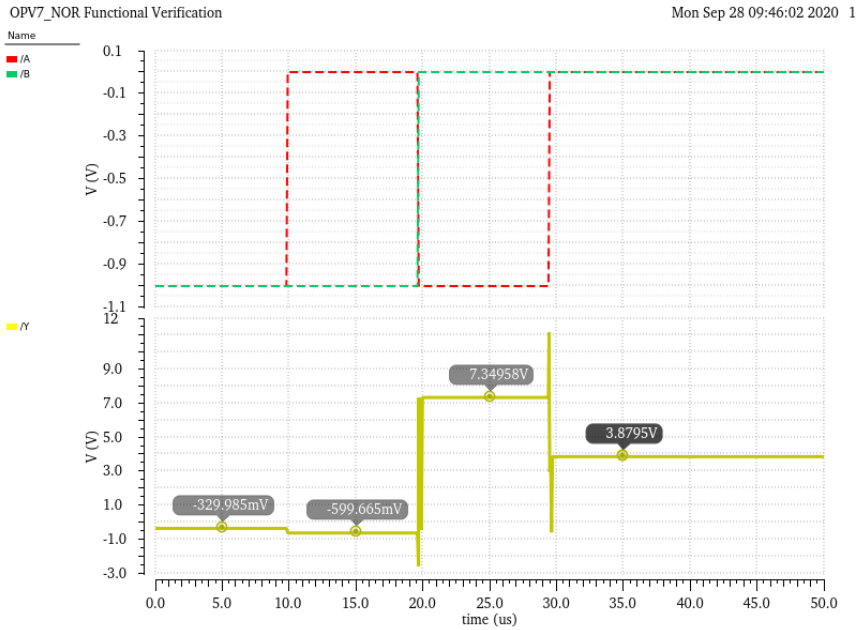


Figure 6.23: Functional verification simulation of OPV7 molFET-based NOR.

OPV7 XOR. In figure 6.23 is shown the *Spectre* simulation results of OPV7-based XOR. In this simulation its behaviour is tested. Actually, after a certain point, the convergence is no more reached due to presence of too high non-linearities. Anyway, in the time interval in which the convergence is reached the circuit does not correctly behaves at all for any configuration. Steady state values are really out of the dynamic of the circuit. Thus, also in this case, clearly OPV7-molFETs as they have been engineered, are not good candidate to be assembled in order to form a XOR gate.

PCP XOR. In figure 6.24 the behaviour of PCP molFET-based XOR is tested. It correctly behaves with acceptable noise margins, even if are not so good enough comparing w.r.t. the gates tested previously. Now the situation will be worst for degradation of noise margins, by putting in cascade a stage. In particular the noise margin for the high logic value is reduced of around 5.5%, whereas the one for the low logic value can be deteriorated of a percentage ranging from 23.8% to 27.8%.

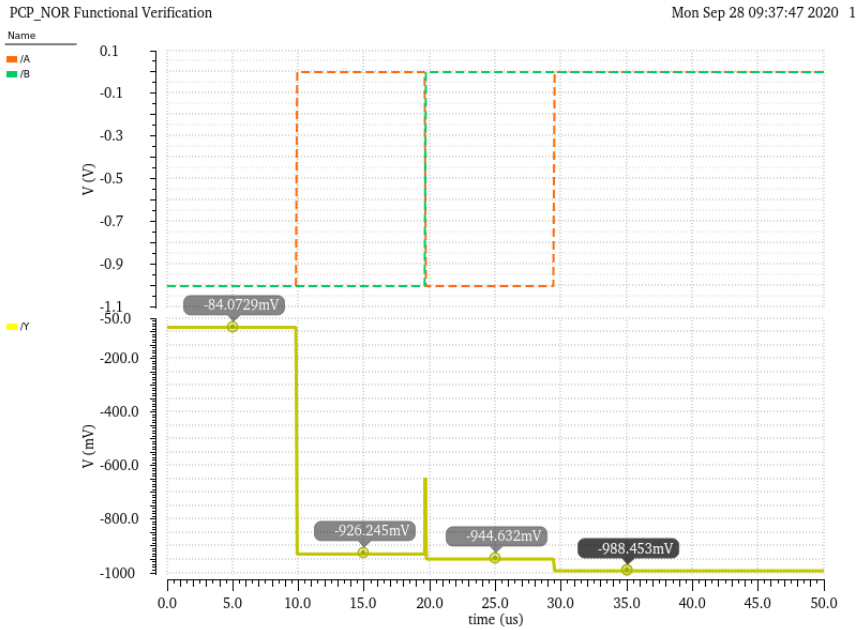


Figure 6.24: Functional verification simulation of PCP molFET-based NOR.

OPV7 HA. In figure 6.27 is shown the *Spectre* simulation results of OPV7-based HA. In this simulation its behaviour is tested. In this case the convergence is always reached. However the circuit does not correctly behaves at all for any configuration, for both outputs. Indeed, since already the XOR gate of which is made of the sum path, does not work as shown previously, the same is expected for the whole HA. Moreover, also in this case, especially for the output signal of the sum S, steady state values are really out of the dynamic of the circuit.

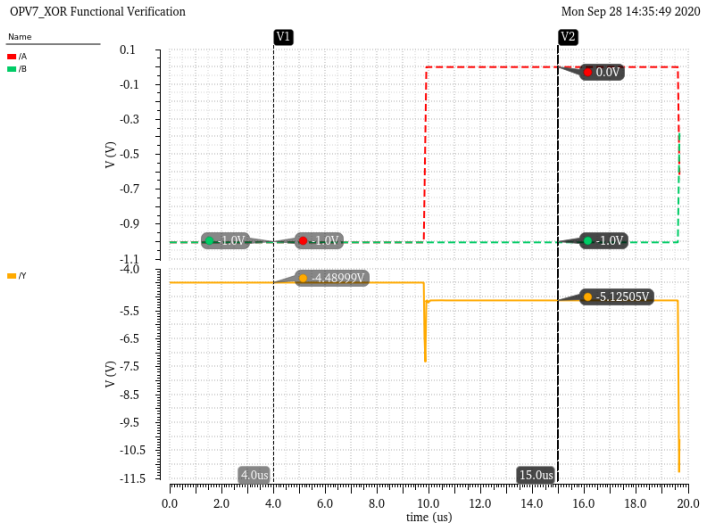


Figure 6.25: Functional verification simulation of OPV7 molFET-based XOR. After 20 µs the convergence is no more reached.

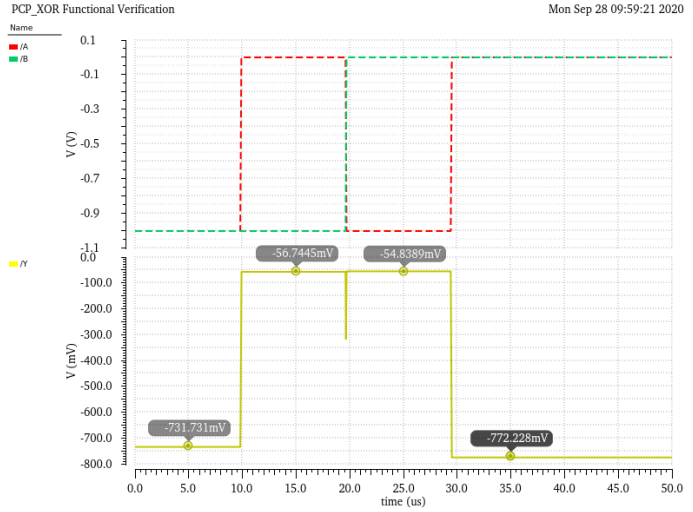


Figure 6.26: Functional verification simulation of PCP molFET-based XOR.

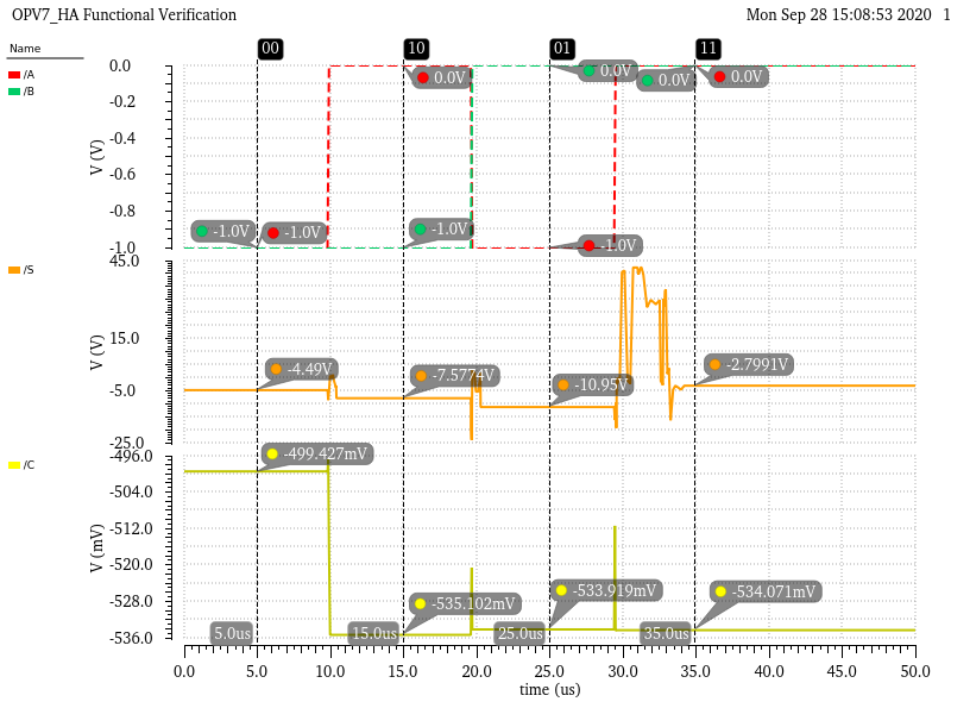


Figure 6.27: Functional verification simulation of OPV7 molFET-based HA.

PCP HA. In figure 6.28 the behaviour of PCP molFET-based HA is tested. It correctly behaves with acceptable noise margins, even if are not so good enough comparing w.r.t. the gates tested previously. So also in this case, the situation will be worst for degradation of noise margins, by putting in cascade a stage, as will be done next for FA. In particular the noise margin for the high logic value is reduced of around 5.5% for the S signal, and of around 16.8% for the C signal. Concerning the noise margin for the low logic value, it can be deteriorated of a percentage ranging from 8.1% to 9.9% for the C signal and from 22.8% to 27% for for the S signal .

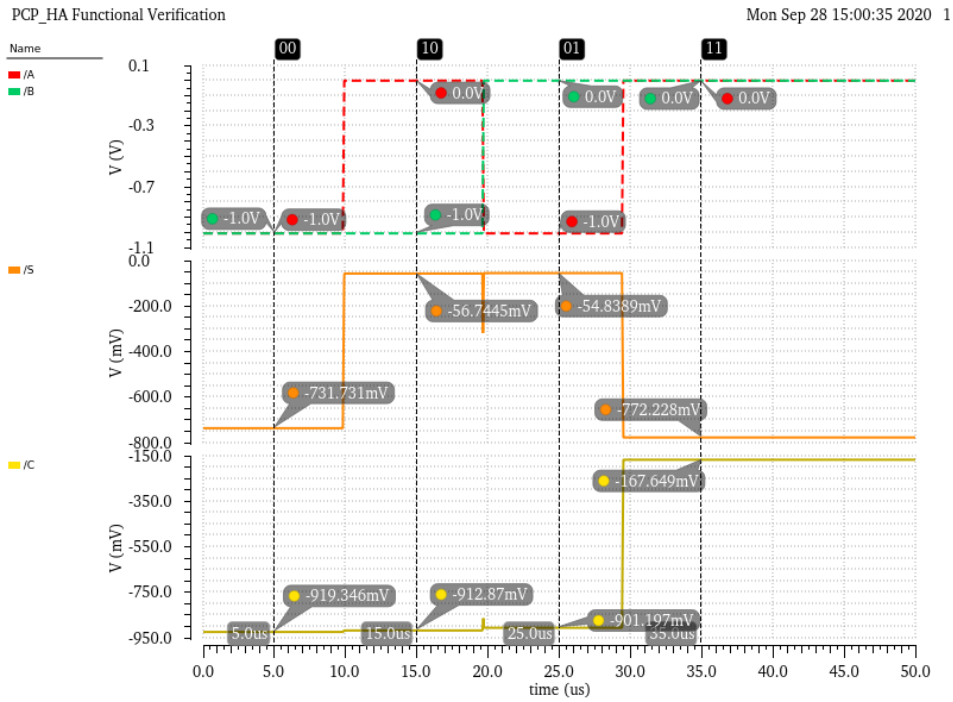


Figure 6.28: Functional verification simulation of PCP molFET-based HA.

OPV7 FA. In figure 6.29 is shown the *Spectre* simulation results of OPV7-based FA. In this simulation its behaviour is tested. Actually, after a certain point, the convergence is no more reached due to presence of too high non-linearities. Anyway, in the time interval in which the convergence is reached the circuit does not correctly behaves at all for any configuration. Some steady state values are really out of the dynamic of the circuit.

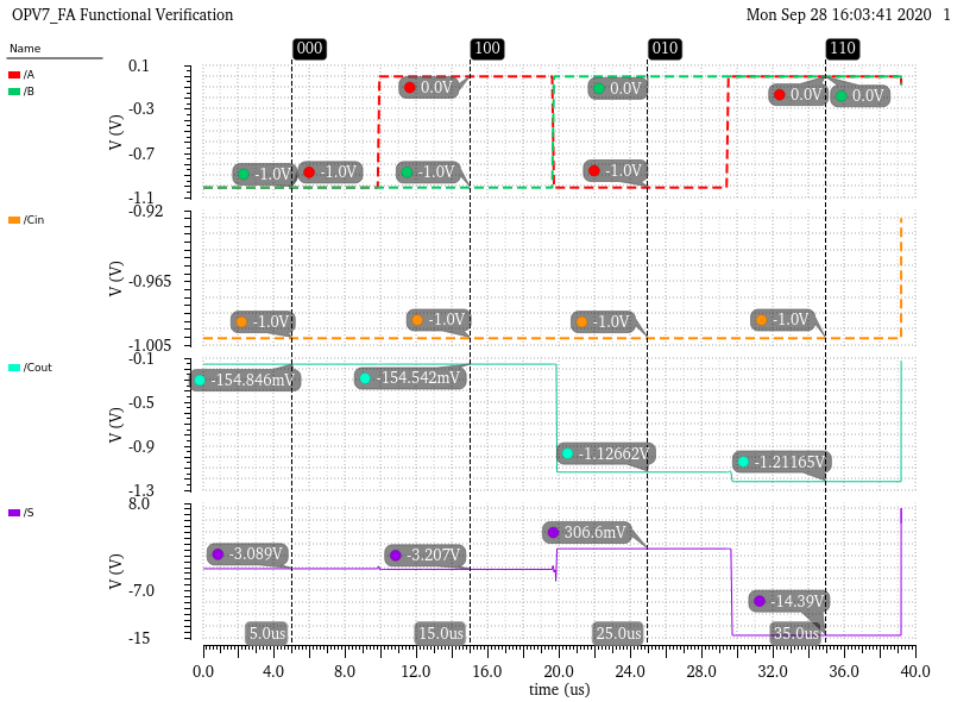


Figure 6.29: Functional verification simulation of OPV7 molFET-based FA.

PCP FA. In figure 6.30 the behaviour of PCP molFET-based FA is tested. It correctly behaves with acceptable noise margins, even if are not so good enough comparing w.r.t. the gates tested previously. In particular the noise margin for the high logic value is reduced of around 5.5% for the S signal, and of around 16.8% for the C signal. Concerning the noise margin for the low logic value, it can be deteriorated of a percentage ranging from 8.1% to 9.9% for the C signal and from 22.8% to 27% for for the S signal .

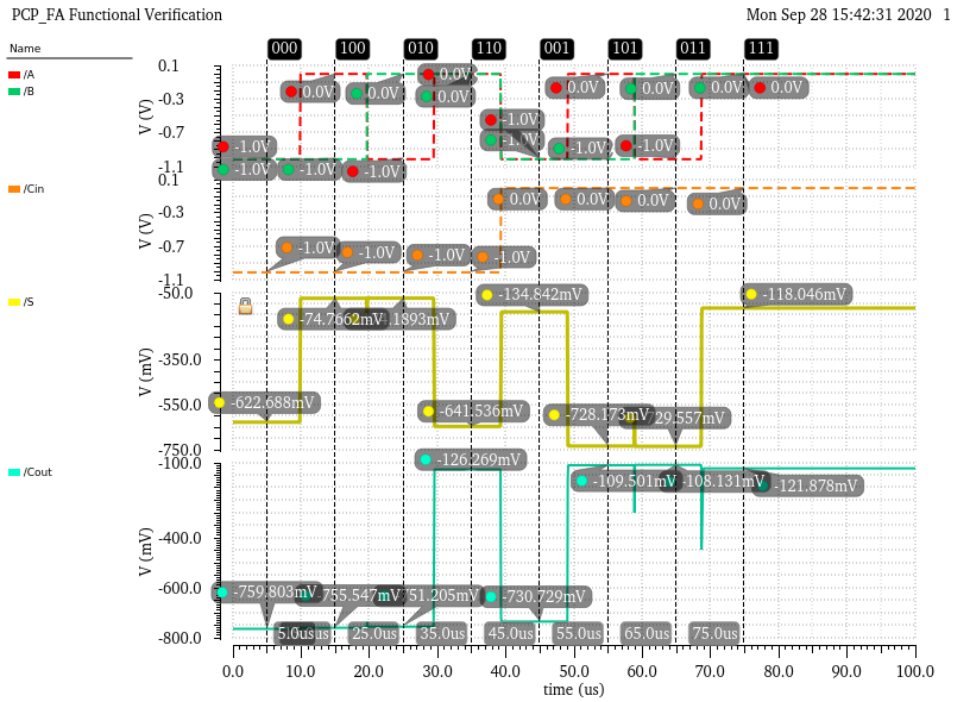


Figure 6.30: Functional verification simulation of PCP molFET-based FA.

6.4 Analysis and comparison

The analysis performed in this subsection are related to the area occupation, static and dynamic power dissipated and maximum operating frequency achievable for both MOSFET and MolFET technologies. The analysis is directly made on the FA circuit, since is the more complex logic circuit investigated in this thesis.

6.4.1 Area occupation

Area is the main advantage of molecular electronics, moreover it is an important parameter to be analyzed. To keep the analysis as simple as possible, only the occupied area of the transistors was considered in the comparison, neglecting the areas of the lines. Nevertheless this comparison is reasonable in terms of the area ratios between the different technologies, since the lines

contributions are essentially the same in all technologies, thus leaving the proportionality between areas essentially unchanged. Of course this will not be a precise and realistic estimation of areas, but just an estimation of the order of magnitude, a first attempt in a quantitative comparison.
In the following tables the different technologies are compared in terms of occupied area in the case of a gate terminal with a reduced value of width depending on molecule (from ATK geometries: 5 Å was set for OPV7, and 10 Å for PCP). This case is of course an idealization, especially with the actual state of art. It is indeed absolutely unfeasible to have a single molecule FET with a gate width of few Å, far below 1 nm. By the way it is reported to further underline the potentiality in terms of scalability of molecular devices.

Technology	Area (nm^2)
MOSFET	1817.83
OPV7-MolFET	9.89
PCP-MolFET	10.14

Table 6.1: Estimated consumed area for the overall number of transistors in FA circuit (42).

6.4.2 Power dissipation

The aim of this subsection is to characterize the FA in terms of power consumption for both technologies. First of all a general remark which holds for every implemented architecture with whatever technology is needed to get a common lexicon. There are mainly two different kinds of dissipated power that must be distinguished in:

– **Static power**: is that kind of power dissipated in stationary regime, useful to feed all electrical components in order to let the circuit work properly, thus due to bias currents. In particular the static dissipated power includes the dissipated power when transistors are kept in the ON state, but also the dissipated power when transistors are kept in the OFF state. The latter contribution is called «leakage power» since it is due to leakage currents that occur when the device is operating in sub-threshold region. Static power contribution is a measure of the dissipated power to keep a given network state (for a given amount of time). Generally electronic circuits change their state during the

computation, and a certain amount of energy should be dissipated to perform the commutations (this is the dynamic power - below).

– **Dynamic power**: is the power dissipated during functioning, i.e. when transistors commute in order to change the network state. Dynamic power is mainly due to two contributions:

  - consumed energy in **charging/discharging** transistor parasitic capacitances in order to turn ON/OFF the transistor itself. A certain time τ_{sw} is required to perform this operation. The more is the exchanged (dissipated) energy in the unit time, the smaller will be this time interval, thus the faster will be the transient device response. Anyway the dynamic power will be greater.

  - **short circuit power:** is the contribution of power dissipated during the commutation transient, when a direct connection between the supply V_{dd} and ground is created due to the contemporary activation of the pull-up and pull-down networks. Notice that this contribution is not present in our architecture since pull-up networks are absent.

Once clarified which kind of power can be dissipated in a given topology,the sources of power dissipation must be identified in this case. Static CMOS logic based circuit operates in static regime. It was exactly coinceived to dissipate very few static power thanks to PU and PD network. Anyway the majority of the power consumption occurs during transient, when the toggle from one state to the other one is in progress. In this transient both of them are ON, thus a direct path from the supply to ground exists. Therefore in order to characterize the circuit in terms of power the static and transient regimes must be both analyzed. Notice that in all the cases, the power values at stack are avarage power values.

Concerning MolFET, for the static regime, the above mentioned LUT model can be used to estimate the static dissipated power, whereas in the transient regime the dynamic one must be adopted (LUT+RC) in order to estimate the dynamic power. Since the powers at stake in the two regimes are separable (and thanks to linearity the total dissipated power is the sum of the two), it is much easier to analyze them separately.
Notice that, to keep things as simple as possible, lines were supposed to be ideal, thus the power contribution due to their parasitics was not considered. By the way, in sight of a MOSFET vs molFET, line dissipated power is not of essential relevance. Indeed, by supposing to have quite similar interconnections in the two cases (a part of the metal 1 contacts level whose power contribution we suppose to be negligible), the line dissipated power would be essentially the same in both the technologies (additive constant contribution); and thus it would not strongly affect the analysis.

DC regime
In stationary conditions it is possible to evaluate the static power contribution. The only contribution of static power dissipation considered in this analysis is the one due to transistors, when they OFF (leakage: P_{leak}). Since the circuit considered is a static CMOS logic based circuit so based on PU and PD networks, this analysis is really simple:

- the current flowing toward ground when the output of FA is at logic 1 is the one related to leakage I_{leak}. By multiplying I_{leak} by the voltage at the output node (that is roughly the one dropping on the PD network that is OFF) the leakage power in the case of output equal 1, is obtained.
- the current flowing from the supply when the output of FA is at logic 0 is the one related to leakage I_{leak}. By multiplying I_{leak} by the voltage drop between the supply and the output (that is roughly the one dropping on the PU network that is OFF) the leakage power in the case of output equal 0, is obtained.
- since there are many inputs configurations for the FA that leads to logic 1 or 0 at the output, the total static power dissipated that is a total leakage power is obtained avaraging all the case, thus multiplying each different contribution of leakage power for 1/8 (since there are 8 different configurations of the input, assumed equiprobable) and finally diving by 8 (number of different cases).

In table 6.2 the resulted leakage power dissipation contributions are reported for each technology: MOSFET and PCP -based. The implementation with OPV7-MolFETs is not considered since it does not pass the functional test, and therefore since the measured values would be not the correct ones to properly make the circuit working, it would be useless to analyze it.

Technology	$P_{leak}\|_S$ (nW)	$P_{leak}\|_{C_{out}}$ (nW)	P_{leak}^{TOT} (nW)
MOSFET-based FA	4.801	2.804	7.605
PCP-MolFET	6.866	7.440	14.306

Table 6.2: Estimated avarage leakage power for the FA circuit. It is estimated for the two available outputs (S and C_{out}), and the total one is estimated as the sum of thes two.

Transient regime

The transient regime analysis is useful for dynamic power estimation, that was performed by a *Matlab* script reported in appendix, section B.2. The dynamic power, as already mentioned, is the energy (per unit time) dissipated for charging and discharging electrostatic capacitances of the device. It was evaluated per pulse, i.e. for a single charge plus a single discharge of stray capacitances, for a single transistor. Then the dynamic power per pulse was multiplied by the number of transistors excited by the pulse in the circuit, to get the total dynamic power per pulse, for the whole FA.
In order to evaluate the power (i.e. energy per unit time) it is necessary to know how long the commutation takes, i.e. the time in which the energy $E = \frac{1}{2} \cdot C \cdot V^2$ is transferred toward/from the capacitances to charge/discharge them. The pulse rising and falling transient time length were set to 0.5 ms.

