Politecnico di Torino

ETF College

Nanoelectronics Systems Notes

Mariagrazia Graziano, Gianluca Piccinini

November 14, 2023

Version V1.2

Preface

The present document is a draft version intended as an integration, where possible, of the SLIDES used during the lectures of the Nanoelectronics Systems course: please refer to the SLIDES as main reference document for the exam preparation.
Large part of the present document is derived by other documents as Master Theses of students that worked on the course subjects during the thesis project, to which goes my gratitide. Some other parts are derived and integrated from students' notes whose contribution is as well extremely important. Other parts of this document are derived by "external" documents as Journal papers and Master or PhD theses and books of researchers and students that contributed to the scientific field touched by the course subjects.

I am aware that this document will need a lot of refinements before becoming a good reference book, and I am since now thankful for the reader's understanding and patience, and grateful to those who will be kind and interested enought to contribute underlining errors, missing points and suggesting possible additions.

I hope, in any case, that this document could be useful for the students approaching the study of this course subects.

Though the documents used as sources will be cited wherever possible, I would like to mention some of the most importat contributions to this document: though many parts have been elaborated or adjusted, the starting contribution remains. So I would like to thank in particular:
Gianluca Piccinini, Chiara Elfi Spano [1], Fabrizio Mo [2], Ali Zahir [?], Yuri Ardesi [3], Giuliana Beretta [4], Ruyu Wang [?], Fabrizio Riente [5], Giovanna Turvani [6], Marco Vacca [7], Salvatore Vistato, Chiara Cannavo' [8], Paolo Pagliarulo [9], Elio Pio Reveruzzi [10], Giorgio Alemanno [11] Mario Simoni, Luca Gnoli, Umberto Garlando [?], Fabrizio Cairo [?], Annachiara Apollonio, Elisa Plotzner, Taulant Marpepa, Antonino Ferrara, Alessandro Morè, Vincenzo Di Lorenzo, ... to be continued.....

- **Version V1.0**: chapters 1,2,3,4,5, 17,18,19,20 in draft with some missing parts. To be rivised.
- **Version V1.1**: chapters 1,2,3,4,5,12,13,14,15,16,17,18,19,20 in draft with some missing parts. To be rivised.
- **Version V1.2**: chapters 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20 in draft with some missing parts. To be rivised.

The index will be updated as new versions of the chapters will be refined and completed.

Contents

CHAPTER 1

Introduction: the driving force

Nano-systems have seen significant improvement: when we talk about new technologies we refer mainly to nanotechnologies. However, when a new technology can be really accepted in the industrial relaity?

Currently the degree of acceptance of a new technology is really poor. This is mainly because industry is heavily reliant on silicon. The industry have invested so much money, time and effort in silicon that it is becoming almost impossible to move abruptly away from the silicon technology. However, if we take a cursory look at the past we can be pretty confident that new nano-technology, under certain conditions, are expected to have an impressive success story in the future. History have taugth us that out of the box thinking (***Keypoint***) have always resulted a the break-through in the industry.

1.1 A bit of historical perspective

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 [12].

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 (see Figure1.1). It contained about 18000 vacuum tubes, which were fragile, very power-demanding, and produced huge amount of heat. Moreover the ENIAC occupied about $170\,\mathrm{m}^2$ 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.

The next important step in the evolution was the invention of (**BJT**) bipolar junction transistor(see Figure1.2) by Shockley, Bardeen and Brattain in 1949 (they won the Nobel prize for this). The first transistor was invented when Shockley, Bardeen and Brattain were studying new material called semiconductor. They discovered that there was an increase in the current from the source to the drain. This structure called the bi-polar transistor became very popular in the industry. 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. The sad side of the story was that integration was not known in the industry.

In 1958 the first integrated circuit has been fabricated by Jack Kilby at Texas Instruments. This integrated circuit consisted of 6 bipolar trasistor all solded into a single plateform as shown in (see Figure 1.3) The MOS was already in exitance but it had some challenges mainly because at that time the technological processes necessary in order to result in the integration was completely absent. In

Figure 1.1: The first digital arithmetic machine.