These considerations hold true also for dynamic power analysis in molFET-based FA. In general the total exchanged energy per commutation (for a single charging or discharging phenomenon) is the sum of the effects on C_{gs}, C_{ds}, C_{gd}:

$$E = \frac{1}{2} C_{gs} V_{gs}^2 + \frac{1}{2} C_{ds} V_{ds}^2 + \frac{1}{2} C_{gd} (V_{ds} - V_{gs})^2 \tag{6.1}$$

Anyway, concerning MOSFET-based FA, the dynamic power dissipated per pulse per single transistor was computed only taking into account the effect of the gate capacitance Cgs (capacitance whose value was computed as a parallel plate capacitance by means of $DTOX$, TOX_E, TOX_P extracted from SPECTRE model file *.scs*), mainly because it was quite tricky to derive exactly the values of C_{gd}, C_{ds} from the parameters file *.scs*. However, for the resolved purposes, which mainly consist in comparing MOSFET-based and molFET-based FA performances, it is not so important to know the exchanged energy per commutation due to these capacitances, since the dynamic power dissipated to charge/discharge C_{gs} of MOSFET, that is a partial contribution, is already much greater w.r.t the total dynamic one in case of molFET. Therefore MOSFET-based FA surely shows worse performance than the molFET one in terms of dynamic power, when all contributions are considered. Moreover for the application which this architecture is aimed for, it is not so significant to derive those contributions since the total power, that is the sum of the static and dynamic one, is greatly dominated by the static one (of the order of nW). The dynamic one is indeed the dynamic power per pulse (of the order of or pW) multiplied by the commutation frequency of the circuit, that could be at maximum of order of GHz.

Concerning MolFETs, differently from the MOSFET case, the total exchanged

energy per commutation was computed taking into account the effect of all the electrostatic capacitances C_g, C_s, C_d (exploiting the average values Cs_AV, Cd_AV) that were computed by means of *ad-hoc MatLab* scripts reported in appendix, section B.1. Thus with reference to the molFET capacitive model in figure 2.24, the dissipated energy per commutation was evaluated as:

$$E = \frac{1}{2}C_g V_{cg}^2 + \frac{1}{2}C_s V_{dot}^2 + \frac{1}{2}C_d(V_{ds} - V_{dot})^2 \tag{6.2}$$

where $V_{dot} = \frac{C_d}{(C_s+C_d)}V_{ds}$ is the voltage drop across C_s (C_g is not considered in this capacitive divider since it is assumed that V_{ds} acts on channel only with no coupling to the gate), and where $V_{cg} = \Delta V_{gs} - V_{dot}$ is the voltage drop across C_g. Notice that $\Delta V_{gs} = |V_{gsON} - V_{gsOFF}|$, i.e. it is the difference between the gate to source voltage that is applied to switch ON the molFET and the value to switch it OFF. The average (over all considered bias points) drain and source capacitances, i.e. Cs_AV and Cd_AV, were used.
Obviously the quantum capacitance is not taken into account, indeed it is not an electrostatic capacitance to be charged and discharged during commutation, but it is useful in correctly estimate C_s and C_d. Notice that, since C_s is assumed equal to C_d a simplification occurs and $V_{dot} = \frac{1}{2}V_{ds}$.

In table 6.3 the resulted dynamic power dissipation contributions are reported for each technology: MOSFET and PCP -based. The implementation with OPV7-MolFETs is not considered since it does not pass the functional test, and therefore since the measured values would be not the correct ones to properly make the circuit working, it would be useless to analyze it. Notice that the considered case in this comparison is the worst one in which all transistors commute simultaneously, i.e. when all 42 transistors of the FA commute simultaneously.

Technology	$P_{dyn}\|_S$ (pW)	$P_{dyn}\|_{C_{out}}$ (pW)	$P_{dyn_{TOT}}$ (pW)
MOSFET-based FA	59.87	41.78	110.65
PCP-MolFET	6.91	6.47	13.376

Table 6.3: Estimated avarage dynamic power per pulse for the FA circuit in the worst case. It is estimated for the two available outputs (S and C_{out}), and the total one is estimated as the sum of these two.

6.4.3 Speed

In order to have an idea of the whole performances of a given technology, it can be useful to estimate, even roughly, a maximum possible frequency of operation. There exists an inferior limit to pulse duration, and thus a superior limit to the maximum frequency of operation, intended as the maximum number of spikes in a second.
In order to estimate it, a significant test was performed, all based (and derived from) the concept of **intrinsic time**. The MOSFET technological node intrinsic time is defined as the time needed to fully discharge a load capacitance present at drain node, with the initial condition that the capacitance is charged and the voltage drop across it is equal to the V_{DD} supply voltage. This load capacitance is a capacitance of value equal to a transistor gate capacitance, emulating the next stage input, plus the transistor output drain capacitance. Other assumptions are that the transistor width should be of 1 µm and the gate voltage should be set equal to supply voltage V_{DD}.
To the end of testing the FA circuit in a significant way, a sort of modified intrinsic time as figure of merit was used. In particular the transient duration of FA was estimated in the case with an output load capacitance of value equal to the FA input capacitance (since it is reasonable to have a situation where many FA circuits are connected in cascade), initially charged such that the voltage drop across it was equal to V_{DD}.
The exponential transient duration is theoretically infinite, but usually it is practically assumed to be $3 \div 7$ τ long, where τ is the time constant. The transient duration was assumed to be $7 \cdot \tau$ long, and since in a pulse there are two transients (rising/falling edges) a minimum pulse duration (to ensure correct behaviour of the FA circuit) was assumed of $14 \cdot \tau$. Consequently the maximum operating frequency (corresponding to the inverse of the minimum pulse duration) was assumed to be the maximum frequency of operation with no time interval between two consecutive pulses.
The simulations were again performed in *Cadence Virtuoso®* environment and the collected data were then post-processed in *MatLab* to estimate the value of the time constant and thus of the maximum operating frequency. The full *MatLab* scripts are reported in appendix, section B.3. Notice that the operative definition (widely used in circuit theory) of time constant was exploited to find its value: the time constant was assumed to be the time interval between the transient initial time instant (for us $t = 0$) and the time instant at which the straight line, that interpolates the exponential decaying curve during its initial phase, intersects the time axis (abscissa axis). Then starting from this time constant value τ the whole pulse duration and the maximum operating frequency were found as already explained.

Of course the above reported analysis is not at all significant in order to have a realistic estimation of the commutation time and of the maximum operating frequency for a real FA circuit. In fact, in this analysis the stray capacitances and resistances of the lines were completely neglected, even if they certainly strongly affect the commutation time. Thus the results reported in table **??** should not be considered as a refined estimation of the speed performances of the FA, but more like a useful starting point in sight of a speed comparison between MOSFET and molFET technologies. Indeed the estimated time constants depend only on the device intrinsic capacitances and on the load. The same procedure was then performed also in the molFET case. Finally in order to compare the different technology-based FA in terms of speed, the maximum operating frequency achievable by the circuit were compared, derived from the transient evaluation in the case of a load equal to the equivalent input capacitance of FA. Results are shown in table **??**.

Technology	f_{max} (MHz)
MOSFET	2.871
PCP-MolFET	0.198

Table 6.4: Estimated maximum frequency comparison.

6.4.4 Conclusions

The effort spent on molecular device modeling and molecular-based circuit designing was justified by the expectation for widely better performances in terms of dissipated power and occupied area of molFET case if compared to MOSFET-based technology. Obviously is expected also a toll to pay in terms of speed, thus in maximum operating frequency (related also to the smaller currents of molecular devices w.r.t. MOSFET ones).
From the reported comparisons it is evident that migrating from conventional MOSFET technology to a molecular FET technology could be favourable depending on which molecule is exploited as channel. Using OPV7 molecule functional problems are relevant. This is due to extremely low I_{ON}/I_{OFF} ratio of order of 40.
Whereas, exploiting PCP molecule could lead to improvements if the MolFET is carefully enginereed. Indeed, functional problems are partially solved and thus speak of a comparison between MOSFET technology, in this specific application, makes sense.
Concerning the occupation of area, it is extremely evident the advantage of migrating from conventional MOSFET technology to a molecular FET one:

2 order of magnitude in reduction can be achieved. As expected the molecular FETs technology is well promising in terms of scalability. The situation is worst concerning the dissipated power. The leakage power turns out to be almost doubled for MolFET, due to higher leakage currents in that operating points for that specific circuit. Whereas, due to lower I_ON and load capacitance values the dissipated dynamic power per pulse is reduced of about one order of magnitude in the case of MolFET technology. However the main drawback of MolFET technology in this application, is the maximum operating frequency achievable, that is reduced more than one order of magnitude, due obviously to the smaller ON current, thus to the smaller driving force, even if the load capacitances at stakes are 3 order of magnitude less. Final remarks should be pointed out for what is concerning a possible «real world» implementation. The main drawback of the PCP-molFET is the small ON current. In a real implementation this would strongly affect and penalize the maximum operating frequency. Indeed, with the actual state of art, the interconnections cannot be though to be much smaller than the ones used with actual MOSFET technologies. A such tiny ON current should thus be used to charge very big line capacitances (w.r.t. device ones), in order to actuate a complete commutation of state. This would require a much longer time, thus leading to a much lower maximum operating frequency. The above reported comparison would be reasonable if we suppose to be able to effectively scale interconnections in order to avoid this issue without impacting too much on the parasitic resistances deriving from scaled cross section of the interconnections (especially the local ones), thus acting for example on low resistive innovative materials or on innovative interconnection processes. Indeed the lines time constant does not depends on the parasitic capacitances only, but on the RC product.
Nevertheless, in summary, it has been shown that there exist advantages in moving from a conventional MOSFET technology toward molFET ones, especially in terms of area and in dynamic power. On the other hand there also drawbacks, and above all the biggest issue to deal with would likely be the maximum operating frequency of molFET.

Chapter 7

Memory implementation: SRAM

The final aim of this chapter is to design, test and characterize a molecular technology-based SRAM cell. To achieve this goal, a methodology based on the following steps is adopted: (A.) firstly a review on working principles of SRAM is done in section 7.1, mainly in order to outline the key points for a correct and effective functioning of its single basic cell. In this step the design constraints of the single cell are theoretically clarified; (B.) a design procedure of the popular CMOS 6T SRAM cell is performed in Cadence Virtuoso® exploiting a quite recent technology, the FD-SOI MOSFET (28 nm) one. This step actually is not so strictly necessary to achieve the goal of the chapter but is performed in order to acquire a more practical understanding on what are the effects on some important figures of merit of the cell when the design constraints outlined in the previous step are satisfied or partially are not. This would be very helpful in the next step for interpret the simulation results in the case of the molFET implementation of the cell. Indeed, in this case, due to molFET behavioural deviation from standard MOSFET an due to the difficulties in satisfying the 6T cell design constraints, simulation results are expected to be really different from the ones obtained for MOSFET implementation. So, in subsections 7.2.1, 7.2.2, the MOSFET-based 6T SRAM cell is respectively designed, and verified (C.) in subsections **??**, **??**, the design and testing phases of the previous step are repeated for molFET technology. In particular only two molecular transistors are exploited for the 6T SRAM implementation, the OPV7 and PCP ambipolar transistors enginereed and characterized in part II of the thesis; Differently from the previous chapter, no analysis of power, speed and area are provide: Indeed since the molecular

implementation suffers of functionality issue it would be useless compare the two technologie.

7.1 SRAM review and metrics

The SRAM is a volatile random access type of memory which still plays an important role in VLSI circuits. The term static refers to the fact that no periodic refresh mechanism is necessary as in DRAM. SRAM is faster but more expensive than DRAM, that is why is typically used for CPU caches while DRAM is used for the main memory. Moreover it allows an effective integration with logic circuits, because of its high compatibility of process and supply voltage. One of the most popular SRAM cell topologies is the *6T* cell shown in figure 7.1, made by six transistors, two of which are pass-transistors that enable the access to the actual cell that retains the data stored. The others form two CMOS back-to-back inverters, also known as *bistable ring*. The bistable ring is able to retain the stored information as long as the cell is supplied, differently from DRAM, where a refresh cycle is needed to allow the data retention. It can be seen also as a latch that in order to work as a storage cell it must work in stable points to avoid metastability issues. In figure 7.2b the transcharacteristic of the bistable ring highlights the presence of three equilibrium points, but only two of them are stable. Indeed, the red dot in the transcharacteristic is unstable since a small variation of the input leads to a change in the output state. In other words, the bistable ring is able to store data far away from this point, thus for variations of input voltage lower than $\frac{V_{cc}}{2}$, with V_{CC} the supply voltage. For sake of completeness a typical 1D memory array is also reported in figure 7.3, where the 6T SRAM cells are arranged in an 1D array of rows and columns: vertically the cells are connected to the bitlines, horizontally they all share a single line, the *wordline*. Obviously, more advanced architectures (2D, 3D) not shown here, are possible and widely exploited for enhance performances and reduce static power consumption. The SRAM cell has three main tasks:

- HOLD: retain the data stored, i.e. when the wordline is not asserted and the bit line can assume any value.
- WRITE: write the data loaded on the bit line into the cell, by asserting the wordline and thus making possible the data on the bitline flow into the cell. This operation requires a greater driving electrical strength w.r.t. the one of the cell, that is beacause the operation of writing is said to occur by means of conflict. The driving force is provided by two buffers on the bitlines that are activated during this operation.

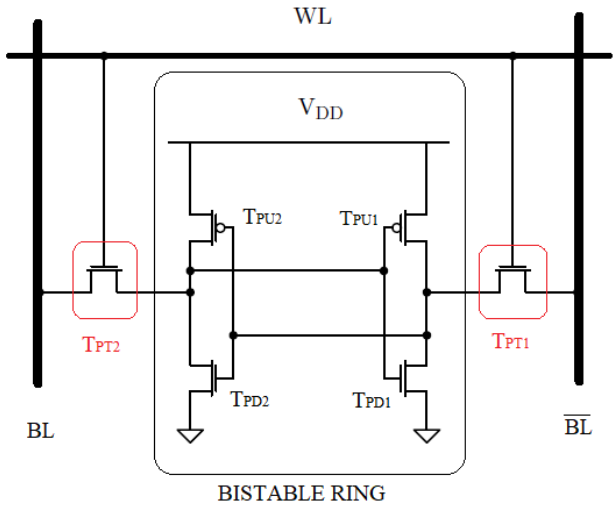
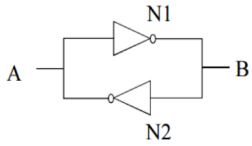


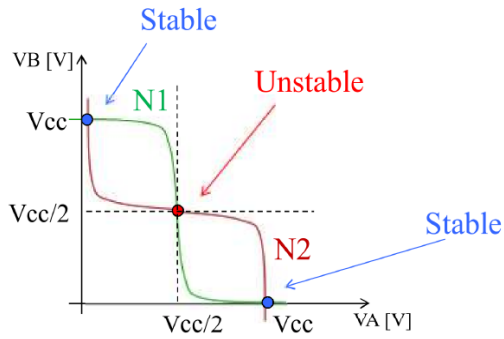
Figure 7.1: A well known 6T SRAM cell. The wider lines are the bit lines (straight and complemented BL, $\overline{BL}$) where data travel; whereas the thinner line is the word line that is activated only if the cell is selected for reading or writing operation. When WL is activated the pass-transistors are switched ON in order to let the bit stored to be read or overwritten with a new data coming from the bit lines. The bistable ring is made by two cross-coupled inverters thus by two pull-up p-MOSFETs and two pull-down n-MOSFETs.

– READ: read the data from the cell, by asserting the wordline and letting the bitlines to sense the data, i.e. precharging and letting them floating. During this operation bitlines are no more externally driven by buffers, but precharged at a weak logic value '1' (typically $V_{DD}/2$) in such a way that the stored data in the cell "wins" w.r.t. the value of the precharged bitlines without conflict.

To evaluate the quality and the robustness of these three operative conditions, some metrics are identified. A useful metric for all the three operations is the *Static Noise Margin* (SNM) a parameter which quantifies the noise immunity of the cell from noise during the different operations (hold, write, read). It practically gives information on how much noise can be tolerated without making the content stored on the cell flip. And therefore is defined as the maximum noise that can be tolerated in order to still achieve a correct retention, writing, reading of data. Moreover, during the hold phase another



(a) Latch as core of 6T SRAM cell.



(b) Transcharacteristic of the bistable ring. It is the result of the superposition of the two inverters transcharacteristics,

Figure 7.2

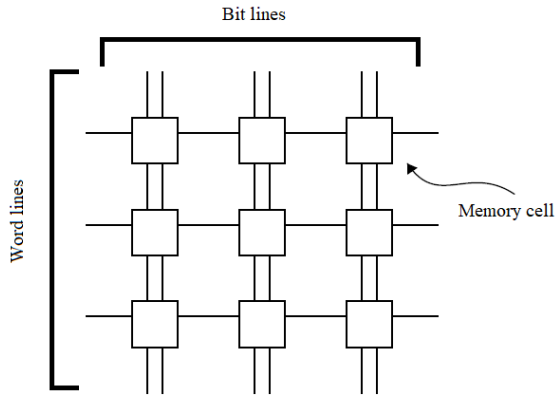


Figure 7.3: A typical in principle schematic of 1D memory array. Each square is a 6T memory cell. Additional circuitries like address decoder or I/O circuitry are omitted.

figure of merit is usually taken into account, the *Data Retention Voltage* (DVR) which gives information about the minimum cell supply voltage which allows that the stored data in the cell does not change. Indeed, when the memory cell is not selected, its pass-transistors are disabled and the cell is said to be in "data retention mode". Even if the cell is not working, it is

not possible to lower the cell supply voltage (for e.g. reduce static power consumption due to leakages), otherwise the bistable ring is no more able to retain the stored data and consequntly looses the information. Thus, a lower bound of the cell supply voltage, the DVR, must be ensure to correctly retain the data.

Obviously, the way the SNM is related to the different memory operations is different and it is presented in the following subsections.

7.1.1 Read operation

A read operation is performed in two steps:

1. bitlines (BL) are pre-charged to the logic '1', which typically corresponds to the supply voltage *VDD*.
2. the cell to be read is selected by asserting the wordline (WL), enabling, in this way, the pass-transistors of the cell. As a consequence, one of the two bitlines gets discharged by the cell (i.e. the one connected to the inverter storing a logic '0'.)

The read operation is critical in this kind of memories, especiallly because sometimes drives to a condition of metastability. Let's see briefly why and how to avoid this issue. Let's refer to figure 7.4 where the pull-down transistor (T_{PD}) is ON (i.e. a logic '0' is stored). When the pass-transistor is activated by asserting the wordline, the bitline is pulled down to ground. Anyway, since the bitline is a very long metal line crossing all the height of the memory array, it has a very high parassitic capacitance C_{BL}. As a consequence, during this operation, a charge-sharing phenomenon occurs: some charge is transferred from the bitline to the storage node Q, that actually is not pulled to ground voltage as it should be. The voltage drop across Q (V_Q) depends on the voltage divider between the pass and pull-down transistors:

$$V_Q = \frac{R_{PD}}{R_{PD} + R_{PT}} V_{BL} \tag{7.1}$$

Since the output of the corresponding inverter $N1$ is connected to the input of the other one $N2$, it may happen that V_Q rises above the the threshold voltage of the pull-down transistor of the inverter $N2$ (T_{PD2}) letting it switching ON and as a consequence pulling down the input gate signal of T_{PU1}, making in turn, as final result, the content of the cell flip from '0' to '1'.

In other words, the read operation may lead to content overwrite, making impossible to read the original value. This situation is depicted in figure 7.5 and compared with the correct behaviour of the cell during a read operation

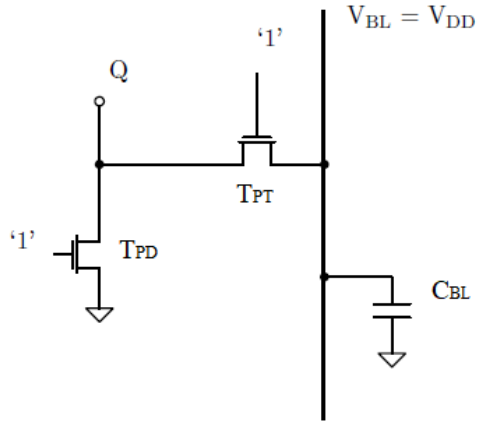


Figure 7.4: C_{BL} is the sum of the parasitic access capacitance of the cells connected to the line (i.e. the drain capacitance of the pass transistor) and the parasitic capacitance of the metal line.

in figure 7.6, where V_Q does not reach the inverter trip-point and, so, the data is read correctly (i.e. the bitline is discharged to the logic '0'), without corrupting it.

To avoid this phenomenon, V_Q must be taken under control and it can be proved that this can be done by ensuring a certain condition on the ON resistances of pull-down and pass transistors: $R_{PD} < R_{PT}$, which practically means making a proper transistor sizing in such a way that, once chosen a specific technological node (i.e. the channel length of the transistors), the channel width of the pull-down transistor is larger than the one of the pass-transistor:

$$W_{PD} > W_{PT} \tag{7.2}$$

This constraint, obviously leads to a cell area increment.

7.1.2 Write Operation

The write operation is simply performed asserting the wordline in order to select the cell and by putting the data to be written on the bitlines. Also for perform a successful writing operation a constraint in transistors sizing must be ensured, as for reading operation.

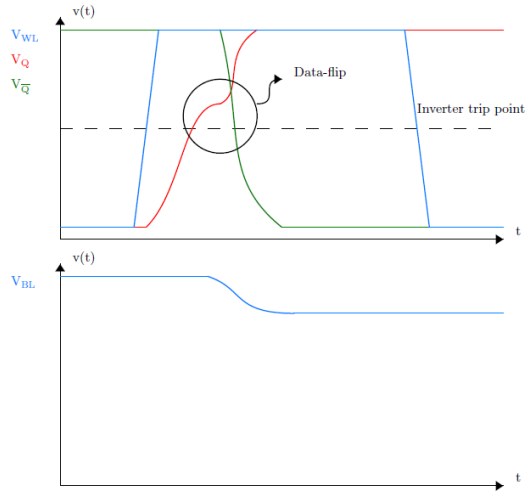


Figure 7.5: Data flip of 6T SRAM. The signal V_{WL} is the voltage applied to the WL (i.e. to the gate of pass transistors), V_Q the above mentioned voltage on Q node, and $\overline{V_Q}$ the complemented one (i.e. the voltage on the opposite inverter $N2$). V_{BL} is the voltage signal on the bitline connected to the node Q through the pass transistor T_{PT1}.

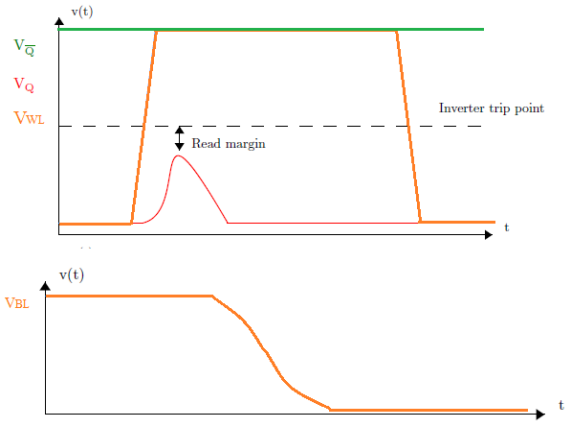


Figure 7.6: No data flip of 6T SRAM. The bitline V_{BL} is discharged to the logic '0'.

It has been stated, for reading operation, that sizing constraint on pull-down and pass transistors ensures V_Q to be taken under control. Unfortunately if a logic '1' is applied to the bitline in order be written in the cell, this constraint creates a problem because V_Q cannot be increased through the pass-transistor T_{PT}. Hence, a logic '1' cannot be forced on the inverter on which a logic '0' is stored. To overcome this problem, only by forcing a logic '0' on the inverter on which a logic '1' is stored, the cell content can be overwritten. This possibility is achieved in the differential structure of the cell, where two bitlines, one straight and the other complemented, are connected to different cell sides.

Anyway also in this case (figure 7.7) a voltage divider is present between the pull-up and pass transistors. Since the pull-up is connected to V_{DD} while the pass transistor is connected to ground, the voltage at the node Q is:

$$V_Q = \frac{R_{PT}}{R_{PU} + R_{PT}} V_{DD} \tag{7.3}$$

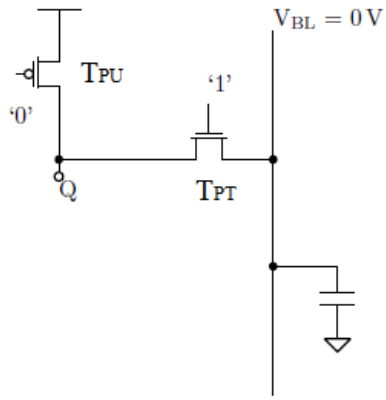


Figure 7.7: Voltage divider between the pull-up and pass transistors.

Hence, V_Q results to be larger than 0V, and it may turn on the pull-down transistor of the other side making possible again the data flip, and not performing a successful writing operation of the data (figure 7.8).

To avoid this phenomenon, it has to be ensured that: $R_{PT} < R_{PU}$. This practically means making a proper transistor sizing in such a way that, once chosen a specific technological node (i.e. the channel length of the transistors), the channel width of the pull-up transistor is larger than the one of the

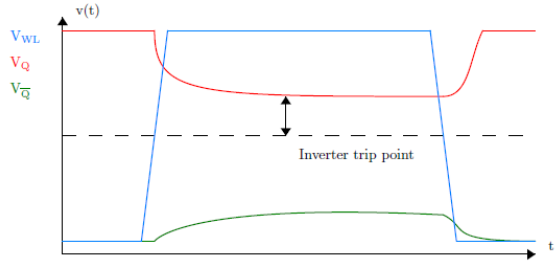


Figure 7.8: Data flip in writing operation. If V_Q does not lower below the bistable ring trip-point, the data is not written to the cell.

pass-transistor:

$$W_{PT} > W_{PU} \tag{7.4}$$

By satisfying this sizing constraint, a successful write operation can be performed, as it is shown in figure 7.9.

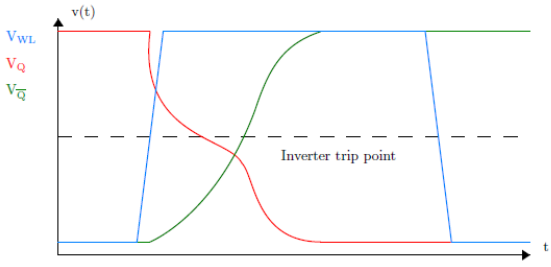


Figure 7.9: Correct writing operation.

7.1.3 Cell robustness

Therefore in conclusion, in order to ensure a certain reading and writing stability, the pass transistor should be sized with an aspect ratio larger than the pull-up and the pull-down with an aspect ratio larger than the pass transistor. Moreover the aspect ratio of the latter should be large enough to provide a minimum value of the cell current to charge/discharge properly the bit lines. Moreover process variations affects variations on the threshold voltage of the transistors affecting in this way the read and write margins of the cell. For example, if the pull-down is designed with an aspect ratio

larger than the pass- transistor one but the latter has a lower V_{th} because of the process variations, the readmargin obtained is lower than the desired one; it may even happen that the cell is unable of performing a correct read operation. The analogous situation might happen for writing operation. Anyway if the SNM of the cell is not enough alternatives topologies such as 8T or 10T cells can enhance it.

7.2 MOSFET-based 6T SRAM

In this section the 6T SRAM cell is implemented with a quite recent technology, exploiting ST FD-SOI 28nm technological library in Cadence Virtuoso. In subsections 7.2.1, 7.2.2, the memory cell and surrounding driving circuits are respectively designed and verified Notice that in order to simplify the design and analysis phases, especially in the case of molecular technology implementation, no parassitics related to the interconnections are taken into account. Anyway, this neglect is not so important since only a single cell is analysed and the metal lines in this case are not at all long, thus don't affect too much the behaviour and performances of the cell.

7.2.1 Design

6T SRAM cell. From the previous section (7.1) it has been stated that in order to obtain a correct reading and writing operation of the cell some constraints on transistors sizing should be satisfied:the pass transistor should be sized with an aspect ratio larger than the pull-up and the pull-down with an aspect ratio larger than the pass transistor. In particular, once chosen a technological node to be exploited, i.e. once fixed the channel length of the transistor (L_{ch}), to modify the aspect ratio it is possible to act only on the channel width (W_{ch}). Hence, the sizing constraints translate into:

$$W_{PU} < W_{PT} < W_{PD} \tag{7.5}$$

The following design chosen values are reported in the table below 7.1: As shown in the figure 7.10 and in the Cadence schematic in figure 7.11: the ratio between pull-down and pass-transistor is chosen equal to 2, in order to avoid the overwriting of the cell content during a read operation, whereas the ratio between pass-transistor and pull-up is chosen equal to 2, in order to allow the write operation to be correctly performed. Moreover notice that,

Transistor	Aspect ratio (AR)	Width (nm)	Length (nm)
Pass-transistor T_{PT}	2	160	30
Pull-up T_{PU}	1	80	30
Pull-down T_{PD}	4	320	30

Table 7.1: Design chosen values for transistors in 6T SRAM cell. The minimum aspect ratio for this technology is $(W = L)_{min} = \frac{80\,\text{nm}}{30\,\text{nm}}$ and the aspect ratio of each transistor is considered as a multiple of this minimum aspect ratio.

since:

$$R_{MOS} = \frac{1}{\mu_{n/p} \cdot C_{OX} \cdot (AR) \cdot (V_{DD} - V_{TH_{n/p}})} \tag{7.6}$$

with $AR = W/L$ (aspect ratio) and:

$$\frac{(AR)_{PD}}{(AR)_{PT}} = \frac{(AR)_{PT}}{(AR)_{PU}} = 2 \tag{7.7}$$

it follows that:

$$\frac{R_{PD}}{R_{PT}} = 2 \qquad , \qquad \frac{R_{PT}}{R_{PU}} = 4 \tag{7.8}$$

meaning that the driving capabilities between the pull-up and the pass transistor is larger than the one between the pull-down and the pass transistor and as a consequence the write margin of the cell is larger than the read one.

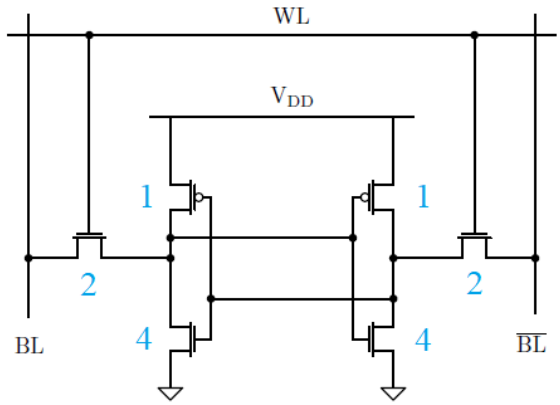


Figure 7.10: 6T SRAM cell schematic with the chosen transistor aspect ratios.

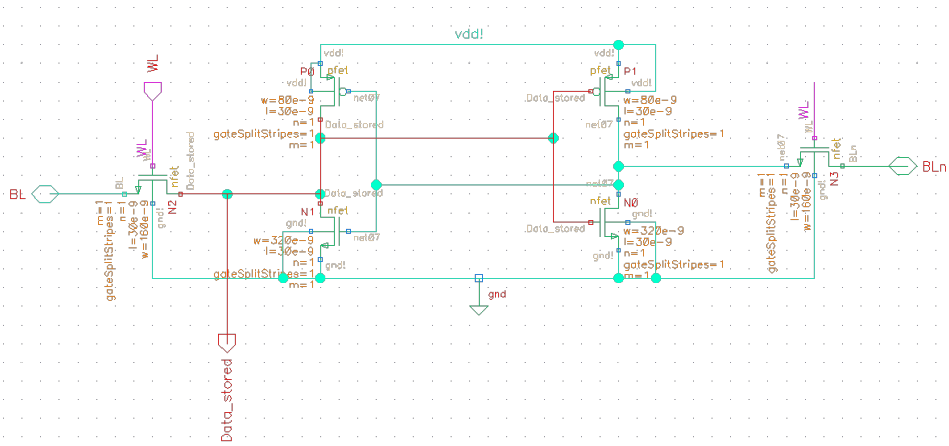


Figure 7.11: 6T SRAM cell Cadence schematic.

Driver circuits. During, for e.g., reading operation the bitlines are firstly precharged and then left floating. These conditions are not easy to be forced as stimuli at cell pins, therefore in order to test the memory cell in real conditions, also driver circuits must be designed and included. Notice that the inverter stages needed to get complemented signals are omitted in figures for sake of clarity. Moreover, since the aim of this chapter is to design, test and analyze the single cell, sense amplifier circuits, essential in an array of cells, are not considered in these phases.

– **Bitline drivers** In figures 7.12, 7.13 are shown the in-principle and Cadence schematics properly designed in terms of transistor sizing. These components drive the bitlines of the memory cell. The input pins are stimulated by ideal voltage sources in the Cadence Virtuoso testbench. Notice that the output section of the circuit is designed to be a tristate inverter. Indeed, when we want to drive the bitlines of the cell to make a write operation, the output stage of the driver is enabled ($EN = `1'$), thus connected to the bitlines. Instead during read operation, the bitlines must sense the data coming from the cell, thus need to be electrically separated from the drivers. This is done by de-asserting the enable signal ($EN = `0'$). Similarly during the precharge cycle, in order to be charged to V_{DD}.

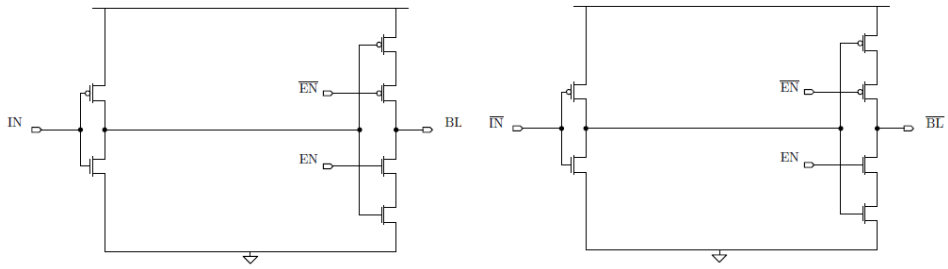


Figure 7.12: Schematic of bitline drivers.

Obviously, notice that in order to test an array of cells, a bitline driver for each couple of bitlines (i.e. for each memory column), needs to be instantiated. Anyway this is out of scope of this thesis.

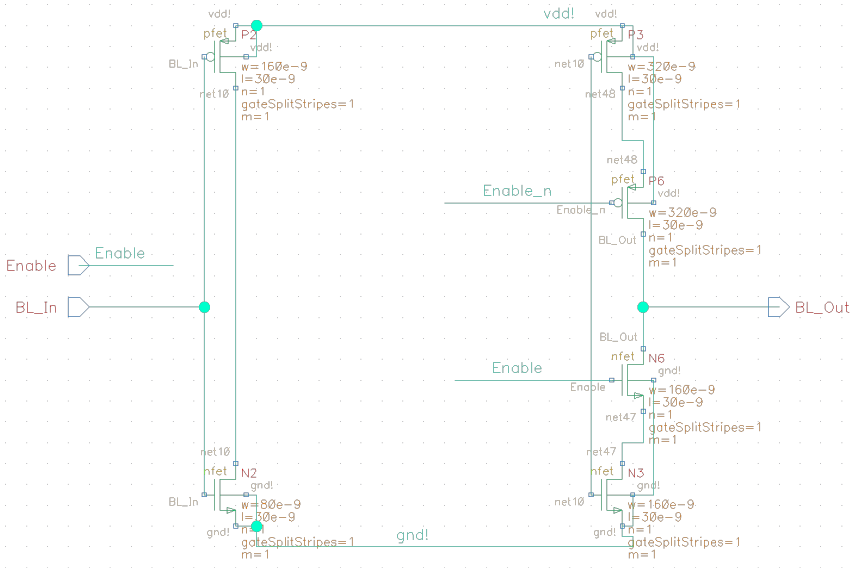


Figure 7.13: Cadence schematic of straight bitline driver.

– **Wordline drivers.** In figures 7.14, 7.15 are shown the in-principle and Cadence schematics properly designed in terms of transistor sizing of the wordline driver. It is simply made of inverter stages, since there is no need to electrically isolate the driver from the wordline. Indeed the wordline voltage has to be always well defined at each clock cycle, in such a way that the cells can be isolated from the bitlines when not selected (i.e. when the wordline is de-asserted).

  Actually this circuit, for the aim of this chapter focused on the design and test of only a single cell, can be omitted and the wordline signal can be simply defined as stimulus at each clock cycle in the input pin WL of the cell. Anyway it is included for sake of completeness.

– **Precharge circuit.** In figures 7.16, 7.17 are shown the in-principle and Cadence schematics properly designed in terms of transistor sizing of the precharge circuit. This component is exploited to precharge to the supply voltage the bitlines before a read cycle. When the precharge signal is de-asserted (i.e. $\overline{PRE} = \text{`0'}$), the bitlines are connected to $V_{DD}$ through the two side pMOS transistors and equalized by the central pMOS; whereas, when it is asserted ($\overline{PRE} = \text{`1'}$), all the transistors are

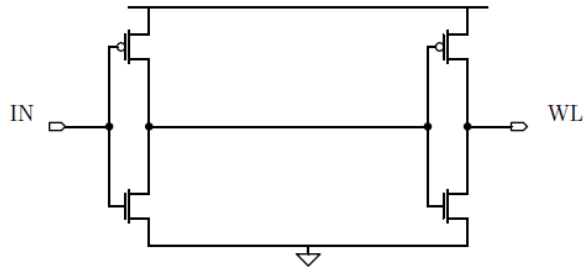


Figure 7.14: In-principle schematic of wordline driver.

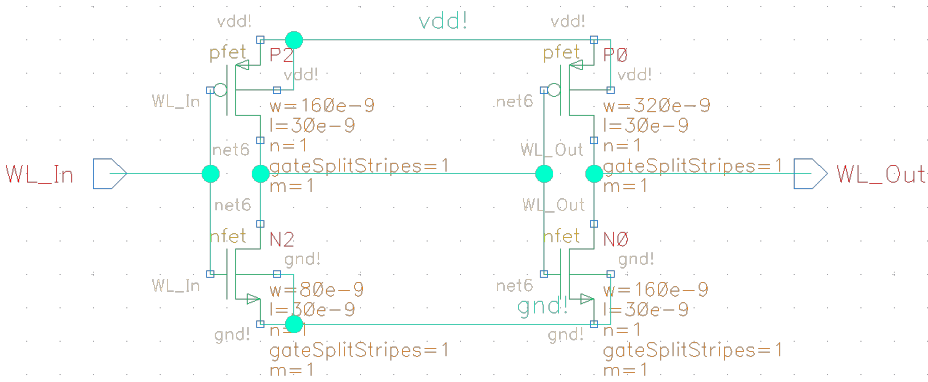


Figure 7.15: Cadence schematic of wordline driver.

disabled and the circuit is electrically isolated from the bitlines.

7.2.2 Testing

In order to test the single 6T SRAM cell, all the previous discussed components are firstly individually tested using the *SPECTRE* simulator embedded in Cadence Virtuoso. Simulation results and comments are reported in the following.

– **Bitline driver.** As expected, when the output stage of the driver is enabled (EN = '1'), the bitlines follow the input in which the data is forced. Instead during read operation, the bitlines must sense the data coming from the cell, thus need to be electrically separated from

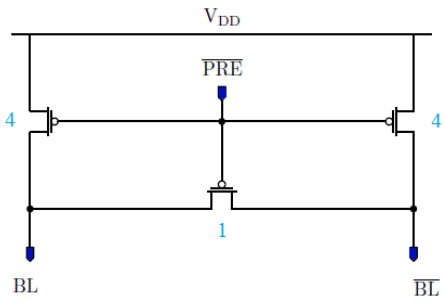


Figure 7.16: In-principle schematic of precharge circuit.

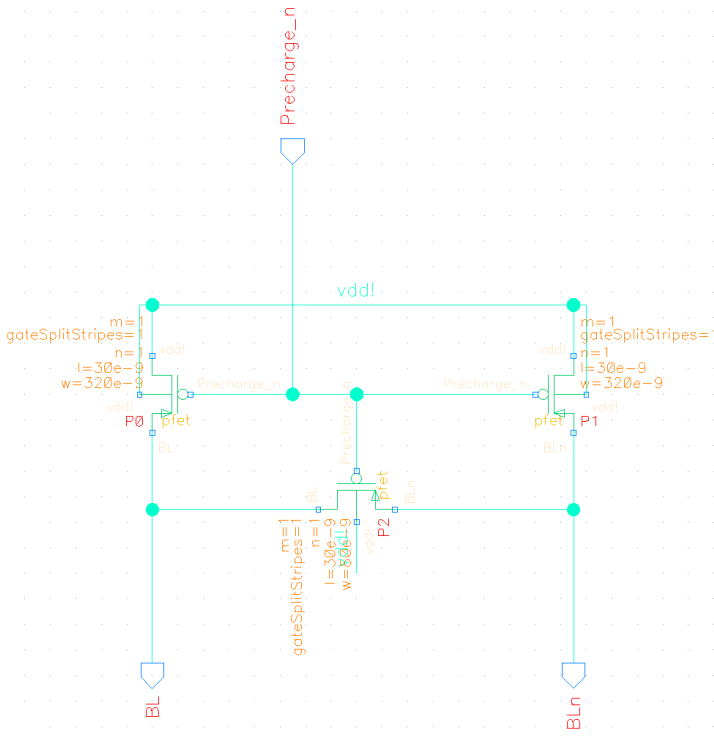


Figure 7.17: Cadence schematic of precharge circuit.

the drivers. This is done by de-asserting the enable signal (EN = '0'). Similarly during the precharge cycle, in order to be charged to V_{DD}.

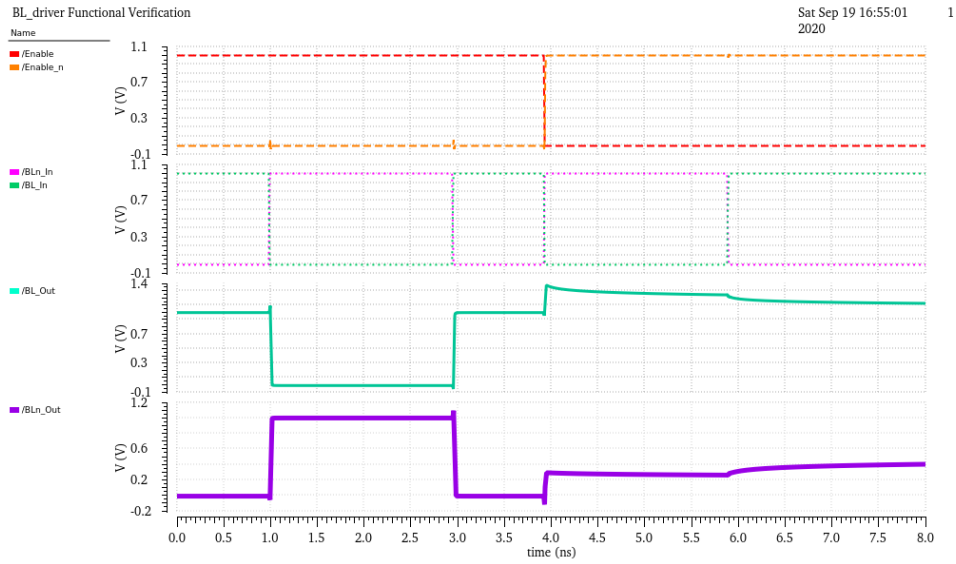


Figure 7.18: Spectre functional verification of bitline driver circuit implemented with MOSFET technology.

– **Wordline drivers.** As expected, when the input signal IN = '1' the wordline is asserted (i.e. goes to high logic value corresponding to V_{DD}). Whereas, when IN = '0' the wordline is de-asserted (i.e. goes to low logic value corresponding to ground) in such a way that the row cells corresponding to wordline in question, can be isolated from the bitlines when not selected.

– **Precharge circuit.** As expected, when $\overline{PRE}$ = '0'), the bitlines are precharged to $V_{DD} = 1V$; and the circuit is electrically isolated from the bitlines when $\overline{PRE}$ = '1').

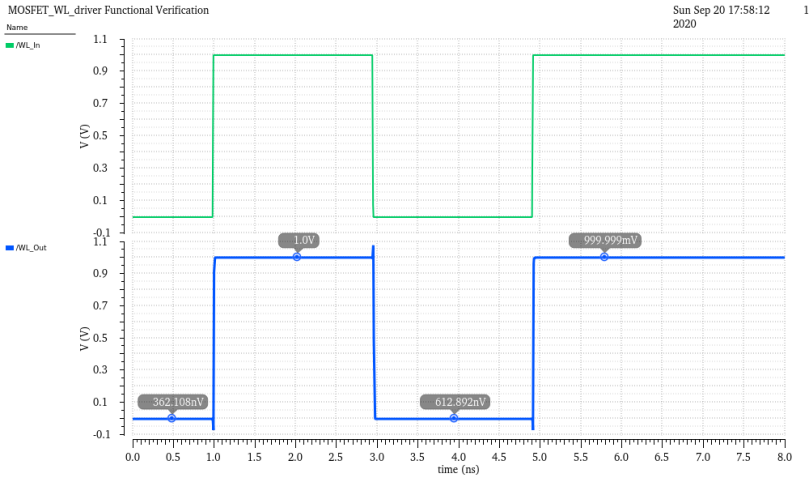


Figure 7.19: Spectre functional verification of wordline driver circuit implemented with MOSFET technology.

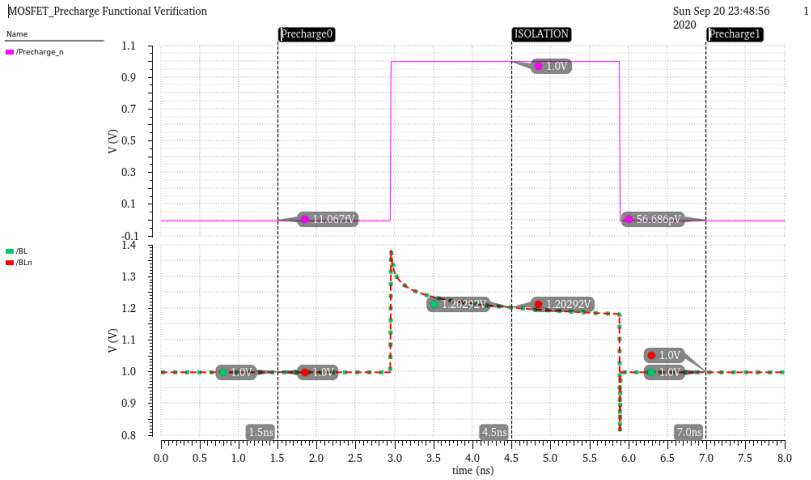


Figure 7.20: Spectre functional verification of precharge circuit implemented with MOSFET technology.

Once individually tested, they are properly connected all together in a unique RTL view schematic as shown in figure 7.22 and finally the SRAM cell is tested in the most significant conditions. Notice that sense amplifier circuits

are omitted since only a single cell is tested and thus they are not so strictly significant. A significant simulation waveform is reported in figure 7.22, where

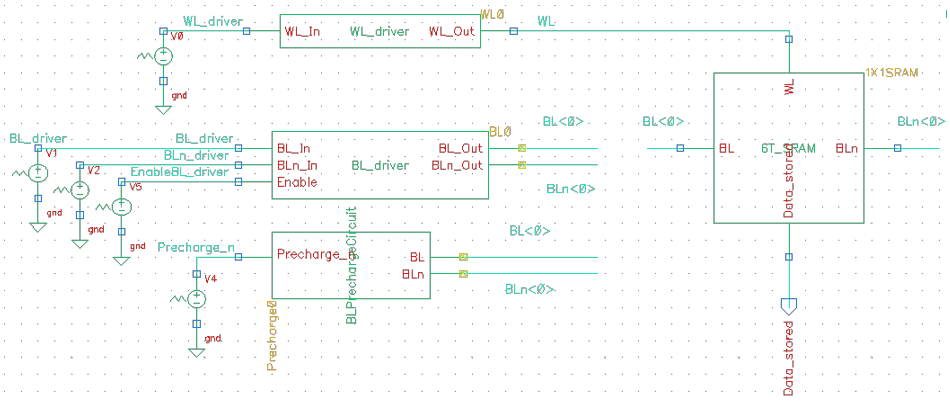


Figure 7.21: Cadence RTL schematic of a single cell of SRAM and surrounding driving circuitries.

all the possible states of the memory are shown. Firstly the EN_BL (pink line) and WL_IN (dash violet line) signals are not asserted allowing the bitlines and wordlines to not be enabled. Moreover, the $Precharge\_n$ signal is asserted so the cell is not in the precharge state, but is in idle state, thus the bitlines are floating and the data in the cell is undefined. Then the EN_BL signal is asserted and thus the bitlines are enabled and follow the input data BL_In, BLn_In forced thanks to the bitlines driver circuit, but only when also the wordline is enabled, thanks to the assertion of WL_IN signal, the cell becomes accessible moving from idle to the write phase. Since in the bitline a logic 1 is forced, this is correctly written in the cell with practically negligible reduction of the dynamic thanks to the high robustness of the designed cell. When the bitlines and wordlines are disabled (EN_BL=‘0’, WL_IN=‘0’), the cell is no more accesible and the bitlines are correctly precharged to V_{DD} by de-asserting the $Precharge\_n$ signal, as shown in the time interval from 4-6 ns. Notice that obviously in this phase, since the cell is no more accesible, the logic 1 written in the previous phase is kept stored in the cell, and thus is correctly read in the next cycle by enabling again the wordline (WL_IN=‘1’) and keeping disabled the bitlines driver circuit (EN_BL=‘0’). Notice that the read operation is not destructive, in the sense that the data is still kept stored in the cell, and the bitlines correctly gets charged/discharged.