Figure 1.2: The first bipolar junction transistor.

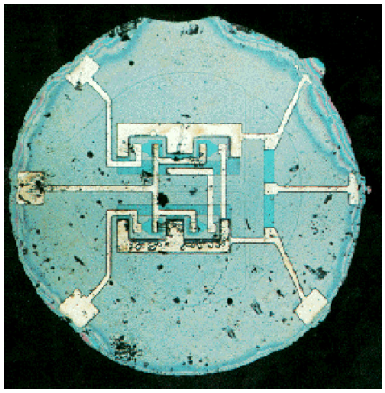


Figure 1.3: The first integrated circuit.

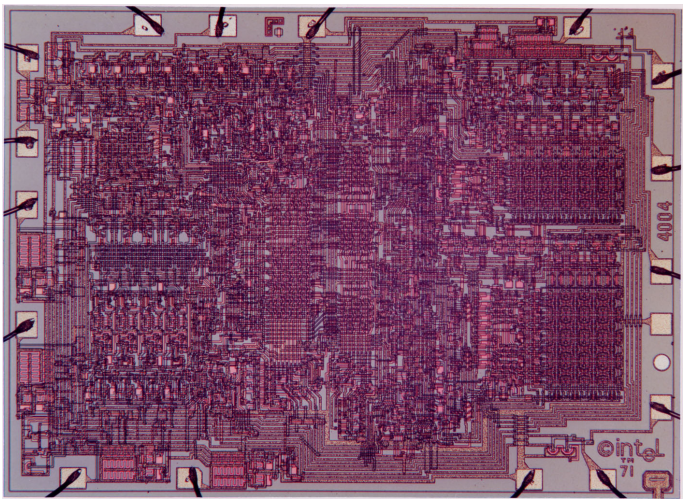


Figure 1.4: The first integrated microprocessor: 4004

1960 the MOS was invented in which the gate was made up of a heavily doped polysilicon. Federico Faggin in 1968 worked at Fairchild Semiconductor (Today called Intel) as an intern, at that time intel was the world leading producer of memories. Faggin work on a project for a japanese company who wanted to develope the architecture for a microprocessor. In 1971 he joined Intel, where he was able to exploit the new silicon gate technology to define a unique intergrated circuit which become the first microprocessor architecture the 4004 in 1971. The micro-graph is reported in Figure 1.4.

When the 4004 integrated microprocessor was developed in intel the company (fairchild) main specialization was in memory but we now know that intel have become the world lastest vendor of microprocessors. Federico Faggin found a way to integrate the MOS transistor using the polysilicon as it's gate technology. He is also credited with the invention of tourchpad at Zilog. The 4004 architecture consisted of 2300 transistor (NMOS only with a $10\mu m$ CMOS technology), was based on both static and dynamic logic, the area was 3mm X 4mm, the CPU was a 4 bit one, and could run around 100kHz. 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 now, semiconductor industry has grown by leaps and bounds. Let's take a look at the micograph in Figure 1.5. A pentium 4 architure was introduced in the year 2000. This includes 54 million of transistors, run at 1.5GHz thanks to a $0.18\mu m$ technology. Current processors run at about 3GHz, include more that 100 million transistors and are fabricated using a $65nm$ technology.

In 1965 **Gordon Moore** (Fairchild Semiconductor, and later confounder of Intel Corporation) observed, based on the integration capabilities gained from 1958 to 1965, that the integration complexity were growing and would have grown with an exponential rate. Data in the last decade seem to confirm this trend. 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 [13] and non-memory chip as `Wafer Scale Engine` by Cerebras [14], a deep learning engine, has 1.2 trillion MOSFETs, manufactured using TSMC's 16 nm FinFET process.

Why? What did allow such a performance improvement?