In the next cycle the cell is driven again in idle and then a writing operation

is performed, this time of a low logic value (i.e. 0V). The data is correctly written, kept stored during the precharge phase, and correctly read. Notice that the fact the a low logic value stored in the cell is correctly read is due to the effort spent at design time in sizing properly the channel widths of the transistors according to the sizing constraints of 7.5.

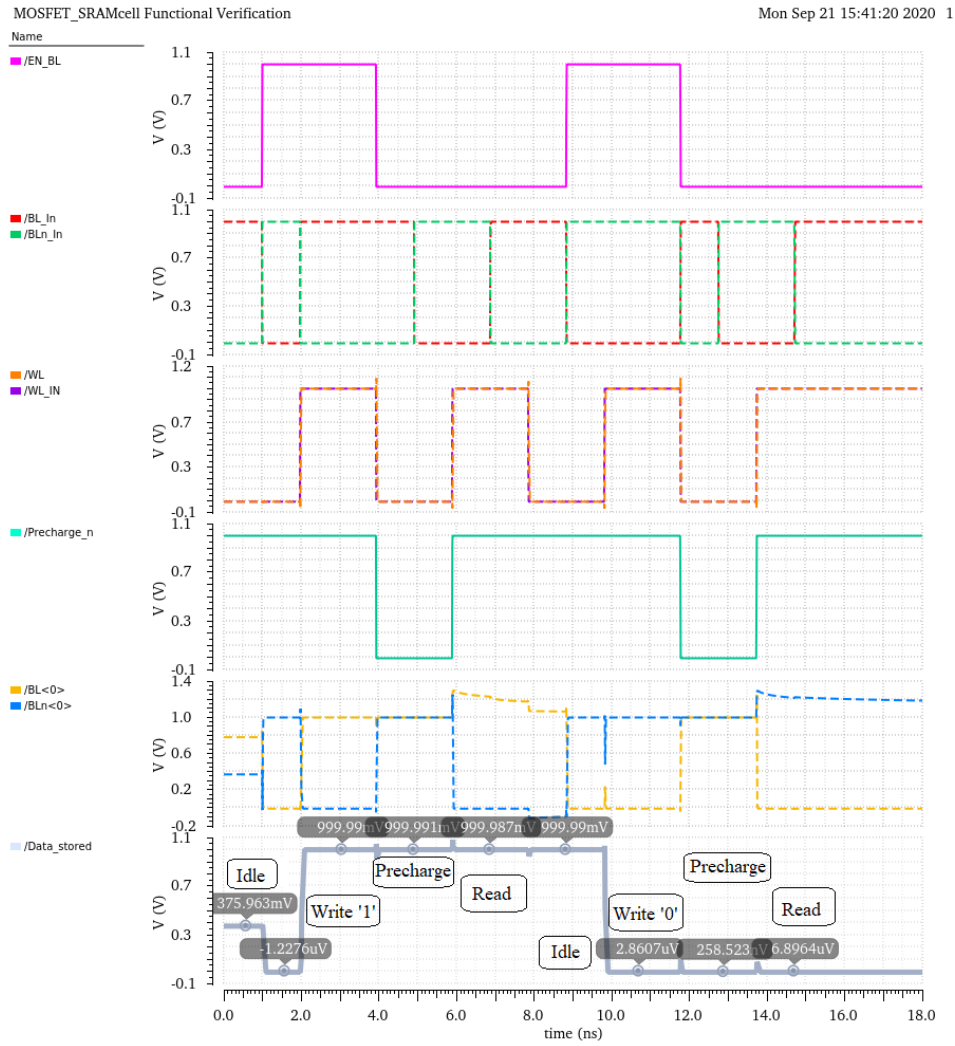


Figure 7.22: Spectre functional verification of a single cell of implemented with MOSFET technology.

7.3 MolFET-based 6T SRAM

7.3.1 Design and testing

The SRAM cell and driver circuits implemented in the previous section in MOSFET technology are now implemented in molecular technology, by simply replacing in each schematic MOSFETs with molecular transistors, and properly adjusting the supply and reference voltages, as done in the design of the logic elements in the previous chapter. So, in summary the high logic value was set equal to 0V, whereas the low logic value to -1V. V_{DD} was set equal to 0V, whereas V_{SS} equal to -1V.

Also in this case it is not possible to perform the transistor sizing procedure as done for MOSFETs-based circuits, due to the limitation in the adopted LUT model for circuits simulations. Thus the sizing constraints for a correct read and write operation cannot be satisfied, and therefore data flips are expected to occur. Moreover since the exploited molFET were engineered to work in static CMOS logic-based circuit and not in a pass-transistor configuration, the resulting SRAM cell will behave even worse than expected.

Also in this case, obviously, in order to test the complete SRAM cell, all the surrounding circutries are firstly individually tested. Then they are properly connected all together in a unique RTL view schematic and this tested in the most significant conditions. Notice that, also in this case, sense amplifier circuits are omitted since only a single cell is tested and thus they are not so strictly significant. Simulation results and comments are reported in the following.

– **Bitlines driver.** In figure 7.23 the behaviour of OPV7 molFET-based bitline driver is tested. It does not correctly behaves at all. First of all, due to the reduced noise margin of the OPV7-based inverter analyzed in the previous chapter, the complemented enable signal $Enable_n$ is very distant from the low logic value. This, as a consequence, already precludes a correct or almost relatively reliable functioning of the circuit, since the output stage cannot be properly disabled. But, even if we are able to correctly complement the *enable* signal, in any case the bitlines, as shown in figure 7.23 does not follow the inputs at all when the output stage of the driver is enabled (EN = '1'). In other words, this circuit does not clearly behave as a bitline driver.

  Different is the case for PCP-based bitlines driver, whose simulation results are shown in figure 7.25. It correctly behaves with acceptable noise margins. Indeed, first of all, in this case a suitable complemented enable signal is generated. Then, when the output stage of the driver

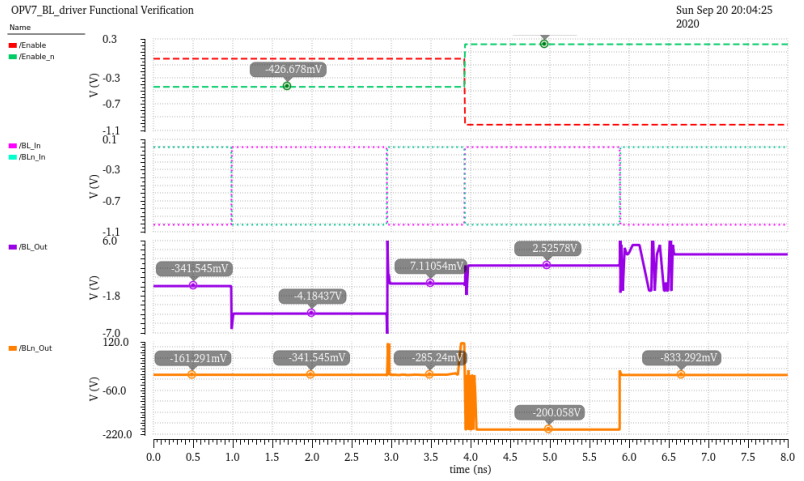


Figure 7.23: Spectre functional verification of bitline driver circuits implemented with OPV7-MolFET technology.

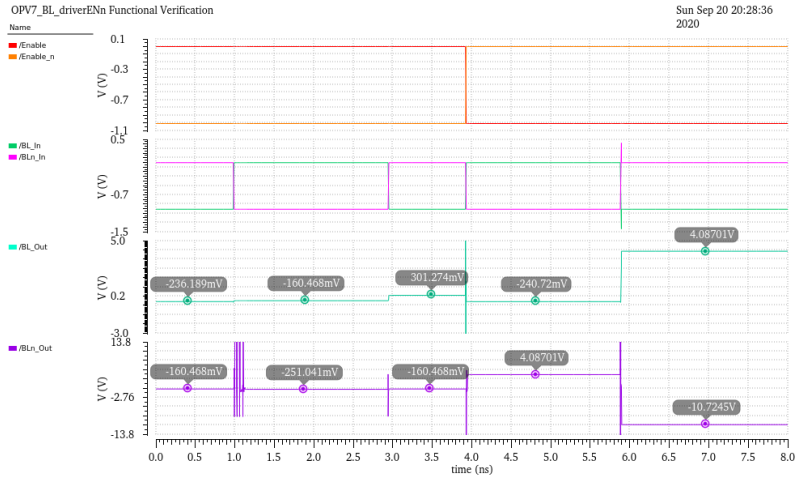


Figure 7.24: Spectre functional verification of bitline driver circuit implemented with OPV7-MolFET technology. In this the $\overline{EN}$ signal is directly generated from an ideal voltage source. Anyway the overall circuit is still not properly working.

is enabled (EN ='1'), the bitlines follow the input in which the data is forced, with a reduced dynamic due to the fact that the pull-up and

down network have not the proper electrical strength. In particular the noise margin for the high logic value is reduced of around 1% to 3%, whereas the one for the low logic value is reduced of around 9%. Moreover during read operation (i.e. when $\overline{EN}$ ='1'), the bitlines is electrically separated from the drivers.

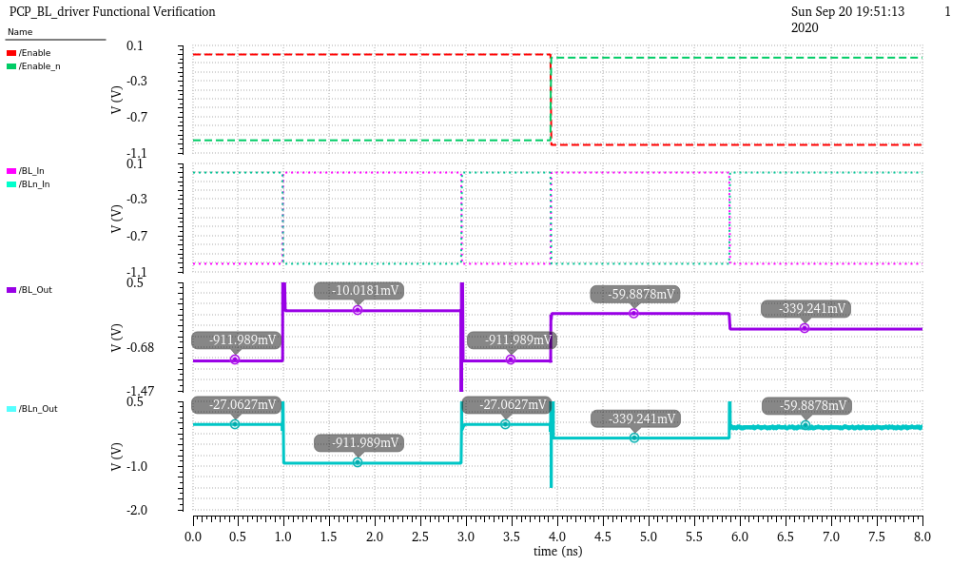


Figure 7.25: Spectre functional verification of bitline driver circuits implemented with PCP-MolFET technology.

– **Wordline drivers.** In figure 7.26 the behaviour of OPV7 molFET-based wordline driver is tested. Neither this simple circuit does not correctly behaves at all. Indeed when the input signal WL_{in} ='1' the wordline is not properly asserted (i.e. does not go to 0). Moreover, when WL_{in}='0' the wordline is not de-asserted (i.e. goes to low logic value corresponding to -1V). The oscillating behaviour starting from 5ns is likely due to convergency problems due to non linearities.

  Different is the case for PCP-based wordlines driver, whose simulation result is shown in figure 7.27. As can be seen, when the input signal WL_{in} ='1' the wordline is now properly asserted with a noise margin reduced of 3.7%, and when WL_{in}='0' the wordline is de-asserted noise margin reduced of 7.2%. Thus, it correctly behaves with acceptable noise margins.

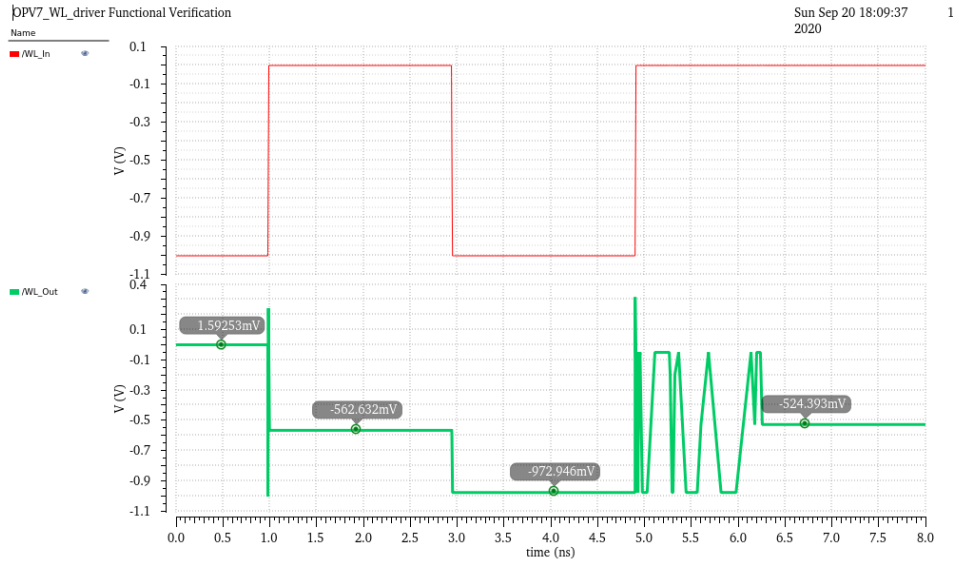


Figure 7.26: Spectre functional verification of wordline driver circuit implemented with OPV7-MolFET technology.

– **Precharge circuit.** In figure 7.28 the behaviour of OPV7 molFET-based precharge circuit is tested. It is evident that, when $\overline{PRE} = `0'$), the bitlines are not precharged to $V_{DD} = 0V$. Thus neither this circuit does not correctly behaves as intended.

  In the case of PCP-based precharge circuit. It is evident, from the simulation result shown in figure 7.29, that when the signal $\overline{PRE} = `0'$, the bitlines are precharged to $V_{DD} = 0V$ with noise margin reduced of 1%, whereas when the signal $\overline{PRE} = `1'$ the bitlines are isolated from the precharge circuit.

Finally the SRAM cell is tested in the most significant conditions, as previously done for MOSFET. A significant simulation waveform is reported in figure 7.30, where all the possible states of the memory are shown. Only PCP-based SRAM cell was tested, since the OPV7-based is not expected to work at all.

Firstly the EN_BL (pink line) and WL_IN (dash violet line) signals are not asserted allowing the bitlines and wordlines to not be enabled. Moreover, the $Precharge\_n$ signal is asserted so the cell is not in the precharge state, but is in idle state, thus the bitlines are floating and the data in the cell is

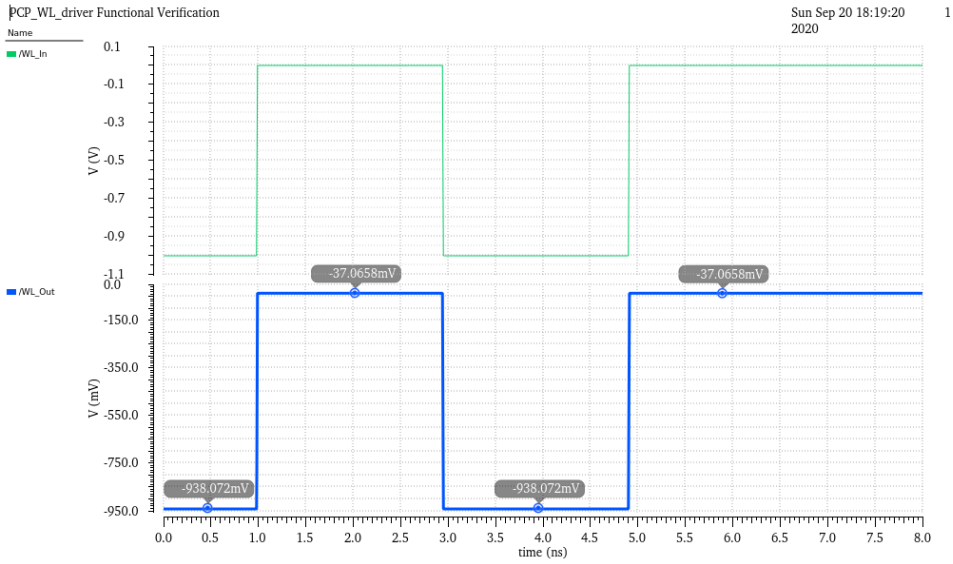


Figure 7.27: Spectre functional verification of wordline driver circuit implemented with PCP-MolFET technology.

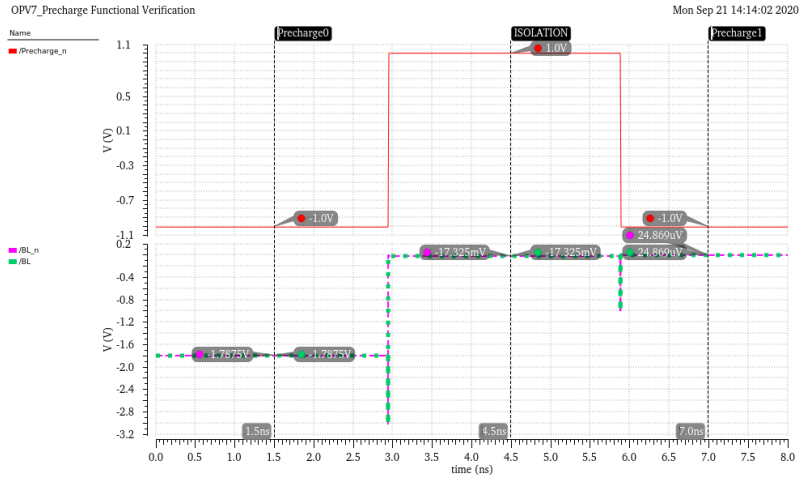


Figure 7.28: Spectre functional verification of precharge circuit implemented with OPV7-MolFET technology.

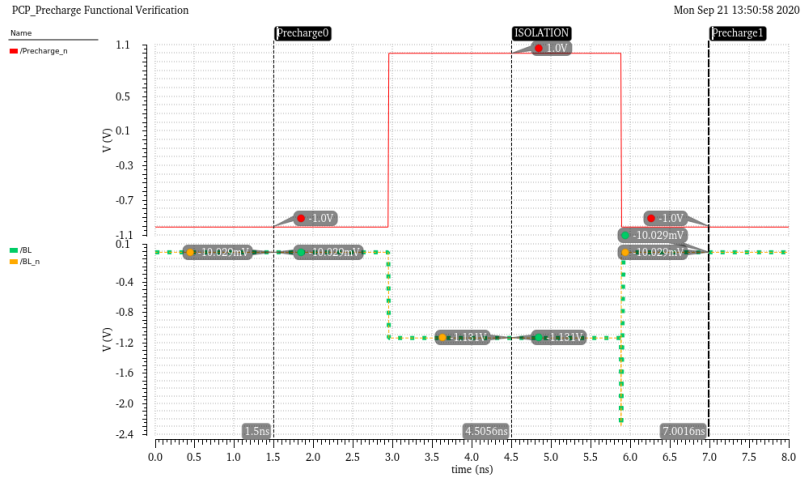


Figure 7.29: Spectre functional verification of precharge circuit implemented with PCP-MolFET technology.

undefined. Then the EN_BL signal is asserted and thus the bitlines are enabled and should follow the input data BL_In, BLn_In. Even if the BLs driver circuit correctly works, it is not able to correctly drive them when the complete circuit is assembled. This is likely due to the connection with the pass-transistors of the cell, that actually does not work effectively as pass-transistor, since the exploited ambipolar molFET was engineered to work as a transistor in a pull-up/down static CMOS network. In any case the complete circuit correctly works, with certain noise margins, in writing an high logic value (0V) in the cell, in keeping it during the precharge and idle cycles and in reading it. Instead does not correctly read a low logic value (-1V) after it has been correctly written. This is the phenomenon of data flip, explained in 7.1.1, which occurs because the sizing constraints for a correct read and write operation cannot be satisfied.

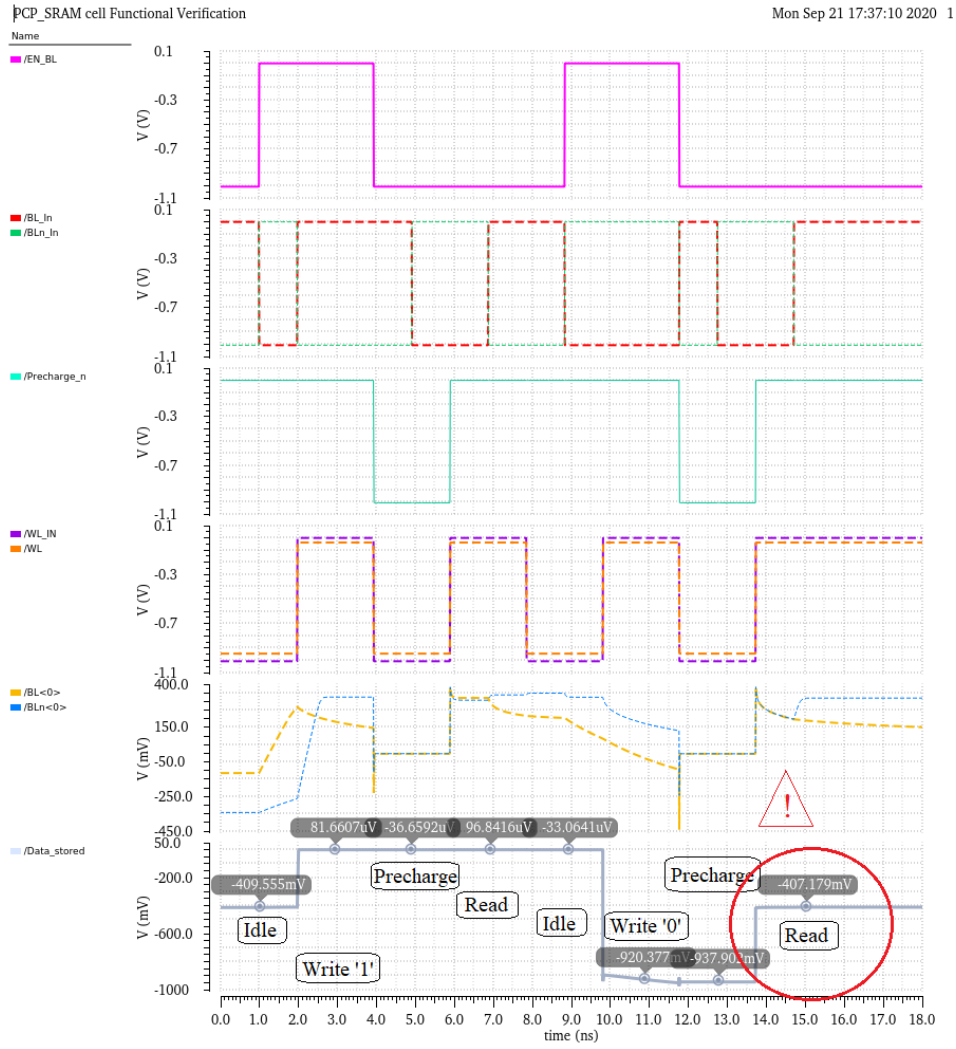


Figure 7.30: Spectre functional verification of a single cell of implemented with PCP-MolFET technology.