One of the keyword is ***abstraction***. When first microprocessors were designed, all the circuit was

Figure 1.5: Pentium 4 micrograph.

done directly at the layout level from teams of designer which know all about their processor, from instruction set down to the smallest transistor layout. The design flow is now a hierarchical one. Single small blocks are carefully designed at full custom level, and afterwards characterised: a few important parameters are defined and only those are used for building bigger blocks. And so on recursively: from device level, to circuit, from circuit to gate level, from gate to macro-blocks and up to system. ***Divide et impera*** has been the enabling strategy. The consequence is that this have resulted in numerous improvement in the metal-oxide semiconductor structure. **Finfet** is the device structure in used today we also expecting the gate all around transistor very soon. In 2015 intel lunched the **15**

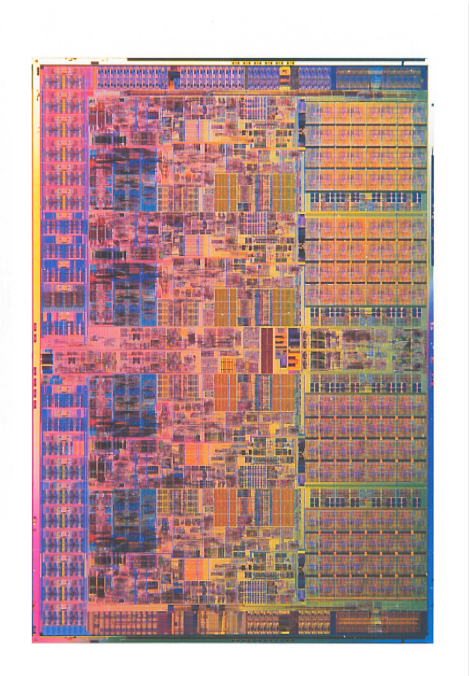


Figure 1.6: A figure illustrating the 15 core Ivybridge ex architecture.

Core Ivybridge Ex Architecture which is known to contain 4.3 billion transitors.

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 [15].

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]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 [16] 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.

1.1.1 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) [16]:

- 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 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 [16]. 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 V, chapter **??**.

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. In summary, some few keypoints that one must be able to satisfy in order for the new technology to be accepted. The first keypoint is that the new technology must solves the problem of power dessipation. The second keypoint is that the new technology must be affordable for high volume production. The third keypoint is that the new technology must allow for integrability. We have the saying that acceptance implies applicability.

1.1.2 What's next then?

In light of the above mentioned fundamental and intrinsic limit, it is legitimated the question: "what will be next?". Indeed it appears that sooner or later an alternative to conventional silicon devices will be necessary. Nobody can answer today to this question, and what the future technology will be depends on many factors, among which the cost will very likely play a central role. Nevertheless the last International Technology Roadmap for Semiconductor (ITRS) report in 2015 [17] and the successive IRDS ones [18], [19] and [20] (in 2017, 2018 and 2020 respectively) try to summarize the possible answers to that question in their sections named "Beyond CMOS". The driving interest of the research in these innovative beyond CMOS technologies is to maintain the circuit scaling cadence and to sustain the performance gain started with CMOS scaling process [17]. This is referred as "more Moore" in the ITRS and IRDS reports. Next to this attempt, also another reason is behind this front of research, that is the "more than Moore", i.e. the trend of accomplishing an heterogeneous integration of multiple functions (not only information processing and storing but even its transmission/reception or sensing and measuring, etc...) as a new paradigm for system architectures [17], [20]. Consequently the beyond CMOS research is not only carried on at device level, but all abstraction levels are simultaneously