Chapter 8

LIM cell implementation

The final practical aim of this work at architectural level, is to assemble logic and memory blocks characterized in the previous chapters in a target digital application, a specific Logic-in-Memory cell. Notice that this step has no the pretense to provide as a final result a MolFET-based optimum and effective design of the cell, since already the costituent parts of the LiM cell suffer of many functional issues. This step, is instead a pretext to: (a.) test the effectiveness of the adopted methodological flow in a more complex circuit; (b) demonstrate the benefits, but above all, the problems in migrating to molecular technology for this specific application. Indeed the problems that will be showed will have the purpose to make the reader qware of the fact that exploiting molecular technology in conventional architectures based on static CMOS logic, is not a convenient path; (c.) propose and briefly discuss ameliorations of the cell able to fully exploit the overwhelming potentials of molecular technology.

The chapter is organized as followed. In 8.1 the drivers for the adoption of an in-Memory-Computing (iMC) paradigm are discussed and then the iMC concept is presented. In 8.1.3, the motivations behind the choice of LiM approach are discussed and an algorithm-specific LiM cell to be implemented is presented, then designed and tested both for a conventional MOSFET technology (FD-SOI 28nm) and MolFET technology. Also in this case no analysis of power, speed and area are provided. Indeed since the molecular implementation suffers of functionality issue it would be useless compare the two technologies.

8.1 in-Memory-Computing paradigm

8.1.1 Drivers

This subsection justifies the choice of a LiM cell as circuital block in the methodology benchmark and proof of molecular technology benefits, by analyzing the drivers in the present technological scenario.

Computing systems are in constant evolution thanks to technological advancement and architectural progress. Over the years, they have gained more and more computational power, allowing to provide support also for very high demanding applications, such as Machine Learning. However, the consequence of the growing complexity of processing units and applications is two-fold: (1) the unbearable impact on power consumption and the rise of the so called *power wall*, which makes impossible to exploit at the same time all the computational resources of the system (*dark silicon*); (2) the performance limitation in terms of bandwidth especially for data-intensive applications systems. Indeed due to different rate of progress between processing units and memories, between which an high rate exchange of data is required, another wall known as the *memory wall* is raised, because memories are not able to provide data as fast as CPUs are able to compute them.

In this contest, a solution for the two above mentioned problems could be go beyond conventional separation between computation and storage by integrating simple logic directly inside the memory cell, thus abandon Von Neumann paradigm on which all modern computing systems are founded and adopt an in-Memory-Computing approach.

Indeed, differently from Von-Neumann architecture where a lot of time is spent in retrieving data from the memory rather than computing them, in the iMC-based architecture data are computed directly inside the memory without the need to move them between the computing and the storage units, allowing to drastically reduce the amount of memory accesses and the associated energy consumption and latency, thus finally shooting down the power and memory walls while also providing high performance thanks to its flexibility and parallelism.

Therefore, in conclusion, the drawbacks related to von Neumann computing model is the main factor which justifies the choice of a LiM cell as circuital block in the methodology benchmark and proof of molecular technology benefits. In addition, this choice allows to fully exploit all potentials of molecular technology and in particular its flexibility in implementing in-memory computing architectures as it intrinsically provides both non volatility and computing capabilities simply by changing physical parameters in the same

device.

8.1.2 Concept review

The state of the art on iMC is vast. The works found in literature differentiate from each other mainly for the role that the memory has in computing data and by means of which technology are implemented (both on standard MOSFET-based technologies on emerging devices, like resistive and magnetic devices). Starting from this observation, a taxonomy for classifying previous works was defined. According to this taxonomy the in-memory computing approaches can be divided in four main categories, also reported, for sake of clarity, in figure 8.1:

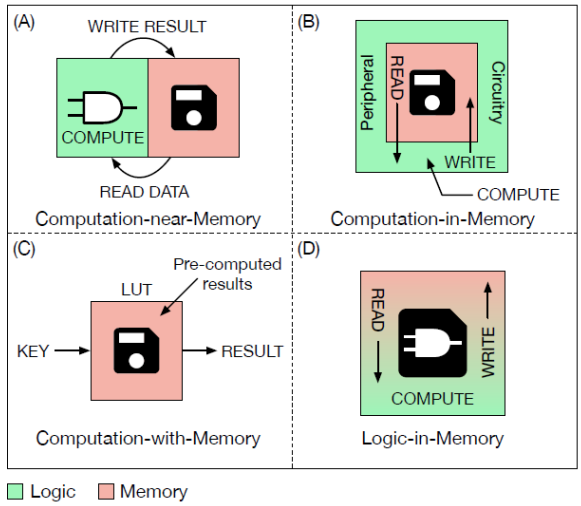


Figure 8.1: Different iMC approaches from [69]. Depending on how the memory is used for computing data, four main in-memory computing approaches can be defined. (A) Computation-near-Memory (CnM), (B) Computation-in-Memory (CiM), (C) Computation-with-Memory (CwM), (D) Logic-in-Memory (LiM).

(A) **Computation-near-Memory** (**CnM**, 8.1 A): thanks to the 3D Stacked Integrated Circuit technology (3D-SIC), computation and storage are brought closer together, by stacking the two units one on top of the other. This technique has the advantage to reduce the length of the interconnections thus widening the memory bandwidth. However, this

approach cannot be considered as true in-memory computing, since computation and storage are still two separate entities, but more as an evolution of conventional architectures based on the von Neumann model.

(B) **Computation-in-Memory** (**CiM**, 8.1 B): the structure of the memory array is not modified, while its intrinsic analog functionality is exploited to perform computation. In particular, in-memory computation is achieved by reading data from the memory which is then sensed by sense amplifiers (SAs). SAs are specifically modified in order to support the computation of a few simple logic operations (AND, OR, ...). The result is then written back in the memory array. Decoders are also adapted in order to read more than one data from the array and execute row-wise (between data on different rows) or column-wise (between data on different columns) operations.

(C) **Computation-with-Memory** (**CwM**, 8.1 C): this approach uses memory as a Content Addressable Memory (CAM) to retrieve pre-computed results by means of a Look Up Table (LUT). The working principle of this kind of computation is that any Boolean function involving two or more inputs can be encoded in a memory by storing its truth table. In particular, input combinations are stored in a LUT, while results are stored in a CAM. Then the LUT is accessed through an input combination and an address is retrieved. These addresses are used to access the CAM and obtain the final result.

(D) **Logic-in-Memory** (**LiM**, 8.1 D): in this case logic is directly integrated inside the memory cell. Differently from the other three approaches, here data are computed locally without the need to move them outside the array (towards a close computing unit as in a CnM approach or towards the peripheral circuitry as in a CiM approach). Internal readings are performed in order to execute operations on data stored in different cells, by exploiting inter-cells connections. Internal writings are executed to locally save the result of the operation. There are a few works belonging to this category, such as [70], [71], [72].

8.1.3 LiM approach

To overcome drawbacks of von Neumann model one possible approach, as presented previously, is the Logic-in-Memory one, which basically consists on conceive circuits where at least part of the computations are executed directly in the memory by adding simple logic inside the cell, that in principle can be also configurable (configurable Logic-in-Memory, cLiM [69]).

There are LiM cells very flexible and general purpose which allows the execution of many different algorithms, without being specifically built for any of them, but for this reason also complex and not optimized. A simpler structure could be thought renouncing to some capabilities of the array. Another possibility could be to specialize the structure of each cell: one memory row, for example, could contain only full adders, while the row after that could be specialised in the computation of the AND/OR logic functions.

Anyway, since the main aim of this chapter is to implement a LiM cell exploiting molecular technology, that is not at all a trivial task, simplicity of an already widely investigated structure constituted with very limited number of transistors was preferred. The chosen LiM cell to be implemented, proposed in [73], is an algorithm-specific cell and not configurable, capable to perform a minimum/maximum search inside the memory array. Details on its MOSFET and molecular implementations are widely discussed in the following sections.

8.2 LiM cell

8.2.1 Algorithm study

The algorithm that the architecture based on the chosen LiM cell to be designed must implement is for the maximum/minimum value search in a memory array. It is based based on the bitwise AND operation between the memory content (i.e. every word stored in memory) and an external datum called «mask vector». The operation is performed in parallel on all memory rows at the same time.

In von-Neumann architecture the algorithm would require the processor to fetchs all the data from the memory and to elaborate them in order to find the maximum/minimum. This operation is time and energy consuming, since many clock cycles are needed to read each word and a large amount of energy is wasted on the busses between the memory and the CPU for the data exchange. For this reason for this type of algorithm, it would be more convenient to perform this operation completely in memory with a LiM architecture, so that no time and energy have to be wasted to access the memory content.

In figure 8.2 a significant example is depicted to explain the steps needed to find the maximum value among unsigned binary data. With NUMBER is denoted the datum stored in the memory rows, while with MASK is denoted the mask vector having the same width of the NUMBER stored in each

memory row and is characterized by a '1' in a position that shifts for each step from MSB to LSB. At each step, the bitwise AND operation is performed between each NUMBER and MASK, and the results of these operations are recorder by the logic that surrounds the array.

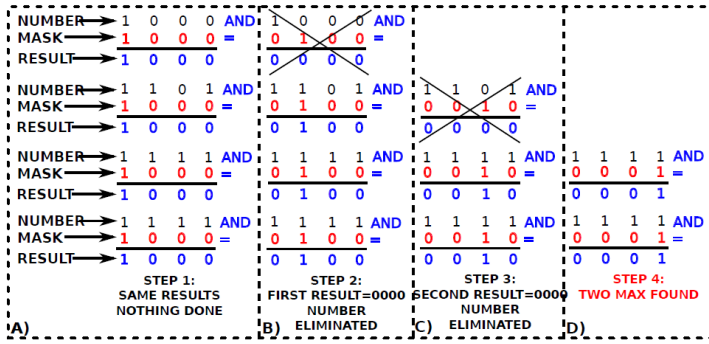


Figure 8.2: Example of the algorithm execution for maximum search from [73].

At the beginning of the algorithm execution, all the word are listed in a search list (in particular, all the rows addresses are considered valid), since each NUMBER is a maximum value candidate. The NUMBER for which the result of bitwise AND operation is zero, are removed from the search list since they are no more maximum candidates, while the others are kept.

In particular, in the example of figure 8.2, the algorithm starts with the mask vector value '1000', that is used to check which rows have the MSB (Most Significant Bit) equal to '1': in fact, considering unsigned binary values, the largest data have the MSB equal to '1', while the others have a smaller encoded value. In this first step no word is removed from the search list since all the AND results have the MSB equal to '1'. In the second step, the same operation is performed changing the mask value: the '0100' value is used to check which words have the $MSB - 1$ equal to '1'. As in the previous case, the words that «match» with the MASK are kept in the search list, while the others are removed. These steps are repeated until the LSB (Least Significant Bit) of the words is reached: at the end of this step, only one word will remain in the search list (unless there are two or more identical data stored in memory, as it happens in the example), and the address of this is provided in output as the maximum value location.

Notice that in the worst case the algorithm, since one bit of every stored word is processed at time (serial algorithm), it takes N steps to detect the maximum (where N is the bit length of the word). However the comparison is performed in parallel on all the numbers. For this reason the execution

time of the algorithm is very small and it is independent of the number of stored data (i.e. memory size). In a von Neumann architecture, instead, the number of data to be read has a strong influence of the search latency, since at least one clock cycle is needed to read each word and to store it in the processor registers.

Moreover, the algorithm is also able to detect minimum values. By excluding, at each step, from the list those numbers for which the results of the comparison are different from '0', the minimum value is detected.

8.2.2 Hardware implementation

A possible hardware solution able to perform the above explained algorithm is a custom memory array based on enhanced memory cells which embed logic functionalities. Obviously, including logic functionalities inside memory cells increments the area (very insignificant in molecular implementation), but greatly increases the performance. In particular every LiM cell performs a logic AND between its stored value and a mask bit. Its conceptual structure is reported below in figure 8.3, where the MEM CELL block denotes the memory core, that can be implemented with a SRAM (6T) or DRAM (1T). Then an AND logic gate is connected between a MASK bit and the output of the storage element and a *wired OR* scheme is exploit to connect the output of all AND gates in one memory row, to detect if at least one has an output equal to '1'.

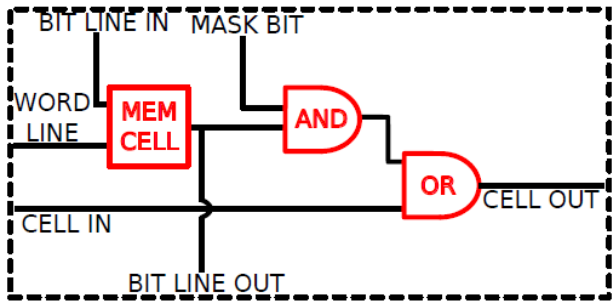


Figure 8.3: LiM cell conceptual schematic from [73].

Each LiM cell is placed inside the memory, based on a standard array structure, as shown in figure 8.4. Each memory cell in a row has an input bit line and an output bit line, in order to store and read data, and is arranged in a chain of OR gates, whose result is recorded by the logic that surrounds the

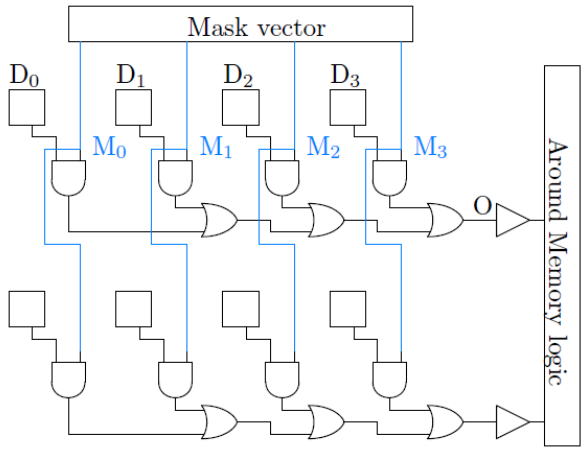


Figure 8.4: Memory array conceptual schematic, where D_i denotes the i-th bit of the datum stored and M_i the i-th bit of the mask vector.

memory array and used by this to update the search list during the maximum/minimum value search. In the algorithm presented previously, a cell is selected by setting the corresponding mask bit to '1', while the other cells are disabled (i.e. the corresponding AND gate output is forced to '0') by setting the associated mask bits to '0'. This implies that the result of the OR chain depends only on the result of the AND between the selected cell content and the corresponding mask bit, which is set to '1'. Hence, the result depends exclusively on the selected cell content.

Finally row and column sense amplifiers are used to read data inside the memory and to read the result of the wired OR line. Moreover, notice that the proposed circuit is intended to function normally as a memory and sometimes to perform logic operations on stored data.

Hereafter we will deepen into the LiM cell only, since the aim of this chapter is to design, test, analyze a single LiM cell implemented with CMOS and non-CMOS (molecular) technologies. Different topological configuration can be chosen both for the memory and logic circuit, but the criteria in the choice are again: simplicity and reduced number of transistors (or electronic devices in general) to ensure the feasibility and functioning of the cell, especially in the molecular case. According to these criteria only one memory cell, the typical 6T SRAM cell characterized in the previous chapter, in conjunction with two different logic cells are investigated: one based on static CMOS logic and a special purpose one optimized for the specific algorithm to be

implemented. Therefore two different LiM cell configurations, reported for clarity in figure 8.5, are designed, tested and analyzed for both technologies.

It is important to highlight that this first technological migration attempt for this specific applications from CMOS technology to molecular one does not do justice of the potentialities of molecular technology, since a direct substitution of MOSFETs with molFETs was done. As a consequence, problems especially in the functioning of the cell and not so revolutionary benefits are expected. For this reason after this first attempt, an enhanced cell able to exploit molecular technology potentialities and molecular memory element with inherent logic capabilities to greatly improve performances and area occupation, is proposed in section **??**.

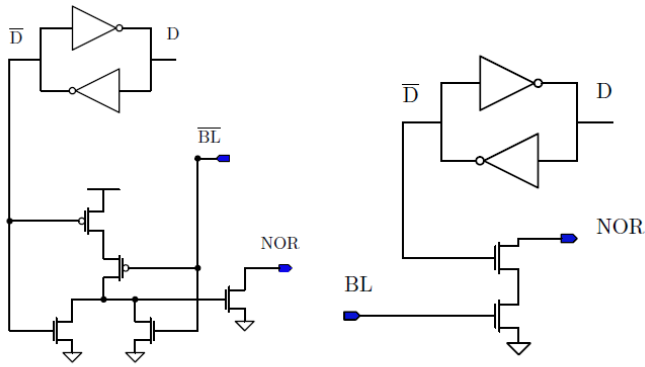


Figure 8.5: The two different LiM cells configurations that will be considered in this chapter. In this figure only the MOSFET-based implementation is shown. The configuration on the left is the one labelled in the following as *Static CMOS-like* since is constituted with AND and OR gates implemented in static CMOS logic. The configuration on the right is the one labelled in the following as *Special purpouse*.

8.3 MOSFET-based LiM cell

In this section the two different LiM cells are implemented exploiting the MOSFET-based SRAM cell and logic components characterized in the previous chapters. In subsections 8.3.1, 8.3.2, the LiM cells are respectively designed and verified and analyzed in terms of power, speed and area.

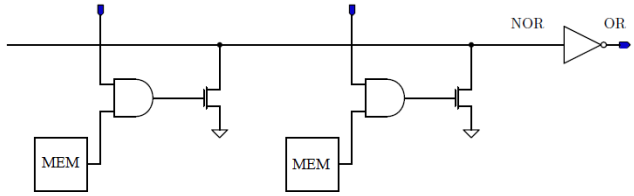


Figure 8.6: A row of memory array where all the LiM cells share the same wired-OR line.

8.3.1 Design

Static CMOS-like LiM cell. The topology shown in figure 8.7 and designed in Cadence Virtuoso (figure 8.8), is characterized by a static CMOS-like logic AND version. The nMOS placed in the final stage has the purpose of pull down the wired OR line shared between all the LiM cells present in a memory row, when the output of the AND gate is '1' as shown in figure 8.6.

The transistors have been sized according to the standard design rules for CMOS static logic: since the two pMOS transistors are in series, an aspect ratio equal to 4 (i.e. 4 times larger than the minimum aspect ratio) has been chosen instead of 2; concerning the nMOS, a minimum aspect ratio has been chosen, since only one MOSFET per pull-down path is present.

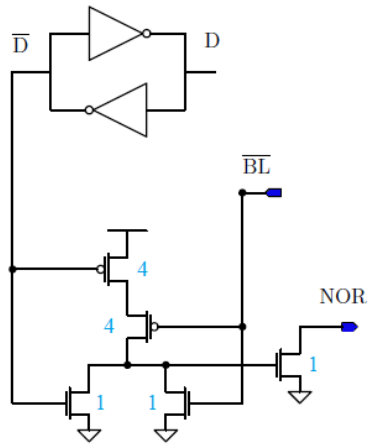


Figure 8.7: Static CMOS-like LiM cell with chosen AR of transistors channel width. Only the bistable ring of the SRAM cell is shown, pass transistors and surrounding WLs and BLs are omitted for sake of clarity.

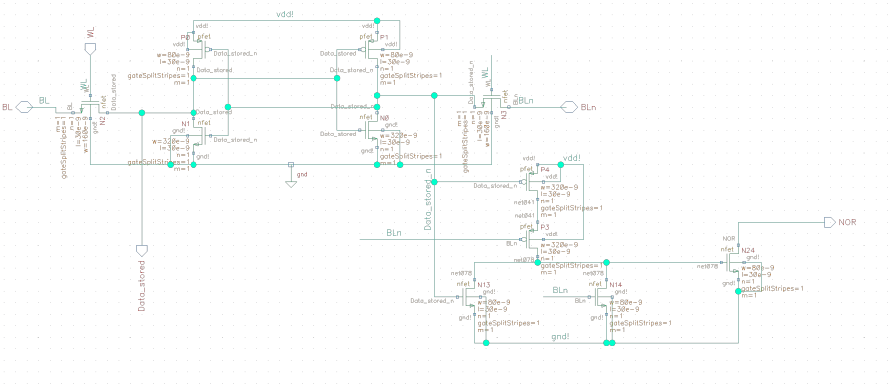


Figure 8.8: Static CMOS-like LiM cell Cadence schematic with chosen AR of transistors channel width.

The behaviour of the cell as function of the input is trivial and reported for sake of clarity in the following table:

Special purpose LiM cell. The LiM topology shown in figure 8.9 and designed in Cadence (figure 8.10), is specific for the algorithm explained in subsection 8.2.1.

Cell behaviour					
D	BL	$\overline{D}$	$\overline{BL}$	AND	NOR
0	0	1	1	0	$0 \rightarrow 1$
0	1	1	0	0	$0 \rightarrow 1$
1	0	0	1	0	$0 \rightarrow 1$
1	1	0	0	1	0

Table 8.1: Static CMOS LiM cell truth table.

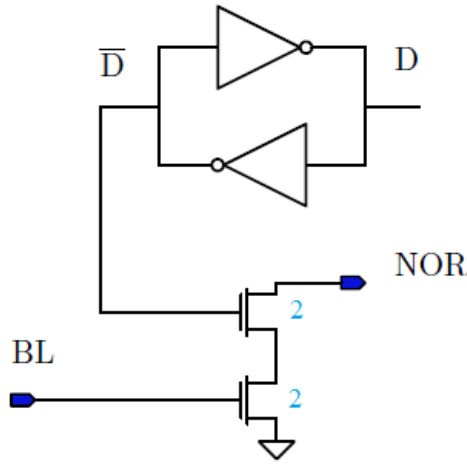


Figure 8.9: Special purpose LiM cell with chosen AR of transistors channel width. Only the bistable ring of the SRAM cell is shown, pass transistors and surrounding WLs and BLs are omitted for sake f clarity.

Indeed since the AND operation is exploited to select the cell whose content has to be checked, a full dedicated AND gate inside the cell is not strictly necessary. With this simple special purpose logic, the cell is selected by asserting the bitline (BL='1'), while it is deselected by de-asserting it (BL='0'). In this way, the deselected cells on the row are unable to discharge the NOR line, since their footer transistor is disabled, and the result of the operation is completely determined by the selected cell content.

Thus, when:

– the cell is storing a logic '1' ($\overline{D}$='0'): the pull-down path of the cell is disabled and the NOR line is charged to V_{DD}: a logic '1' is obtained as result.

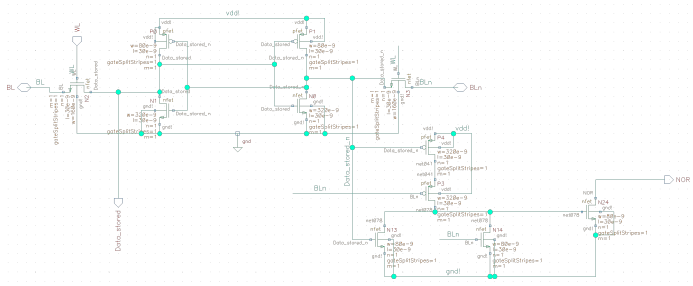


Figure 8.10: Special purpose LiM cell Cadence schematic with chosen AR of transistors channel width.

– the cell is storing a logic '0' ($\overline{D}$='1'): the pull-down path is enabled and the NOR line is not charged: a logic '0' is obtained as result.

Notice that, at the end of the story, the NOR line actually behaves like an OR line because: when the cell content is equal to '0', the line remains at '0'; whereas when the cell content is equal to '1', the line is charged to '1'. Hence, there is no need for the inverting stage at the end of the row as shown in figure 8.6.

In the following table the cell behavior as function of the inputs is reported. Notice that the cell content is evaluated only when BL='1', otherwise the pull-down path of the cell is always disabled and, so, the NOR line is always charged. When BL='1', the result of the AND operation (in practice, the cell content since BL='1') is reported in output.

Cell behaviour				
D	BL	$\overline{D}$	$D \cdot BL$	NOR
0	1	1	0	0
1	1	1	1	$0 \rightarrow 1$
X	0	X	X	$0 \rightarrow 1$

Table 8.2: Special purpose LiM cell truth table.

8.3.2 Testing

Static CMOS-like LiM cell.