involved. Novel computing paradigms for beyond CMOS technologies were investigated in the last years [18], [19], [20], and some of them (such as big data, internet of things, artificial intelligence, neural networks, etc...) require high performance and efficiency which is increasingly difficult to fulfill with the saturating "more Moore" CMOS technologies [20].
During the last decades, the explored and proposed beyond CMOS solutions are many and various, and the current landscape is extremely ample [20]. In figure 1.7 is summarized a possible taxonomy of emerging beyond CMOS memory devices (top) and there is also a comparison among few of them (bottom-left); while figure 1.8 summarizes a possible taxonomy of emerging beyond CMOS logic devices. As evident several solutions exist in both the fields, even if many of them are just "emerging" proposal, in the sense that they were proved at research level but advantages w.r.t. CMOS may be at moment only potential, i.e. at the present there may be important technological limits that are still unsolved. Nevertheless they are somehow promising or interesting enough to be still considered as a valid research topics for the next years [20]. An introduction to all of them falls outside the purposes of this work (interested readers can refer directly to the IRDS 2020 report [20] and references therein). However they constitute the current background in which "molecular electronic sensors", main subject of this work, find place.
ITRS and IRDS reports clearly state that the technologies they consider are somehow promising substitute of conventional CMOS at short-term, as bridging technologies between current CMOS and future beyond CMOS, but nothing is said about the long-term beyond CMOS technologies [20]. Indeed it is reasonable supposing that the transition between conventional electronics and the future one will be gradual, with, at the beginning, a coexistence between CMOS and beyond CMOS technologies, until (for some reasons) one technology will "win" above the others and will become the dominant one. Thus at the beginning, i.e. during the next years, an important or even discriminant factor for the success of new technologies will likely be the compatibility with standard CMOS technology and with standard CMOS architectures, since a coexistence even within the same IC, will be very probable [20], [21]. Nevertheless this is not the only feature to be considered in comparing the beyond CMOS technologies. Several factors should be accounted for, see e.g. fig. 1.7 (bottom-left), such as the already mentioned process cost, the performances in terms of speed and dissipated power (or energy efficiency), the scalability (that should be -at least potentially- superior to current CMOS one), etc... At least some of these features are required to a new technology to be further considered as a potential substitute of CMOS, and this holds true for all the technologies reported in figures 1.7 and 1.8.
In light of this, promising new technologies for the next future, i.e. at short-term, seem to be e.g. the Tunnel FET (TFET), the Nano-Wire FET (NW FET) and the Carbon NanoTube FET (CNT FET), etc... The TFET appears able to overcome the CMOS thermal intrinsic limit of 60 mV/dec for the subthreshold slope (SS). While there is no difficulty in switch it off (contrary to conventional CMOS) there are at the moment still difficulties in achieving an enough large ON current (due to the ON state barrier through which tunneling happens, that is too wide thus leading to a small tunneling probability - see also chapter 17). It presents also optimal compatibility with CMOS processes and architectures. For memories, the landscape is more extended and ample and here I would just like to cite the macromolecular memories, in which the basic cell is essentially a capacitor whose dielectric is a polymer (i.e. a molecular chain), eventually embedding small molecules or nanoparticles (e.g. the C_{60} fullerene that will be cited again in next chapters of this work). The logic state ("0" or "1") can be stable at room temperature and present a resistive, ferroelectric or capacitive nature depending on the specific nature of the sandwich [20]. The programming power (tens of nW) seems to be enough low for extremely low-power applications [20]. An evolution toward miniaturization of these memories are the macromolecular ReRAM or even the molecular memories. The latter are memory applications of the so-called molecular Single Electron Transistor (SET), that will be cited again in next chapters of this work.
As mentioned, the ample set of all these possible devices essentially constitutes the wide landscape in which nowadays molecular electronics and molecular electronic sensors find place. Before going on in clarifying better what molecular electronics is, where it is precisely placed in terms of its technology features and why it is intimately linked with molecular electronic sensors, a deeper look to the capabilities that are required to novel beyond CMOS technologies is given.