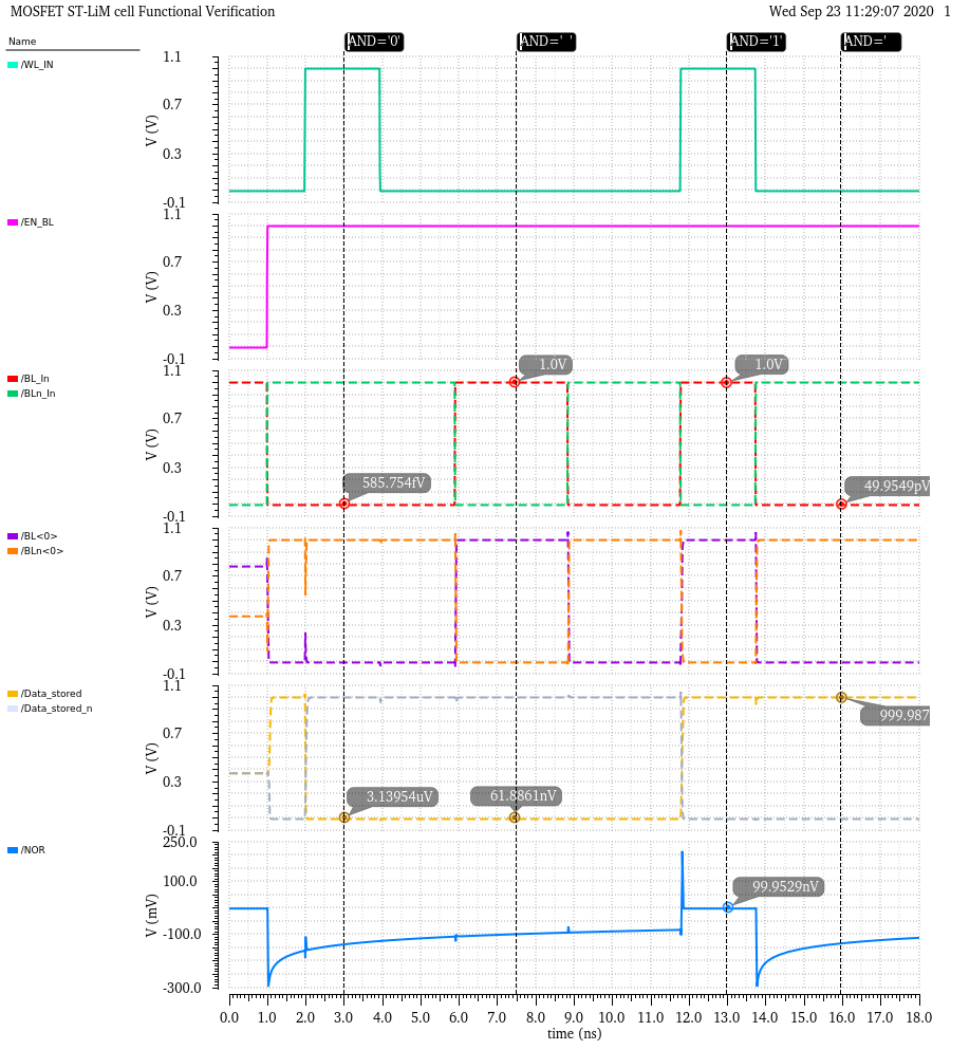


Figure 8.11: Functional verification of the static CMOS-like LiM cell.

Special purpose LiM cell.

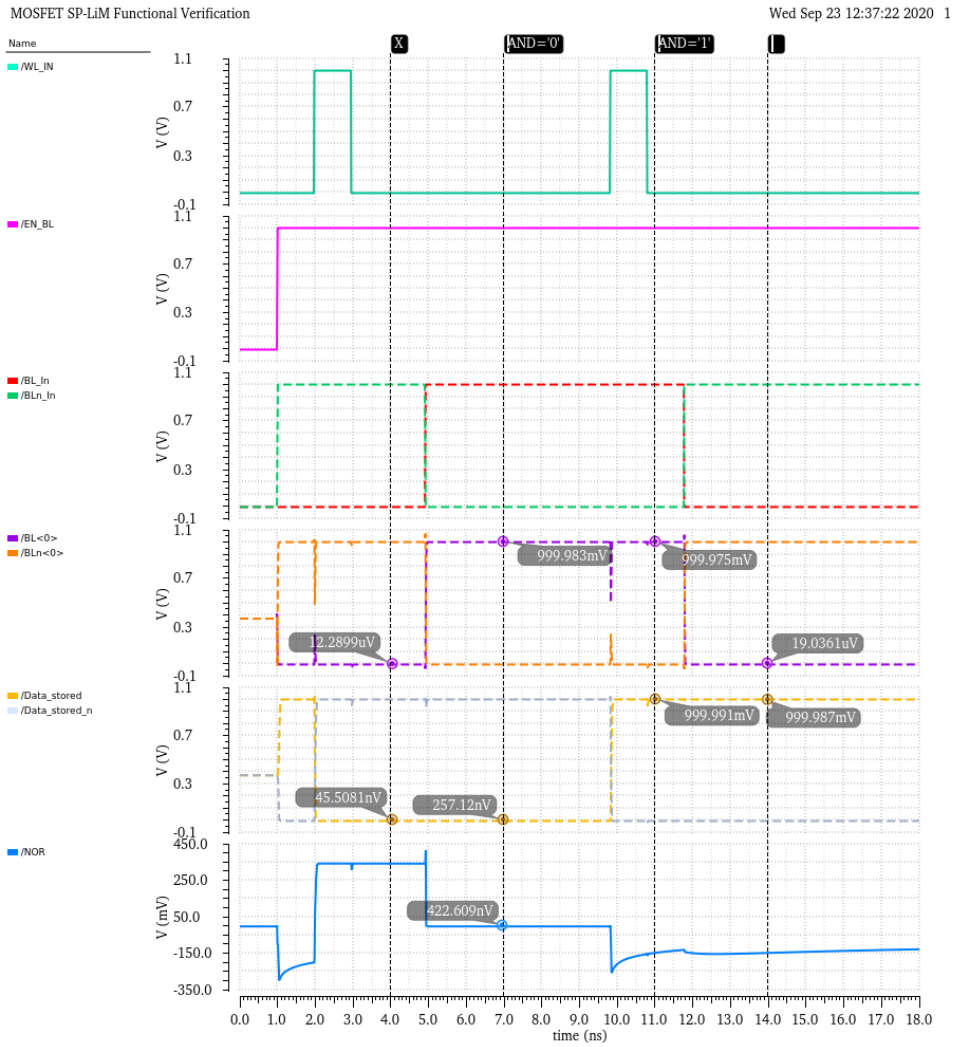


Figure 8.12: Functional verification of the special porpuse LiM cell.

8.4 MolFET-based LiM cell

In this section the two different LiM cells are implemented exploiting the MolFET-based SRAM cell and logic components characterized in the previous chapters. In the following subsection 8.4.1, the LiM cells are respectively designed and verified.

8.4.1 Design and testing

The two LiM cells were designed simply substituting MOSFETs with MolFETs, in particular only PCP-MolFETs for obvious reasons. The testing phase were made on two different types of implementations:

A. full PCP-MolFET-based implementation of drivers, SRAM cells and logic blocks. In this case the LiM cell does not correctly behaves at all (fig. 8.13), as expected. The reasons is in the pass-transistors of the SRAM cell that does not allow the cell to correctly store and write data. As a consequence the overall LiM cell it is not suitable to be integrated in an array that implement the algorithm of min/max search.

B. hybrid MOSFET/PCP-MolFET-based implementation: driver circuits and pass transistors of SRAM cell with MOSFETs, whereas the bistable ring (core of SRAM) and logic blocks with PCP-MolFETs. This implementation was designed in order to overcome the limitations of MolFET as pass-transistor. In this case the LiM cell correctly behaves with acceptable noise margins, confirming again that the problem was in the pass transistors of the SRAM cell. Obviously the number of LiM cells that can be putted in cascade is very limited w.r.t. MOSFET-based LiM, since after a certain point noise margins in the output dynamic of each cell will create serious problems in identifying the correct logic values, thus leading to malfunctioning of the LiM array.

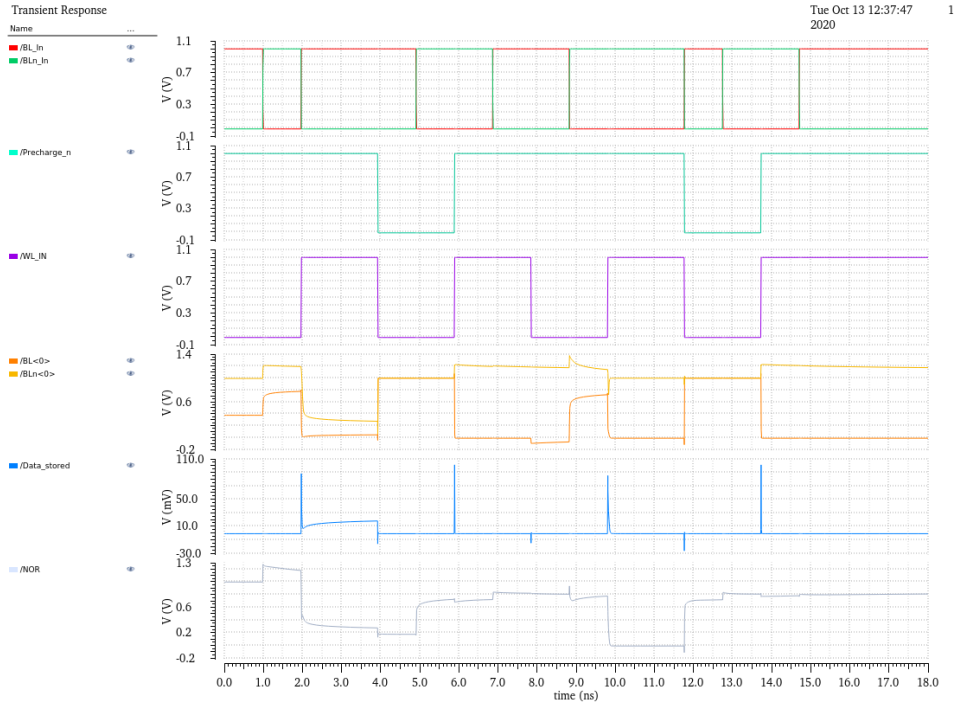


Figure 8.13: Functional verification of the static LiM cell. Case (a).

Chapter 9

Conclusions and outlook

Molecular electronics can be a a viable "Beyond-CMOS" solution, able to face the restless demand in scalability, low power consumption and high performances of nowadays electronic systems. Indeed it is, in principle, able to provide an incomparable integration. Moreover, a smart and cheaper fabrication process based on self-assembly have always drawn the attentio of researchers and companies.
Molecular transistor is one of the cornerstone of this intriguing field, and the effort spent in this thesis is focused on investigate it and to demonstrate if there are any realistic potential advantages in exploit it at circuital level. In particular, the theoretical background useful to understand conduction through this type of device is presented in part I and partially II, and its potential is investigated at device and circuital level at the end of part II and part III, respectively.

The main contributions of this MSc. thesis are in the methodology conceived and employed and in the results achieved.
Concerning the former, a truly engineer approach was adopted. I say "engineer" because of the following key points:

- The deepening into quantum transport and its understanding, essential in designing molecular devices to predict the charge transport behavior, was achieved in a "ready-to-use" manner.
- The effort spent at physical and device levels was exploited in order to practically understand the effectiveness of the device at circuital level, in terms of area, power and speed. To this end, a practical circuital model, widely discussed in 5, is conceived and exploited.

Then, concerning the results achieved: they are relevant if compared with the previous works found in current literature. Indeed, in particular:

- a novel ambipolar MolFET based on PCP molecule is carefully engineer in the device geometry and biasing conditions, and then fully characterized by means of atomistic simulations. An $Ion/Ioff$ ratio of more than 1400 is achieved, making the way for a molecular-based IC a little bit closer.
- a MolFET-based implementation of logic and memory elements and a target Logic-in-Memory application is performed. Design, functional verifications, performances analysis are performed in Cadence Virtuoso® both for molecular ambipolar transistors and FD-SOI MOSFET (28 nm) technologies in order to be compared and demonstrate the benefits and problems, for these specific applications, in migrating from conventional to a molecular technology. The result of this comparison shows some critical functionality issues. However, it has been shown that there exist advantages in moving from a conventional MOSFET technology toward molFET ones, especially in terms of area and in dynamic power (reduction of 2 and almost 1 order of magnitude respectively). On the other hand there also drawbacks in static power consumption (almost doubled), and above all the biggest issue to deal with would likely be the maximum operating frequency of molFET (reduction of more than one order of magnitude).

At last, the stated goal of the thesis is achieved. It reaches an adequate awareness of how nature of molecular technology is often very different from CMOS established one. This deviation leads to circuit malfunctions if it exploited in conventional architectures. As a consequence conventional architectural paradigms cannot really do justice of the overwhelming potentialities of molecular technology. Therefore possible alternative solutions from state-of-art literature can be conceived and investigated. One of the most promising and exotic ideas could be directly substitute logic or storage elements with single molecules as reported in [74], [75].

Appendix A

Atomistic simulations in QuantumWise ATK ®

A.1 *ATK* simulation set up

In this first chapter of the appendix, firsly the set up of all the main parameters, in order to perform atomistic simulations in ATK for molecular transistors, are reported for sake of completeness. Finally, intermediate results simulations for OPV7 and PCP Ambipolar-MolFETs are shown.

In order to analyze transport through a MolFETs, the following steps must be followed to set up the simulation in *QuantumWise ATK*:

1) Build the molecular junction, i.e. the molecule under test (MUT) anchored between two gold (111) surfaces that work as S/D electrodes [76].
2) Set up the device zero-bias calculation [77]. In particular in this step the following set up in the main parameters, show in the table, was chosen:

∗ for OPV7 molecule:

	ATK-SE: Extended Hückel Calculator
Basic Device	Electron temperature: 300 K, Density mesh cut-off: 10 Hartree, k-point sampling: $n_a = 1$, $n_b = 1$, $n_c = 135$, SCF iteration = YES, Left/Right electrode voltage = 0V
Countour integral parameter	default settings
Device algorithm parameters	real self-energy calculator: Krylov complex self-energy calculator: Sparse Recursion
Electrode parameters	default settings
Huckel basis set	Basis type: Hoffmann, Weighting scheme: Wolfsberg
Iteration control parameter	tolerance = 1e-05, maximum steps = 100, algorithm = PulayMixer damping factor = 0.1 history steps = 20
Numerical accuracy parameters	default settings
Poisson Solver	Conjugate Gradient transversal direction A: PBC, transversal direction B: Neumann BC transport direction C: Dirichlet BC

∗ for PCP molecule:

	ATK-SE: Extended Hückel Calculator
Basic Device	Electron temperature: 300 K, Density mesh cut-off: 10 Hartree, k-point sampling: $n_a = 3$, $n_b = 1$, $n_c = 135$, SCF iteration = YES, Left/Right electrode voltage = 0V
Countour integral parameter	default settings
Device algorithm parameters	real self-energy calculator: Krylov complex self-energy calculator: Sparse Recursion
Electrode parameters	default settings
Huckel basis set	Basis type: Hoffmann, Weighting scheme: Wolfsberg
Iteration control parameter	tolerance = 1e-05, maximum steps = 100, algorithm = PulayMixer damping factor = 0.1 history steps = 20
Numerical accuracy parameters	default settings
Poisson Solver	Conjugate Gradient transversal direction A: PBC, transversal direction B: Neumann BC transport direction C: Dirichlet BC

3) Set up the parameters for the equilibrium transmission spectrum analysis. The following configuration were chosen:

   * for OPV7 molecule:

	Analysis: Transmission Spectrum
Energy range	$E_0 = -2.5$, $E_1 = 2$, Points= 101
k-point sampling	$n_a = 1$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

   * for PCP molecule:

	Analysis: Transmission Spectrum
Energy range	$E_0 = -2$, $E_1 = 2$, Points= 101
k-point sampling	$n_a = 3$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

4) Set up the parameters for the non-equilibrium transmission spectrum and IV-curve analysis [77]. The equilibrium analysis file obtained from the previous step must be selected as input for the out-of equilibrium calculation. The following configuration for the IV block were chosen:

   * for n-OPV7 MolFET transchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = 1V$
Voltage gate bias range	$V_0 = -1$, $V_1 = 1$, @$V_{DS} = 0.1V$, @ $V_{DS} = 1V$
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 1$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

   * for n-OPV7 MolFET outchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = 1V$
Voltage drain bias range	$V_0 = 0$, $V_1 = 1$
Voltage front gate bias range	$V_0 = -1$, $V_1 = 3$, Points=9
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 1$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

∗ for p-OPV7 MolFET transchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = -3V$
Voltage gate bias range	$V_0 = -1$, $V_1 = 1$, @$V_{DS} = 0.1V$, @ $V_{DS} = 1V$
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 1$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

∗ for n-OPV7 MolFET outchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = 1V$
Voltage drain bias range	$V_0 = -1$, $V_1 = 0$
Voltage front gate bias range	$V_0 = -3$, $V_1 = 1$, Points=9
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 1$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

∗ for p-PCP MolFET transchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = -3.3V$
Voltage gate bias range	$V_0 = -1$, $V_1 = 1$, @$V_{DS} = 0.1V$, @ $V_{DS} = 1V$
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 3$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

∗ for p-PCP MolFET outchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = -3.3V$
Voltage drain bias range	$V_0 = 0$, $V_1 = 1$
Voltage front gate bias range	$V_0 = -1$, $V_1 = 3$, Points=9
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 3$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

∗ for n-PCP MolFET transchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = 1.2V$
Voltage gate bias range	$V_0 = -1$, $V_1 = 1$, @$V_{DS} = 0.1V$, @ $V_{DS} = 1V$
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 3$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

∗ for n-PCP MolFET outchar:

	Analysis: IV-curve
Voltage back gate	$V_0 = 1.2V$
Voltage drain bias range	$V_0 = -1$, $V_1 = 0$
Voltage front gate bias range	$V_0 = -3$, $V_1 = 1$, Points=9
Energy range	$E_0 = -2$, $E_1 = 2$, Points=101
k-point sampling	Grid type: Monkhorst-Pack $n_a = 3$, $n_b = 1$
Energy zero parameter	Average Fermi level
Infinitesimal	1e-06 eV
Self-energy calculator	Krylov

A.2 *ATK* simulation results

The following simulation results are intermediate results useful in order to accurately engineer the behavior of the two ambipolar MolFETs, by properly choosing the biasing conditions, and in particular by choosing the values of back gate voltages in order to have:

* n-type MolFETs ON when $V_{FG} = +1V$ and OFF when $V_{FG} = -1V$;
* p-type MolFETs ON when $V_{FG} = -1V$ and OFF when $V_{FG} = +1V$.

A.2.1 n-OPV7 MolFET

* @ OFF state: sweep in $V_{BG} = [0,3]V$ @ $V_{FG} = -1V$, for $V_{DS} = [0,1]V$.

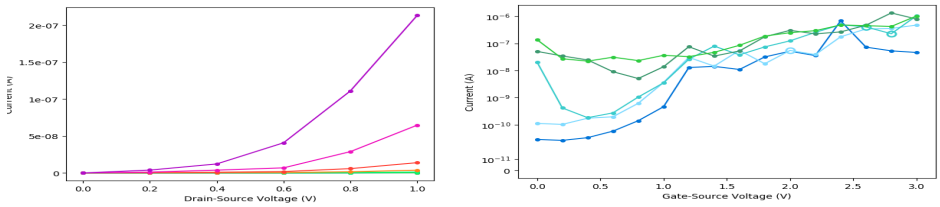


Figure A.1: On the left: parametric Output characteristic of n-OPV7 MolFET when it is in the OFF state, with parameter V_{BG}; On the right: parametric trans-characteristic in log scale of n-OPV7 MolFET when it is in the OFF state, with parameter V_{DS}.

* @ ON state: sweep in $V_{BG} = [0,3]V$ @ $V_{FG} = +1V$, for $V_{DS} = [0,1]V$.

A.2.2 p-OPV7 MolFET

* @ OFF state: sweep in $V_{BG} = [-3,0]V$ @ $V_{FG} = +1V$, for $V_{DS} = [-1,0]V$.
* @ ON state: sweep in $V_{BG} = [-3,0]V$ @ $V_{FG} = -1V$, for $V_{DS} = [-1,0]V$.

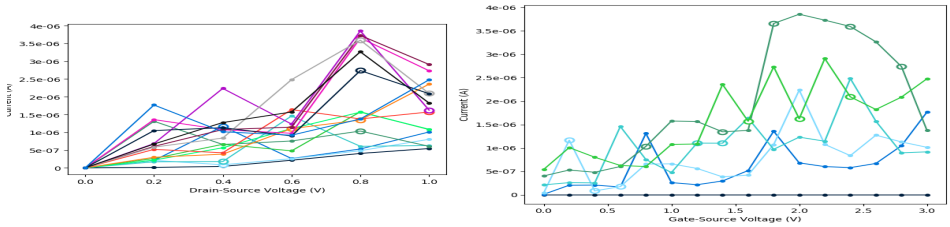


Figure A.2: On the left: parametric Output characteristic of n-OPV7 MolFET when it is in the ON state, with parameter V_{BG}; On the right: parametric transcharacteristic in log scale of n-OPV7 MolFET when it is in the ON state, with parameter V_{DS}.

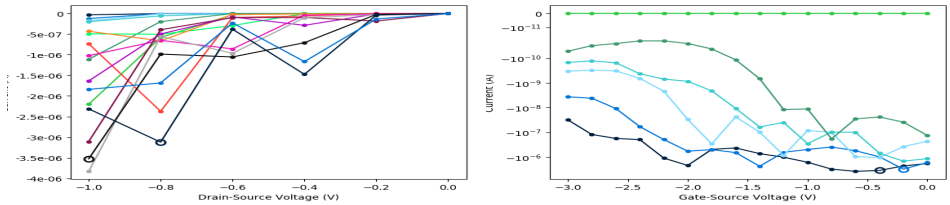


Figure A.3: On the left: parametric Output characteristic of p-OPV7 MolFET when it is in the OFF state, with parameter V_{BG}; On the right: parametric transcharacteristic in log scale of p-OPV7 MolFET when it is in the OFF state, with parameter V_{DS}.

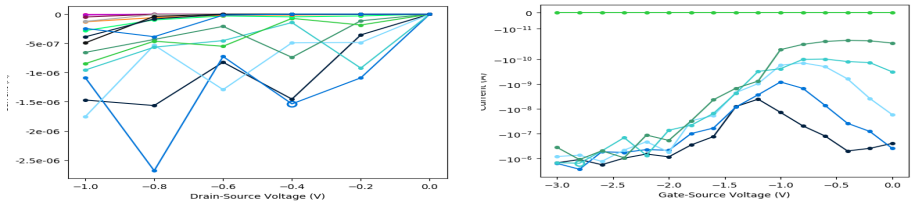


Figure A.4: On the left: parametric Output characteristic of p-OPV7 MolFET when it is in the ON state, with parameter V_{BG}; On the right: parametric transcharacteristic in log scale of p-OPV7 MolFET when it is in the ON state, with parameter V_{DS}.

A.2.3 n-PCP MolFET

* @ OFF state: sweep in $V_{BG} = [0,3]V$ @ $V_{FG} = -1V$, for $V_{DS} = [0,1]V$.

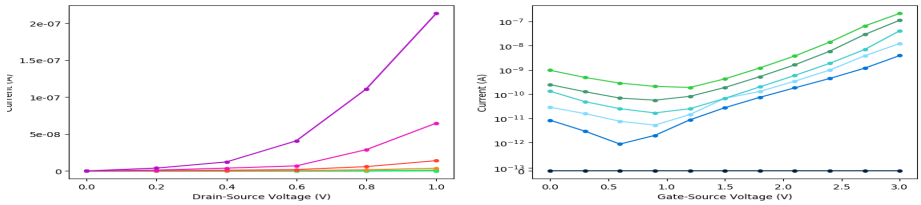


Figure A.5: On the left: parametric Output characteristic of n-PCP MolFET when it is in the OFF state, with parameter V_{BG}; On the right: parametric transcharacteristic in log scale of n-PCP MolFET when it is in the OFF state, with parameter V_{DS}.

* @ ON state: sweep in $V_{BG} = [0,3]V$ @ $V_{FG} = +1V$, for $V_{DS} = [0,1]V$.

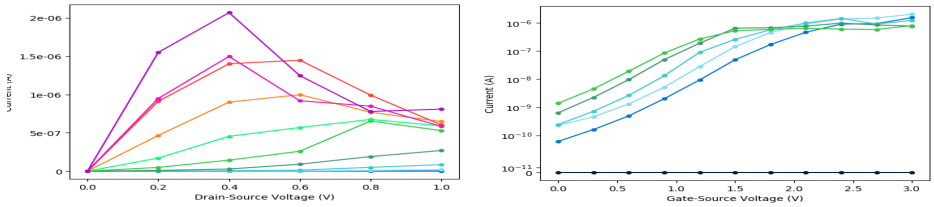


Figure A.6: On the left: parametric Output characteristic of n-PCP MolFET when it is in the ON state, with parameter V_{BG}; On the right: parametric transcharacteristic in log scale of n-PCP MolFET when it is in the ON state, with parameter V_{DS}.