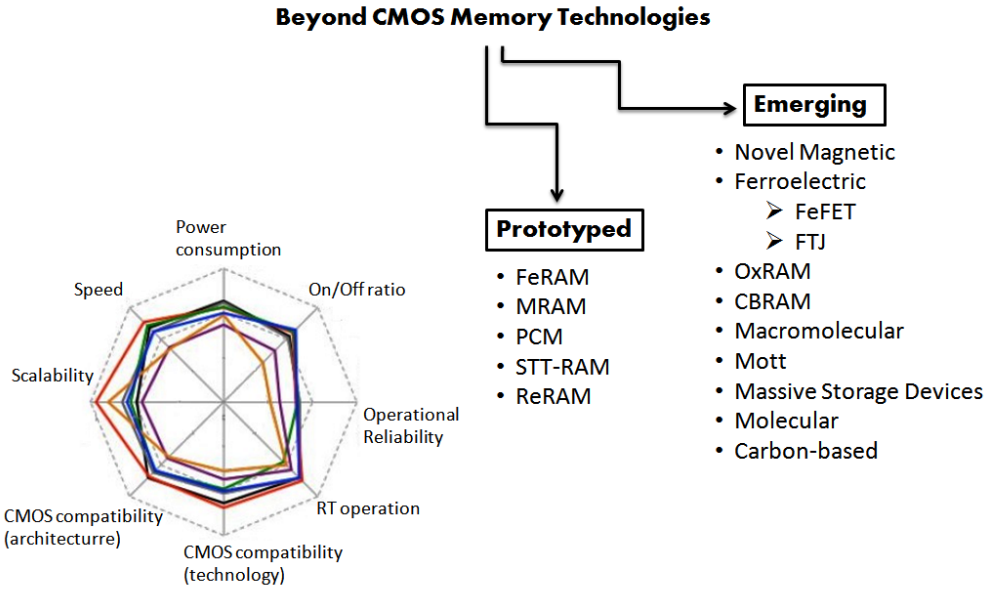


Figure 1.7: Possible taxonomy of emerging memory devices and a comparison among few of them (bottom-left) in terms of relevant features for a future commercialization. Black line: FeFET memory; grey line: FTJ memory; red line: ReRAM memory; green line: Mott memory; purple line: macromolecular memory; orange line: molecular memory; blue line: carbon-based memory. The classification is coherent with the one reported in [20].

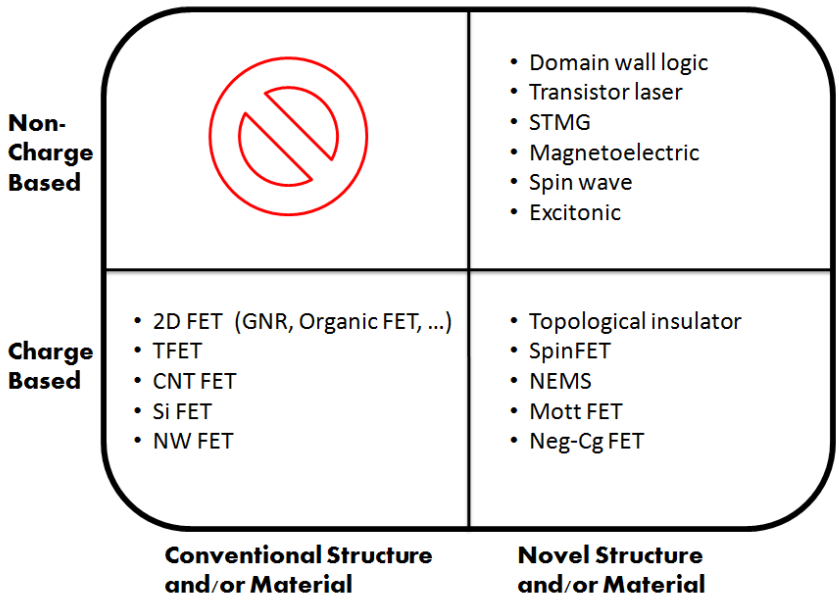


Figure 1.8: Possible options for emerging logic devices. The devices are differentiated according to whether the structure or the material are conventional or innovative (abscissa axis) and whether the information representation is based on electron charge or on some non-charge entity (ordinate axis). Only three groups are possible since a conventional FET material and structure is intrinsically a charge-based device. The classification is in accordance with the one from [20].

For many years the technology advancements were driven by memories (DRAMs especially), in which the demand for increased capability drove to a huge research effort in scalability of the basic cells and thus transistors. Instead during the last years the driving force gradually became the processing unit,

requiring for extremely low power functioning and increased cache memories capability and performances [20]. In these terms, the novel logic switches, in order to be considered potential replacements for current CMOS ones, should possess one or more of the following fundamental capabilities [20]:

- scalability: increase in the device density (and thus decrease the cost) beyond that achievable by ultimately scaled CMOS
- speed: increase beyond CMOS in speed performances (switching speed)
- power: reduction of the switching energy and thus of the overall power consumption
- novel computing schemes: enabling of innovative computing paradigms or information processing functions that cannot be performed as efficiently with conventional CMOS