A.2.4 p-PCP MolFET

* @ OFF state: sweep in $V_{BG} = [-3,0]V$ @ $V_{FG} = 1V$, for $V_{DS} = [-1,0]V$.

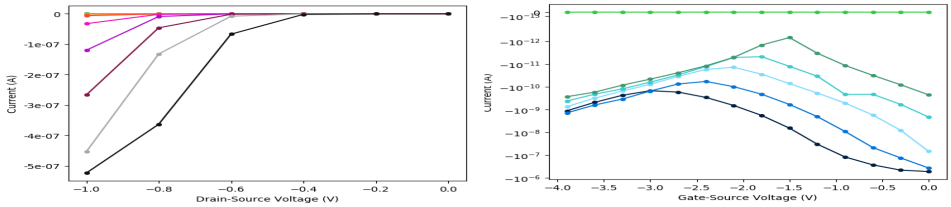


Figure A.7: On the left: parametric Output characteristic of p-PCP MolFET when it is in the OFF state, with parameter V_{BG}; On the right: parametric transcharateristic in log scale of p-PCP MolFET when it is in the OFF state, with parameter V_{DS}.

* @ ON state: sweep in $V_{BG} = [-3,0]V$ @ $V_{FG} = -1V$, for $V_{DS} = [-1,0]V$.

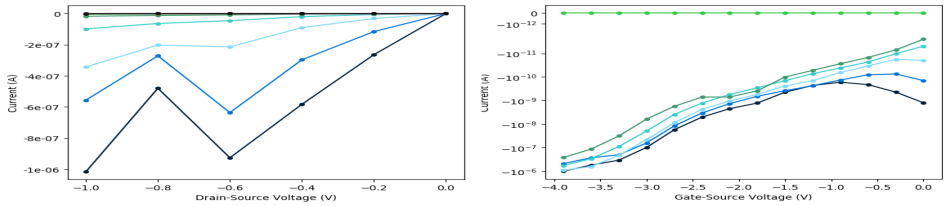


Figure A.8: On the left: parametric Output characteristic of p-PCP MolFET when it is in the ON state, with parameter V_{BG}; On the right: parametric transcharateristic in log scale of p-PCP MolFET when it is in the ON state, with parameter V_{DS}.

Appendix B

Modeling molFET in Cadence Virtuoso®

B.1 MatLab scripts for MolFET RC parameters estimation

```
%% The aim of this script is to provide a quantitative estimation of molFET
%  capacitances Cg, Cs, Cd and resistances Rg, Rs, Rd.
clear all
close all
clc

disp('Welcome in molFET parameters calculator program!');
% Load data:
output_data_folder = 'C:\HUCKEL_IV_suite\output_data';
%%%% %%%% %%%% %%%%     NO  MORE  USER  INPUTS      %%%% %%%% %%%% %%%%
% If no variations to output data storage are made the following paths are
% automatically found (DO NOT MODIFY!):
% "Hrho.mat" structure path:
Hrho_path = strcat(output_data_folder, '\Hrho.mat');
% "device" file path:
device_path = strcat(output_data_folder, '\device');
% "TEV.mat" structure path:
TEV_path = strcat(output_data_folder, '\TEV.mat');
% "HS.mat" structure path:
HS_path = strcat(output_data_folder, '\HS.mat');
% "id" file path:
id_path = strcat(output_data_folder, '\id');
% correct data import:
disp('Data are correctly imported from:');
disp(output_data_folder);
string_to_be_displayed = sprintf('- - - - - - - - - - - - - - - - - - - -
- - - - - \n');
disp(string_to_be_displayed);
%% Cg gate capacitance (assumed to be constant):
% by default in Huckel-IV 3.0 the molecule is supposed to be inside a box
% of dimensions 100x100xL (angstrom) L=length of the molecule,
% so the width W of the gate is W = box_length = 100 angstrom.
% Simulation domain dimension (DO NOT MODIFY!):
box_length = 100; % angstrom
% gate width in angstrom:
W = box_length; % angstrom
% The following function performs the Cg calculation:
[Cg] = Cg_calculator(Hrho_path, device_path, W);
```

```
% Display it:
string_to_be_displayed = sprintf('The electrostatic gate capacitance resulted to be:
Cg = %g F', Cg);
disp(string_to_be_displayed);
%% Approximated estimation of Cs and Cd from parallel plane approximation:
% details on the method are provided directly as comments inside the
% function:
H = W; % contact height is the same of the width

[Cs_approx, Cd_approx] = Csd_calculatorAPPROX(Hrho_path, device_path, W, H);

string_to_be_displayed = sprintf('The approximated (parallel plates) source and drain
                capacitances resulted to be:     Cs = Cd = %g F', Cs_approx);
disp(string_to_be_displayed);
%% Quantum capacitance in function of applied Vds (fixed Vgs):
% The following function evaluates the quantum capacitance Cq(Vds) in
% function of the applied bias Vds, it is an approximate expression based
% on linearized approximation of Cd in the onsidered bias window. Details
% on employed method are in Cq_calulator.m

[Vg, Vds, Cq] = Cq_calculator(TEV_path, Hrho_path);
Cq = Cq*2; % SPIN DEGENERACY: DOS=DOS*2 !!
% Since Cq is evaluated in each Vds bias point, the returned array Vds is
% exactly the aaplied bias Vds array (used in ouput char estimation).
% Also Vg is exactly the same used for finding the current output char.

% Plot Cq(Vds)
figure
hold on
grid on
plot(Vds,Cq,'LineWidth',2.0);
title_string = sprintf('Average linearized quantum capacitance in
function of Vds for Vgs = %g V', Vg);
title(title_string);
xlabel('Vds (V)');
ylabel('Approximated Cq (F)');
set(gca,'LineWidth',2.0,'FontSize',15.0);

% Euilibrium Cq(Vds=0) is now extracted:
equilibrium_index = find( Vds==0 );
Cq_equi = Cq(equilibrium_index);

%% Ceq_0 equivalent electrostatic capacitance at equilibrium:
% Ceq_0 is the equivalent electrostatic capacitance of molFET at
% thermodynamic equilibrium (i.e. when contacts and molecule Fermi energy
% levels are aligned). It is the sum of:
%  > Cg, already evaluated from MOm physical geometry, thus indipendent from Vgs
%  > Cd_0, drain capacitance at equilibrium (i.e. Vds=0 V). Indeed it is a
%  non linear capacitance dependent on the applied bias.
%  > Cs_0, source capacitance at equilibrium (i.e. Vds=0 V). Indeed it is a
%  non linear capacitance dependent on the applied bias.
% Notice that if dl=dr then the electrode-molecule coupling is the same for drain
% and source, thus the assumption Cd=Cs holds, and Ceq_0 = 2*Cd_0 + Cg.

% The following function performs the Ceq_0 calculation:
[Ceq_0] = Ceq_0_calculator(Hrho_path, Cq_equi);

% Display it:
string_to_be_displayed = sprintf('The equivalent electrostatic capacitance at
equilibrium resulted to be:
                                        Ceq_0 = %g F', Ceq_0);
disp(string_to_be_displayed);
```

```
% equilibrium drain capacitance calculation (the same holds for Cs_0):
Cd_0 = (Ceq_0)/2;
% OR MAYBE:
%Cd_0 = (Ceq_0 - Cg)/2;
% OR MAYBE:
%Cd_0 = abs( (Ceq_0 - Cg)/2 );

Cs_0 = Cd_0; % true if voltage division factor eta is 0.5
%(i.e. if dl=dr and same contact material -and surface orientation- is used for S and D)

% Display it:
string_to_be_displayed = sprintf('The electrostatic drain/source capacitance at
equilibrium resulted to be:
                                        Cd_0 = Cs_0 = %g F', Cd_0);
disp(string_to_be_displayed);

%% Source and Drain Cs, Cd eletrostatic capacitances in function of applied Vds (fixed Vgs):
% The following function evaluates the source and drain capacitances
% Cs(Vds) and Cd(Vds) in function of the applied bias Vds, supposing Vgs
% fixed: thus the variations wrt Vgs are not considered and the channel
% potential variations due to Vds variations are supposed to be given by a
% capacitive divider: Cd/(Ces+Cq) where Ces=Cs+Cd+Cg.

[Vg_Csd, Vds_Csd, Cs, Cd] = Csd_calculator(Hrho_path, HS_path, device_path,
id_path, Cd_0, Cs_0, Cg, Cq);
% Since Cs and Cd are evaluated as differential capacitances (exploiting a
% sequence of small Vds variations) they are evaluated in different Vds
% bias points: Vds_Cds is different from Vds !!!
% In particular in Vds_Cds there are halfway values of Vds wrt to Vds array
% ones !! Thus in plotting Cd and Cs Vds_Csd should be used !!
figure
hold on
grid on
plot(Vds_Csd,Cs, Vds_Csd,Cd, 'LineWidth',2.0);
title_string = sprintf('Source and Drain capacitances in function of Vds
for Vgs = %g V', Vg_Csd);
title(title_string);
ylim=get(gca,'ylim');
xlim=get(gca,'xlim');
text(xlim(1)+0.1,ylim(2)-0.2e-19,'since voltage division factor is assumed
to be 0.5 they are equal');
legend('Cs', 'Cd');
xlabel('Vds (V)');
ylabel('Approximated Cs and Cd (F)');
set(gca,'LineWidth',2.0,'FontSize',15.0);

%% Gate resistance Rg estimation:
% details on the method are provided directly as comments inside the
% function:
% choose the dielectric material to be used as oxide layer (uncomment the
% desired one):
oxide='ZrO2';
%oxide='SiO2';
% The following function performs the Rg calculation:
[Rg] = Rg_calculator(Hrho_path, device_path, W, oxide);
% Display it:
string_to_be_displayed = sprintf('The approximated gate resistance resulted
to be:     Rg = %g ohm', Rg);
disp(string_to_be_displayed);

%% Source and Drain Rs, Rd resistances in function of applied Vds (fixed Vgs):
% The following function evaluates the source and drain resistances
% Rs(Vds) and Rd(Vds) in function of the applied bias Vds, supposing Vgs
% fixed. Further details on method are provided directly in the function
% code. It returns also the average coupling factors tauS, tauD (again in
% function of the applied bias).
[Vgs_fixed, Vds, Rs, Rd, tauS, tauD] = Rsd_calculator(Hrho_path, HS_path, TEV_path, Cq);
```

```
figure
hold on
grid on
plot(Vds,Rs, Vds,Rd, 'LineWidth',2.0);
title_string = sprintf('Source and Drain resistances in function of Vds
for Vgs = %g V', Vgs_fixed);
title(title_string);
legend('Rs', 'Rd');
xlabel('Vds (V)');
ylabel('Approximated average Rs and Rd (\Omega)');
set(gca,'LineWidth',2.0,'FontSize',15.0);
figure
hold on
grid on
plot(Vds,tauS, Vds,tauD, 'LineWidth',2.0);
title_string = sprintf('Source and Drain coupling factors in function of Vds
for Vgs = %g V', Vgs_fixed);
title(title_string);
legend('tauS', 'tauD');
xlabel('Vds (V)');
ylabel('Approximated average tauS and tauS (s)');
set(gca,'LineWidth',2.0,'FontSize',15.0);
%% Average Cs, Cd, Rs, Rd, Cq, tauS, tauD, are now evaluated:
Cq_AV = mean(Cq);
Cs_AV = mean(Cs);
Cd_AV = mean(Cd);
Rs_AV = mean(Rs);
Rd_AV = mean(Rd);
tauS_AV = mean(tauS);
tauD_AV = mean(tauD);
```

Listing 1: Main_molFET_parameters_RC.m

B.2 MatLab scripts for dynamic power estimation

```
clear all
close all
clc
%% This script estimates the dynamic power of a MOSFET:
% PHYSICAL PARAMETERS OF MOSFETS PROVIDED IN .scs file:
Wch=250e-9; %m, chosen at design time, it is an avarage specific for FA circuit
Lch=10e-9; %m, chosen at design time
Area=Lch*Wch;
eps0 = 8.8542e-12; % F/m
dtox = 4e-10; %from file .scs BSIM4
tox_e = 2.43e-09;  %from file .scs BSIM4
tox_p = tox_e - dtox;  %equation from file .scs BSIM4
eox = 3.9;

%oxide capacitance computed approximatively as parallel plate capacitance
Cox = ((eox*eps0)/tox_p)*Area;
Cg= Cox;
Vgs_ON= 1; %V, chosen at design time
 %useful if Cd, Cs contribution in dynamic power estimation are considered:

    % In order to evaluate the power (i.e. energy per unit time) it is
    % necessary to know how long the commutation takes, i.e. the time in
    % which the energy E=(1/2)*C*V^2 is transfered toward/from the
    % capacitances to charge/discharge them.
    % Pulse rising transient time length:
    risetime = 0.5e-3; % s
    % Pulse falling transient time length:
    falltime = risetime; % s
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Pdyn calculation:
    % The total exchanged energy per commutation (for a single charging or
    % discharging pphenomenon) is the sum of the effects on Cg, Cs, Cd.
    % Anyway for MOSFET we consider only Cg contribution:
    E = 0.5*Cg*(Vgs)^2

    % The total dissipated dynamic power during a complete work cycle, i.e.
    % during a full pulse (rising and falling edges, i.e. charging and
    % discharging) is twice the total exchanged energy per commutation 2*E(k)
    % over the total time interval during which that amount of energy is
    % exchanged:
    Pdyn = (2*E) / (risetime + falltime); % W

% Output display:
string_to_be_displayed = sprintf('\nTotal dissipated dynamic power per single pulse:
Pdyn = %g W', Pdyn');
disp(string_to_be_displayed);
```

Listing 2: Pdyn_FA_MOSFET.m

```
clear all
close all
clc
%% This script estimates the dynamic power of a molFET:
% Pdyn = Dynamic Power is dissipated in charging and discharging (stray)
% electrostatic capacitances of the device. It depends on the capacitange
% values and on the applied voltages.
% This script evaluates the dissipated dynamic power Pdyn per pulse, i.e.
% for a single charge plus a single discharge of stray capacitances, for a
% FA circuit
% 7 cases are considered: the first with no active inputs, the second with
% one single active input, the third with two active inputs and so on...
% THIS SCRIPT DOES NOT REQUIRE ANY USER'S INPUT A PART THE CHOICE OF THE
% MOLECULE TO BE USED !

% Select the desired molecule and load correct data:  %%%%%%%%%%%%%%%%%%%%%
string_to_be_displayed = sprintf('Please press:\n"1" for HDT molFET\n
                        "2" for OPE3 molFET\n
                        "3" for PCP molFET.\n');
disp(string_to_be_displayed);
molecule = input('>> ');

if molecule == 1 % HDT case
    % Useful constants:
    Cg = 5.14648e-18; % F
    Cs = 1.00269e-18; % F, average (over all cosidered Vds) source capacitance
    Cd = Cs; % if voltage division factor id 0.5 Cs=Cd
    % Cq = quantum capacitance is not used, indeed it is not an electrostatic
    % capacitance to be charged and discharged during commutation. It was useful
    % in correctly estimate Cs and Cd.
    VgsON = -2; % V
    VgsOFF = 0; % V
    deltaVgs = abs(VgsON - VgsOFF); % V
    % Vds is an array of 7 values, indeed it changes depending on how many ON
    % transistor are simultaneously activated: Vds(1) is the case of only 1
    % ON transistor, Vds(2) is the case of 2 ON inputs and so on...
    % (up to 7 ON transistors, i.e. Vds(7)).
    % For no active inputs no commutation happens and Pdyn=0 !
     Vds = [968.9e-3 948.5e-3 934e-3 923e-3 914.2e-3 906.9e-3 900.6e-3]; % V
    % In order to evaluate the power (i.e. energy per unit time) it is
    % necessary to know how long the commutation takes, i.e. the time in
    % which the energy E=(1/2)*C*V^2 is transfered toward/from the
    % capacitances to charge/discharge them.
    risetime = 0.5e-3;  % s, rising transient time length
    falltime = risetime; % s, falling transient time length

elseif molecule == 2 % OPE3 case
    Cg = 4.76568e-18; % F
    Cs = 8.85742e-19; % F, average (over all cosidered Vds) source capacitance
    Cd = Cs; % if voltage division factor id 0.5 Cs=Cd
    VgsON = -1; % V
    VgsOFF = +1; % V
    deltaVgs = abs(VgsON - VgsOFF); % V
    Vds = [709.4e-3 643.3e-3 595.5e-3 557.5e-3 526.4e-3 500.1e-3 477.4e-3]; % V
    risetime = 0.5e-3; % s, rising transient time length
    falltime = risetime; % s, falling transient time length

elseif molecule == 3 % OPE3 case
    Cg = 5.08519e-18 ; % F
    Cs = 1.09671e-18; % F, average (over all cosidered Vds) source capacitance
    Cd = Cs; % if voltage division factor id 0.5 Cs=Cd
    VgsON = -2; % V
    VgsOFF = +2; % V
    deltaVgs = abs(VgsON - VgsOFF); % V
    Vds = [878.4e-3 797.4e-3 740.6e-3 696.2e-3 660e-3 629.6e-3 603.8e-3]; % V
    risetime = 0.5e-3; % s, rising transient time length
    falltime = risetime; % s, falling transient time length

else % wrong case:
    string_to_be_displayed = sprintf('Wrong input!! =)\n');
    disp(string_to_be_displayed);
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% output data structures allocation:
%length(Vds) is the number of considered cases
Vdot_s = zeros(1,length(Vds));
Vcg = zeros(1,length(Vds));
E = zeros(1,length(Vds));
Pdyn = zeros(1,length(Vds));
Pdyn_singlepulse = zeros(1,length(Vds));

parallelism=7; %number of columns in the crossbar-array, i.e. redundancy
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Output display:
string_to_be_displayed = sprintf('\nTotal dissipated dynamic power per single pulse:');
disp(string_to_be_displayed);

% Pdyn calculation:
for k=1:length(Vds) % k is the index for the considered case
    % The voltage drop across Cs is Vds/2 since it is equal to Cd, anyway in
    % general it is given by a capacitive divider:
    Vdot_s(k) = (Cd/(Cs+Cd))*Vds(k);
        % The voltage drop across Cg is given by;
    Vcg(k) = deltaVgs - Vdot_s(k);
    % The total exchanged energy per commutation (for a single charging or
    % discharging phenomenon) is the sum of the effects on Cg, Cs, Cd:
    E(k) = 0.5*Cg*(Vcg(k))^2 + 0.5*Cs*(Vdot_s(k))^2 + 0.5*Cd*(Vds(k) - Vcg(k))^2; % J
    % The total dissipated dynamic power during a complete work cycle,
    %is twice the total exchanged energy per commutation 2*E(k) over
    %the total time interval during which that amount of energy is exchanged:
     Pdyn(k) = (2*E(k)) / (risetime + falltime); %Pdyn per pulse for a single transistor
     Ptot(k) = Pdyn(k)*(k*parallelism); %Total Pdyn per pulse for all transistors ON
     % Display the result:
    string_to_be_displayed = sprintf('case of %g simultaneously active inputs:
        Pdyn = %g W for a single transistor,
         Ptot= %g W total', k, Pdyn(k), Ptot(k));
    disp(string_to_be_displayed);
end
% Average over all possible cases:
Pdyn_av = mean(Pdyn);
% Display it:
string_to_be_displayed = sprintf('\nAverage Pdyn per pulse: Pdyn_av = %g W', Pdyn_av);
disp(string_to_be_displayed);
```

Listing 3: Pdyn_FA_molFET.m

B.3 MatLab scripts for time constant estimation

```
%% This script estimates the time constant (step response) of a mol/MOS FET:
clear all
close all
clc
% Once the simulations on Cadence Virtuoso are concluded, please export data
% in .csv format file
DataTable = readtable('Vc_Cg_OPE3.csv');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%    NO  MORE  USER's  INPUTS   %%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
DataNames = DataTable.Properties.VariableNames;
Data = table2array(DataTable);
DataSize = size(Data);

%% Plot data and choice of curve:
string_to_be_displayed = sprintf('\nPlots are generated automatically.\n');
disp(string_to_be_displayed);

for index=1:2:DataSize(2)
figure
hold on
grid on
plot(Data(:,index),Data(:,index+1),'LineWidth',2.0);
VariableName = char(DataNames(index));
VariableName = VariableName(1:end-1);
title_string = sprintf('Imported data: %s', VariableName);
title(title_string);
xlabel('time (s)');
ylabel('Voltages (V) or Currents (A)');
set(gca,'LineWidth',2.0,'FontSize',15.0);
string_to_be_displayed = sprintf('Press "%g" to proceed
in time constant estimation for
                                the "%s" quantity.', index, VariableName);
disp(string_to_be_displayed);
end


%%%%%%% INPUT INDEX AND PROCEED WITH TIME CONSTANT ESTIMATION:
index = input('>> ');

if (index > DataSize(2)) % wrong value
    string_to_be_displayed = sprintf('Value "%g" is NOT allowed. Please restart.\n', index);
    disp(string_to_be_displayed);


elseif (rem(index, 2) == 0) % wrong value, since introduced index is even
    string_to_be_displayed = sprintf('Value "%g" is NOT allowed. Please restart.\n', index);
    disp(string_to_be_displayed);
```

```
elseif (rem(index, 2) == 1)
    VariableName = char(DataNames(index));
    VariableName = VariableName(1:end-1);
    string_to_be_displayed = sprintf('You have chosen: %s', VariableName);
    disp(string_to_be_displayed);
    string_to_be_displayed = sprintf('Time constant in now found by means
    of a linear interpolation.');
    disp(string_to_be_displayed);
    string_to_be_displayed = sprintf('Now a suitable time interval, for
    exponential transient
                                linear interpolation, should be selected.');
    disp(string_to_be_displayed);
    string_to_be_displayed = sprintf('Please insert an initial time instant
    for interpolation of transient: ');
    disp(string_to_be_displayed);
    InitialTimeTrans = input('>> ');
    string_to_be_displayed = sprintf('Now insert a final time instant for
    interpolation of transient: ');
    disp(string_to_be_displayed);
    FinalTimeTrans = input('>> ');
    string_to_be_displayed = sprintf('Now insert an initial time instant
    for interpolation of
                                    regime steady state region: ');
    disp(string_to_be_displayed);
    InitialTimeSS = input('>> ');
    string_to_be_displayed = sprintf('Now insert a final time instant
    for interpolation of
                                    regime steady state region: ');
    disp(string_to_be_displayed);
    FinalTimeSS = input('>> ');


    if (FinalTimeTrans <= InitialTimeTrans | FinalTimeSS <= InitialTimeSS) % error
        string_to_be_displayed = sprintf('Input error: Usually a final time
        instant should be greater
                                            than an initial one.');
        disp(string_to_be_displayed);

    elseif (InitialTimeSS <= FinalTimeTrans) % error
        string_to_be_displayed = sprintf('Input error: Usually the
        steady state is achieved after a transient.');
        disp(string_to_be_displayed);
    else
        NewTimeAxis = linspace(Data(1,index), Data(end,index), 10000);
        timeindex = find( InitialTimeTrans < Data(:,index) & Data(:,index) < FinalTimeTrans );
        %first order polynomial fit (least-squares is used) for exponential transient:
        Pexp = polyfit(Data(timeindex,index), Data(timeindex,index+1), 1);
        % Evaluate the fitted polynomial Pexp in the whole time interval:
        StraightLine = polyval(Pexp, NewTimeAxis);

        SSindex = find( InitialTimeSS < Data(:,index) & Data(:,index) < FinalTimeSS );
        % first order polynomial fit (least-squares is used) for SS=steady state:
        Pss = polyfit(Data(SSindex,index), Data(SSindex,index+1), 1);
        % Evaluate the fitted polynomial Pss in the whole time interval:
        StraightLineSS = polyval(Pss, NewTimeAxis);

        %find the intersection point (i.e. time constant) between the straight lines:
        error = abs(StraightLine - StraightLineSS); % error array
        minimum_error = min(error); % minimum error corresponds to interstection point
        TimeConstIndex = find(error == minimum_error);
        % the value for which the error is minimum is the intersection point, i.e. time constant
```

```
        % The found time value is saved and then displayed:
        TimeConstant = NewTimeAxis(TimeConstIndex) - InitialTimeTrans;
        % given by the difference the time value in intersection point and
        transient initial time instant
        string_to_be_diplayed = sprintf('\nFound time constant value: Tconst
        =%g s', TimeConstant);
        disp(string_to_be_diplayed);
        string_to_be_diplayed = sprintf('Conventional transient duration: 7*Tconst
        = %g s', 7*TimeConstant);
        disp(string_to_be_diplayed);
        string_to_be_diplayed = sprintf('Min period: 2*7*Tconst = 14*Tconst
        = %g s', 14*TimeConstant);
        disp(string_to_be_diplayed);
        string_to_be_diplayed =sprintf('Max operating frequency: (14*Tconst)^-1
        = %g Hz', (14*TimeConstant)^-1 );
        disp(string_to_be_diplayed);

        % Plot:
        figure
        grid on
        hold on
        plot(Data(:,index),Data(:,index+1),'LineWidth',2.0);
        plot(NewTimeAxis, StraightLine, 'Color', [0 0 0]/255, 'LineWidth', 1.25);
        % black curve
        plot(NewTimeAxis, StraightLineSS, 'Color', [130 130 130]/255, 'LineWidth', 1.25);
        %gray curve
        % plot the intersection point:
        plot(NewTimeAxis(TimeConstIndex), StraightLine(TimeConstIndex),'or','MarkerSize',10);
        VariableName = char(DataNames(index));
        VariableName = VariableName(1:end-1);
        title_string = sprintf('Time constant evaluation: %s(t)', VariableName);
        title(title_string);
        xlabel('time (s)');
        ylabel('Voltage (V) or Current (A)');
        set(gca,'fontsize', 14);

    end

else
    string_to_be_displayed = sprintf('Something went wrong during input acquisition.
    Please restart.\n');
    disp(string_to_be_displayed);

end
```