Molecular electronics gained particular attention during last three decades from scientific community because of its intrinsic interesting capabilities [22], [23]. In December 2005, the ITRS officially noted that molecular electronic components will need to be integrated into silicon manufacturing in order for Moore's law to continue to hold true [24], [22].
With the term "molecular electronics", sometimes also called "moletronics" [22], [25], it is conventionally intended that field of electronics in which the fundamental units for computing are devices based on single molecules or, alternatively, small packets of molecules [26]. It is an interdisciplinary field that includes physics, chemistry, materials science and engineering [25]. In particular all the electronic components can be realized by means of single (or small packets of parallel) molecules, thus transistors, diodes, resistors, capacitors, etc... can be all implemented as molecular devices [25].
As a title of example in figure 1.9 is reported a scheme of a molecular transistor. Notice that source, drain and gate contacts are quite conventional, a gate dielectric is still present (in general a front and a back gates can be present) and the main distinctive feature is that the channel is constituted by a single molecule, chemically bonded with the source and drain contacts and "suspended" above the gate dielectric. Thus the active portion of the device is a single molecule.

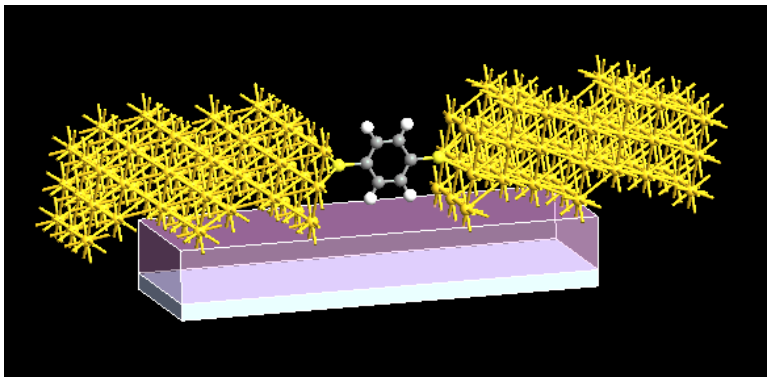


Figure 1.9: Example of a molecular FET structure. In this case the PDT (phenyl dithiol) molecule is used. The yellow atoms are the gold source and drain contacts; the two dark yellow atoms are the sulfur anchoring groups; grey atoms are carbon and white hydrogen. The light purple region is the gate insulator (continuum approximation) while the white/light grey region is the metal gate contact. The figure was generated in *Quantum-Wise ATK* software.

Sometimes with the term "molecular electronics" is indicated a much wider class of devices, including some macromolecular-based (or polymer-based) devices and transistors, and also optoelectronic devices such as organic light-emitting displays (or even liquid crystal displays), organic photovoltaic devices, chemical sensors and so on... that are built starting from (organic) molecular processes, but present a much larger active region than the one of small molecule devices, like the transistor mentioned so far [25], [27]. These devices were the first practical results of research on the field of molecular electronics mainly pioneered by Forrest Garret [22] during the 1980s.