Listing 4: molFET_RC_TimeConstantStraightLine.m

```
%% This script estimates the time constant (step response) of a mol/MOS FET:
clear all
close all
clc
% Once the simulations on Cadence Virtuoso are conluded, please export data
% in .csv format file
DataTable = readtable('molFET_HDT_IntrinsicTimeVd_Cg.csv');
%%%%%%%%%%%%%%%%    NO  MORE  USER's  INPUTS   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
DataNames = DataTable.Properties.VariableNames;
Data = table2array(DataTable);
DataSize = size(Data);
%% Plot data and choice of curve:
string_to_be_displayed = sprintf('\nPlease insert the tolerance you want to use.\n
                It is proportional to the maximum signal variation tolerated
    in steady state.\nTolerance:');
disp(string_to_be_displayed);
tolerance = input('>> ');

for index=1:2:DataSize(2)
figure
hold on
grid on
plot(Data(:,index),Data(:,index+1),'LineWidth',2.0);
VariableName = char(DataNames(index));
VariableName = VariableName(1:end-1);
title_string = sprintf('Imported data: %s', VariableName);
title(title_string);
xlabel('time (s)');
ylabel('Voltages (V) and Currents (A)');
set(gca,'LineWidth',2.0,'FontSize',15.0);
%%%%%%  DERIVATIVE CALCULATION:
DataDerivative = diff(Data(:,index+1)); % signal (time) derivative
DataDerivativeTimeAxis = Data(1:end-1,index); % adjusting time scale
% If the derivative is lesser than tolerance then it is assumed the
% transient is estinguished, if instead it is greater then the transient is going on:
TimeIndeces = find(abs(DataDerivative) >= tolerance);
% i.e. TimeIndeces stores the indeces of time array for which there is the transient
ConsideredTimeInstants = DataDerivativeTimeAxis(TimeIndeces);
% time instants for which DataDerivative >= tolerance (i.e. still transient)
TimeConstant = max(ConsideredTimeInstants) - min(ConsideredTimeInstants);
% time constant, s (time array is always increasing positive)
% Plot/Display it (only if it is found a strictly positive interval is found:
 if isempty(TimeIndeces) % not possible display it (emtpy vetor):
    string_to_be_displayed = sprintf('The evaluated time constant for "%s" was not found,
    thus it is not displayed.', VariableName);
    disp(string_to_be_displayed);
 else % in this case it can be displayed:
    string_to_be_displayed = sprintf('The evaluated time constant for "%s" data is: %g s',
                                     VariableName, TimeConstant);
    disp(string_to_be_displayed);
    % plot the transient starting point:
    plot(Data(TimeIndeces(1),index), Data(TimeIndeces(1),index+1),'or','MarkerSize',10);
    % plot the transient starting point:
    plot(Data(TimeIndeces(end),index), Data(TimeIndeces(end),index+1),'or','MarkerSize',10);
 end
end
string_to_be_displayed = sprintf('\nThe evaluated Time intervals can be interpreted like:
                    \na) exponential transient time constants,
                    \nor\nb) time intervals needed to estiguish the whole transient.
                    \nDepending on selected tolerance value you can get one of these
                    two values.\n');
disp(string_to_be_displayed);
string_to_be_displayed = sprintf('Now please check with the help of the graphs if the
                results are satisfactory or not.\nThis depends also on how you interpret
                these time intervals (case a) or (b) ).\nEventually restart the execution
                with an updated tolerance if you are not satisfied.  =)');
disp(string_to_be_displayed);
```

Listing 5: molFET_RC_TimeConstantDerivativesMethod.m

Bibliography

[1] Deloitte analysis, 2020. https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/technology-media-telecommunications/deloitte-cn-tmt-semiconductors-the-next-wave-en-190422.pdf.

[2] Samsung makes 1TB flash eUFS module, 2019. https://www.electronicsweekly.com/news/business/samsung-makes-1tb-flash-module-2019-01/.

[3] Cerebras, 2019. https://fuse.wikichip.org/news/3010/a-look-at-cerebras-wafer-scale-engine-half-square-foot-silicon-chip/

[4] ASML ANNUAL REPORT 2013. https://www.sec.gov/Archives/edgar/data/937966/000119312514046822/d546896d20f.htm.

[5] International Roadmap for Devices and Systems (IRDS) 2020 Edition. https://irds.ieee.org/editions/2020.

[6] Mark Ratner. A brief history of molecular electronics. *Nature Nanotechnology*, 8(6):378–381, Jun 2013.

[7] JAMES M. TOUR, WILLIAM A. REINERTH, LEROY JONES II, TIMOTHY P. BURGIN, CHONG-WU ZHOU, C. J. MULLER, M. R. DESHPANDE, and MARK A. REED. Recent advances in molecular scale electronics. *Annals of the New York Academy of Sciences*, 852(1):197–204, 1998.

[8] Dong Xiang, Xiaolong Wang, Chuancheng Jia, Takhee Lee, and Xuefeng Guo. Molecular-scale electronics: From concept to function. *Chemical Reviews*, 116(7):4318–4440, 2016. PMID: 26979510.

[9] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel. Surface studies by scanning tunneling microscopy. *Phys. Rev. Lett.*, 49:57–61, Jul 1982.

[10] G. Binnig, C. F. Quate, and Ch. Gerber. Atomic force microscope. *Phys. Rev. Lett.*, 56:930–933, Mar 1986.

[11] Valentin Dubois, Shyamprasad N. Raja, Pascal Gehring, Sabina Caneva, Herre S. J. van der Zant, Frank Niklaus, and Göran Stemme. Massively parallel fabrication of crack-defined gold break junctions featuring sub-3

nm gaps for molecular devices. *Nature Communications*, 9(1):3433, Aug 2018.

[12] S. Datta. Non-equilibrium green's function (negf) method: a different perspective. In *2015 International Workshop on Computational Electronics (IWCE)*, pages 1–6, 2015.

[13] Na Xin and Xuefeng Guo. Catalyst: The renaissance of molecular electronics. *Chem*, 3(3):373 – 376, 2017.

[14] Junyang Liu, Xiaotao Zhao, Jueting Zheng, Xiaoyan Huang, Yongxiang Tang, Fei Wang, Ruihao Li, Jiuchan Pi, Cancan Huang, Lin Wang, Yang Yang, Jia Shi, Bing-Wei Mao, Zhong-Qun Tian, Martin R. Bryce, and Wenjing Hong. Transition from tunneling leakage current to molecular tunneling in single-molecule junctions. *Chem*, 5(2):390 – 401, 2019.

[15] Hyunwook Song, Youngsang Kim, Yun Hee Jang, Heejun Jeong, Mark A. Reed, and Takhee Lee. Observation of molecular orbital gating. *Nature*, 462(7276):1039–1043, Dec 2009.

[16] Chuancheng Jia, Agostino Migliore, Na Xin, Shaoyun Huang, Jinying Wang, Qi Yang, Shuopei Wang, Hongliang Chen, Duoming Wang, Boyong Feng, Zhirong Liu, Guangyu Zhang, Da-Hui Qu, He Tian, Mark A. Ratner, H. Q. Xu, Abraham Nitzan, and Xuefeng Guo. Covalently bonded single-molecule junctions with stable and reversible photoswitched conductivity. *Science*, 352(6292):1443–1445, 2016.

[17] Brian Capozzi, Jianlong Xia, Olgun Adak, Emma J. Dell, Zhen-Fei Liu, Jeffrey C. Taylor, Jeffrey B. Neaton, Luis M. Campos, and Latha Venkataraman. Single-molecule diodes with high rectification ratios through environmental control. *Nature Nanotechnology*, 10(6):522–527, Jun 2015.

[18] Huseyin Atesci, Veerabhadrarao Kaliginedi, Jose A. Celis Gil, Hiroaki Ozawa, Joseph M. Thijssen, Peter Broekmann, Masa-aki Haga, and Sense Jan van der Molen. Humidity-controlled rectification switching in ruthenium-complex molecular junctions. *Nature Nanotechnology*, 13(2):117–121, Feb 2018.

[19] Jonathan E. Green, Jang Wook Choi, Akram Boukai, Yuri Bunimovich, Ezekiel Johnston-Halperin, Erica DeIonno, Yi Luo, Bonnie A. Sheriff, Ke Xu, Young Shik Shin, Hsian-Rong Tseng, J. Fraser Stoddart, and James R. Heath. A 160-kilobit molecular electronic memory patterned at 1011 bits per square centimetre. *Nature*, 445(7126):414–417, Jan 2007.

[20] C Joachim, JK Gimzewski, and A Aviram. Electronics using hybrid-molecular and mono-molecular devices. *Nature*, 408(6812):541â548, November 2000.

[21] Arieh Aviram and Mark A. Ratner. Molecular rectifiers. *Chemical Physics Letters*, 29(2):277 – 283, 1974.

[22] George M. Whitesides and Mila Boncheva. Beyond molecules: Self-assembly of mesoscopic and macroscopic components. *Proceedings of the National Academy of Sciences*, 99(8):4769–4774, 2002.

[23] EladâD. Mentovich, Itshak Kalifa, Alexander Tsukernik, Ariel Caster, Natalie Rosenberg-Shraga, Hanit Marom, Michael Gozin, and Shachar Richter. Multipeak negative-differential-resistance molecular device. *Small*, 4(1):55–58, 2008.

[24] Pramod Reddy, Sung-Yeon Jang, Rachel A. Segalman, and Arun Majumdar. Thermoelectricity in molecular junctions. *Science*, 315(5818):1568–1571, 2007.

[25] Stefan Thiele, Franck Balestro, Rafik Ballou, Svetlana Klyatskaya, Mario Ruben, and Wolfgang Wernsdorfer. Electrically driven nuclear spin resonance in single-molecule magnets. *Science*, 344(6188):1135–1138, 2014.

[26] Tao Wang and Christian A. Nijhuis. Molecular electronic plasmonics. *Applied Materials Today*, 3:73 – 86, 2016.

[27] James M Tour. *Molecular Electronics*. WORLD SCIENTIFIC, 2003.

[28] QuantumATK version Q-2019.12, Synopsys QuantumATK (www.synopsys.com/silicon/quantumatk.html).

[29] Supriyo Datta. *Quantum transport: atom to transistor*. Cambridge University Press, 2005.

[30] Marc Baldo. *Introduction to Nanoelectronics*. MIT OpenCourseWare Publication, 2011.

[31] Ismael Rattalino, Paolo Motto, Gianluca Piccinini, and Danilo Demarchi. A new validation method for modeling nanogap fabrication by electromigration, based on the resistanceâvoltage (râv) curve analysis. *Physics Letters A*, 376(30):2134 – 2140, 2012.

[32] Bingqian Xu and Nongjian J. Tao. Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science*, 301(5637):1221–1223, 2003.

[33] M. Kiguchi, R. Stadler, I. S. Kristensen, D. Djukic, and J. M. van Ruitenbeek. Evidence for a single hydrogen molecule connected by an atomic chain. *Phys. Rev. Lett.*, 98:146802, Apr 2007.

[34] Linda Zotti, Beatrice Bednarz, Juan Hurtado-Gallego, Damien Cabosart, Gabino Rubio-Bollinger, Nicolas Agrait, and Herre Zant. Can one define the conductance of amino acids? *Biomolecules*, 9:580, 10 2019.

[35] Takanori Harashima, Yusuke Hasegawa, Satoshi Kaneko, Manabu Kiguchi, Tomoya Ono, and Tomoaki Nishino. Highly reproducible formation of polymer singleâmolecule junction for wellâdefined current signal. *Angewandte Chemie International Edition*, 58, 04 2019.

[36] Elad Mentovich, Bogdan Belgorodsky, and Shachar Richter. Resolving the mystery of the elusive peak: Negative differential resistance in redox proteins. *The Journal of Physical Chemistry Letters*, 2:1125â1128, 04 2011.

[37] Limin Xiang, Julio L. Palma, Yueqi Li, Vladimiro Mujica, Mark A. Ratner, and Nongjian Tao. Gate-controlled conductance switching in dna. *Nature Communications*, 8(1):14471, Feb 2017.

[38] Paven Thomas Mathew and Fengzhou Fang. Advances in molecular electronics: A brief review. *Engineering*, 4(6):760 – 771, 2018.

[39] Wenjun Xu, Edmund Leary, Songjun Hou, Sara Sangtarash, Teresa GonzÃ¡lez, Gabino Rubio-Bollinger, Qingqing Wu, Hatef Sadeghi, Lara Tejerina, Kirsten Christensen, NicolÃ¡s AgraÃ¯t, Simon Higgins, Colin Lambert, Richard Nichols, and Harry Anderson. Unusual lengthâdependence of conductance in cumulene molecular wire. *Angewandte Chemie*, 04 2019.

[40] Yanxi Zhang, Saurabh Soni, Thijs Krijger, Pavlo Gordiichuk, Xinkai Qiu, Gang Ye, H. Jonkman, Andreas Herrmann, Karin Zojer, Egbert Zojer, and Ryan Chiechi. Tunneling probability increases with distance in junctions comprising self-assembled monolayers of oligothiophenes. *Journal of the American Chemical Society*, 140, 10 2018.

[41] Thijs Stuyver, Tao Zeng, Yuta Tsuji, Paul Geerlings, and Frank De Proft. Diradical character as a guiding principle for the insightful design of molecular nanowires with an increasing conductance with length. *Nano Letters*, 18(11):7298–7304, Nov 2018.

[42] A. Zahir, S. A. A. Zaidi, A. Pulimeno, M. Graziano, D. Demarchi, G. Masera, and G. Piccinini. Molecular transistor circuits: From device model to circuit simulation. In *2014 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pages 129–134, 2014.

[43] Wanda Andreoni, Alessandro Curioni, and Henrik GrÃ¶nbeck. Density functional theory approach to thiols and disulfides on gold: Au(111) surface and clusters. *International Journal of Quantum Chemistry*, 80(4â5):598–608, 2000.

[44] N. J. Tao. Electron transport in molecular junctions. *Nature Nanotechnology*, 1(3):173–181, Dec 2006.

[45] David L. Cooper, Joseph Gerratt, and Mario Raimondi. The electronic structure of the benzene molecule. *Nature*, 323(6090):699–701, Oct 1986.

[46] Katharine E. Linton, Mark A. Fox, Lars-Olof PÃ¥lsson, and Martin R. Bryce. Oligo(p-phenyleneethynylene) (ope) molecular wires: Synthesis and length dependence of photoinduced charge transfer in opes with triarylamine and diaryloxadiazole end groups. *Chemistry â A European Journal*, 21(10):3997–4007, 2015.

[47] Mariagrazia Graziano, A. Zahir, Ahmed Mahmoud, Azzurra Pulimeno, Gianluca Piccinini, and Paolo Lugli. Hierarchical modeling of opv-based crossbar architectures. 08 2014.

[48] Cornelius Thiele, Lukas Gerhard, Thomas Eaton, David Torres, Marcel Mayor, Wulf Wulfhekel, Hilbert von LÃ¶hneysen, and Maya Lukas. Stm study of oligo(phenylene-ethynylene)s. *New Journal of Physics*, 17, 05 2015.

[49] Chuancheng Jia, Marjan Famili, Marco Carlotti, Yuan Liu, Peiqi Wang, Iain M. Grace, Ziying Feng, Yiliu Wang, Zipeng Zhao, Mengning Ding, Xiang Xu, Chen Wang, Sung-Joon Lee, Yu Huang, Ryan C. Chiechi, Colin J. Lambert, and Xiangfeng Duan. Quantum interference mediated vertical molecular tunneling transistors. *Science Advances*, 4(10), 2018.

[50] Guowen Peng, Mikkel Strange, Kristian S. Thygesen, and Manos Mavrikakis. Conductance of conjugated molecular wires: Length dependence, anchoring groups, and band alignment. *The Journal of Physical Chemistry C*, 113(49):20967–20973, Dec 2009.

[51] Ahmed Mahmoud and Paolo Lugli. First-principles study of a novel molecular rectifier. *Nanotechnology, IEEE Transactions on*, 12:719–724, 09 2013.

[52] Fu Xiaoxiao, Rui Zhang, Guang-Ping Zhang, and Zong-Liang Li. Rectifying properties of oligo(phenylene ethynylene) heterometallic molecular junctions: Molecular length and side group effects. *Scientific reports*, 4:6357, 09 2014.

[53] Linda A. Zotti, Thomas Kirchner, Juan-Carlos Cuevas, Fabian Pauly, Thomas Huhn, Elke Scheer, and Artur Erbe. Revealing the role of anchoring groups in the electrical conduction through single-molecule junctions. *Small*, 6(14):1529–1535, 2010.

[54] Roar SÃ¸ndergaard, Sebastian Strobel, Eva Bundgaard, Kion Norrman, Allan Hansen, Edgar Albert, Gyorgy Csaba, Paolo Lugli, Marc Tornow, and Frederik Krebs. Conjugated 12 nm long oligomers as molecular wires in nanoelectronics. *Journal of Materials Chemistry - J MATER CHEM*, 19, 06 2009.

[55] Muhammed Schukfeh, Kristian Storm, Ahmed Mahmoud, Roar SÃ¸ndergaard, Anna Szwajca, Allan Hansen, Peter Hinze, Thomas Weimann, Sofia Svensson, Achyut Bora, Kimberly Dick, Claes Thelander, Frederik Krebs, Paolo Lugli, Lars Samuelson, and Marc Tornow. Conductance enhancement of inas/inp heterostructure nanowires by surface functionalization with oligo(phenylene vinylene)s. *ACS nano*, 7, 04 2013.

[56] Stanislav Tsoi, Igor Griva, Scott Trammell, Amy Blum, Joel Schnur, and Nikolai Lebedev. Electrochemically controlled conductance switching in a single molecule: Quinone-modified oligo(phenylene vinylene). *ACS*

nano, 2:1289–95, 07 2008.

[57] Ahmed Mahmoud and Paolo Lugli. Atomistic study on dithiolated oligophenylenevinylene gated device. *Journal of Applied Physics*, 116:204504, 11 2014.

[58] Mark W. H. Hoorens, Miroslav Medved', Adèle D. Laurent, Mariangela Di Donato, Samuele Fanetti, Laura Slappendel, Michiel Hilbers, Ben L. Feringa, Wybren Jan Buma, and Wiktor Szymanski. Iminothioindoxyl as a molecular photoswitch with 100 nm band separation in the visible range. *Nature Communications*, 10(1):2390, Jun 2019.

[59] Carson Bruns and J. Stoddart. *Molecular Switches and Machines with Mechanical Bonds*, pages 555–733. 10 2016.

[60] MartinâB. Duriska, SuzanneâM. Neville, Boujemaa Moubaraki, JohnâD. Cashion, GregoryâJ. Halder, KarenaâW. Chapman, ChÃ©rif Balde, Jean-FranÃ§ois LÃ©tard, KeithâS. Murray, CameronâJ. Kepert, and StuartâR. Batten. A nanoscale molecular switch triggered by thermal, light, and guest perturbation. *Angewandte Chemie International Edition*, 48(14):2549–2552, 2009.

[61] Fei Lu, Qi Qin, Yuan Li, and Jiezhi Chen. Computational design of molecular transistor with van der waals gating. *Applied Physics Express*, 13, 07 2020.

[62] D. Hou and J. H. Wei. The difficulty of gate control in molecular transistors, 2011.

[63] Sujit S. Datta, Douglas R. Strachan, and A. T. Charlie Johnson. Gate coupling to nanoscale electronics. *Phys. Rev. B*, 79:205404, May 2009.

[64] Jonas Fransson, O. Bengone, Andreas Larsson, and Jim Greer. A physical compact model for electron transport across single molecules. *Nanotechnology, IEEE Transactions on*, 5:745 – 749, 12 2006.

[65] Dong Xiang, Dongku Kim, Hyunhak Jeong, Takhee Lee, Yongjin Cheng, Qingling Wang, and Dirk Mayer. Three-terminal single-molecule junctions formed by mechanically controllable break junctions with side gating. *Nano letters*, 13, 05 2013.

[66] A. Mahmoud and P. Lugli. Toward circuit modeling of molecular devices. *IEEE Transactions on Nanotechnology*, 13(3):510–516, 2014.

[67] A. Zahir, A. Pulimeno, D. Demarchi, M. Graziano, G. Piccinini, A. Mahmoud, and P. Lugli. Modular framework for molecular-fet device-to-circuit modeling. In *2015 IEEE 15th International Conference on Nanotechnology (IEEE-NANO)*, pages 156–159, 2015.

[68] J.M. Rabaey, A.P. Chandrakasan, and B. Nikolić. *Digital Integrated Circuits: A Design Perspective*. Prentice Hall electronics and VLSI series. Prentice Hall, 2003.

[69] Giulia Santoro, Giovanna Turvani, and Mariagrazia Graziano. New

logic-in-memory paradigms: An architectural and technological perspective. *Micromachines*, 10:368, 05 2019.

[70] H. Jarollahi, N. Onizawa, V. Gripon, N. Sakimura, T. Sugibayashi, T. Endoh, H. Ohno, T. Hanyu, and W. J. Gross. A nonvolatile associative memory-based context-driven search engine using 90 nm cmos/mtj-hybrid logic-in-memory architecture. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 4(4):460–474, 2014.

[71] K. Yang, R. Karam, and S. Bhunia. Interleaved logic-in-memory architecture for energy-efficient fine-grained data processing. In *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 409–412, 2017.

[72] Mario Cofano, Marco Vacca, Giulia Santoro, Giovanni Causapruno, Giovanna Turvani, and Mariagrazia Graziano. Exploiting the logic-in-memory paradigm for speeding-up data-intensive algorithms. *Integration*, 66:153 – 163, 2019.

[73] M. Vacca, Y. Tavva, A. Chattopadhyay, and A. Calimera. Logic-in-memory architecture for min/max search. In *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pages 853–856, 2018.

[74] Ahmed Mahmoud, Alessio Gagliardi, and Paolo Lugli. Atomistic study of three-leg molecular devices. *Organic Electronics*, 24:37 – 42, 2015.

[75] Adam Jaros, Esmaeil Farajpour Bonab, Michal Straka, and Cina Foroutan-Nejad. Fullerene-based switching molecular diodes controlled by oriented external electric fields. *Journal of the American Chemical Society*, 141(50):19644–19654, Dec 2019.

[76] QuantumWise ATK: Building a molecular junction. https://docs.quantumwise.com/tutorials/molecular_junction/molecular_junction.html.

[77] QuantumWise ATK: Molecular Device. https://docs.quantumwise.com/tutorials/molecular_device/molecular_device.html.

Acknowledgements

I would like to thank my colleague and dear friend Fabrizio Mo for his contribution in some topics of this thesis, fruit of a previous teamwork for the workshop “Innovative System” held at Politecnico di Torino in a.y. 2019/20.

I would also like to thank Ph.D student Y. Ardesi, professor M. Vacca for their useful advices and kind availability and professors A. Canu, F. Tanda, G. Piccinini, M. Graziano for their inspirational examples which have defined who I am today.

Last but not least, thanks to my lovely family and roommates for their daily support, despite the distance, in this absurd and restless year 2020.