The use of single molecules as basic building blocks for realizing electronic devices potentially represents the next (or even the ultimate) limit of miniaturization of electronic devices, beyond the CMOS intrinsic limit, and up to the basic "bricks" of matter [28], [29]. Indeed molecules are the fundamental entities of matter, or better of compound materials (atoms should be more correctly the fundamental basic entities of matter but they are so only for single elements). And this is more correct the more the employed molecules are small, e.g. the benzene ring with a dimension of the order of 1 nm (while it may be not true at all, as already mentioned, for the so-called macromolecules that can be hundreds or thousands of times bigger than the benzene ring - think e.g. to proteins, enzymes, DNA, or bio-macro-molecules in general).
Since the beginning of the century several molecular device prototypes were demonstrated in literature [23], [29], even if, actually, the idea of making electronic devices by means of single molecules is not so young. The term "molecular electronics" was coined in 1958 (even if the meaning was slightly different: it was referred to a bottom-up material engineering from molecules to devices), when the US Air Force and Westinghouse Electric Company adopted it to describe their radical joint research program, that unfortunately was a failure [22]. In the mid 1970s the term "molecular electronics" went back to vogue (with the conventional meaning explained above) and in 1974 the first molecular rectifier (actually the first single molecule electronic device) was proposed [30]. This was a purely theoretical device, in which an isolated molecule was supposed to present rectifying properties. The first prototypes were instead successfully realized only after more than twenty years, because of technological limits [22]. The current technology allows for the fabrication of single molecule devices or even for the simultaneous creation of thousands of them, with a reasonable reliability and yield, even if currently the technological process variability is still the main issue in molecular electronics [26], [28], [23], [27], [29]. For example, the manipulation of single molecules is possible by means of Scanning Tunneling Microscopy (STM) methods, that make possible even to act on a single molecule, e.g. by breaking or modifying it to perform the synthesis of specific structures [23], [27], [31]. Moreover the synthesis and anchoring (even with the desired orientation) of millions or billions of molecules is made possible by means of Self-Assembled Monolayer (SAM) techniques, in which a sequence of chemical reactions, under given conditions, is able to self-assembly molecules in between break-junctions or above a substrate [26], [29]. From the standpoint of contacting the molecule the difficulty is mainly due to the creation of reliable and clean (typically metal or graphene) contact gaps, of the order of the nm. For example this is achievable by means of Mechanically Controlled Break-Junctions (MCBJ) or through controlled electromigration, electroburning or etching processes [23], [29], [32], [33], [34].
Hence at moment, after more than twenty years of ameliorations in the techniques for the construction of molecular devices, it is possible to fabricate single molecule devices with a yield of the order of $25\% \div 40\%$ (better performances are often obtained with graphene contacts) [32], [33], [34]. As mentioned, the low yield and the high process variations are at the moment the main unsolved issues in molecular electronics from the technological standpoint, leading to low reliability of fabricated devices. The high process variations, moreover, lead to a wide variance on measured current-voltage characteristics. Other problems concern the large parasitics, especially for the interconnection, that are actually the main Achilles' heel of this technology (precisely like in the ultra-scaled CMOS case). Another very remarkable drawback is the usually poor I_{ON}/I_{OFF} ratio (for conventional molecules it is of the order of few tens [23] even if improvements up to few hundreds seem possible [35], [36]) and the too low I_{ON} value. Instead, the main potential advantage w.r.t. to CMOS is the huge scalability: molecules as fundamental entities of matter push the scalability to the ultimate intrinsic limit of miniaturization, potentially beyond the 8 nm thermal limit of silicon devices [22], [26]. The SAM processes are compatible with the conventional CMOS process. Even if they require few extra steps to be performed w.r.t. to standard CMOS process, these extra steps are generally few and can be performed such that the remaining steps of standard processes are not affected at all [29]. Moreover the parallel and simultaneous fabrications of billions of molecular devices, accessible with the SAM processes, allows to potentially further decrease the cost per unit below the CMOS one. Due to the small dimensions and currents, molecular devices have other potential advantages w.r.t. CMOS: they seem to be intrinsically low power, even if compared with conventional CMOS [25], [35], [23]; and moreover due to the small amount of electrons (thus mass) that are moved through a so tiny device,

molecular electronic components are potentially faster that CMOS ones (potentially smaller intrinsic relaxation times), even if parasitics constitute currently a big conflicting effect to this benefit. In the words of Mark A. Reed and James Tour: "molecular electronics is the ultimate step in small, cheap and fast" [22]; nevertheless, mainly due to the previously cited technological still unsolved problems, these are currently only potential advantages. Table 1.1 summarizes the molecular electronics pros and cons illustrated up to now.

Table 1.1: Advantages and drawbacks of moletronics (i.e. molecular electronic technology) w.r.t. conventional CMOS technology.

Advantages of moletronics w.r.t. CMOS	Drawbacks of moletronics w.r.t. CMOS
huge scalability	process variations
low cost (thanks to SAM)	low yield and reliability
CMOS process compatible	large parasitics
low power	poor Ion/Ioff ratio

In summary, in terms of the previously discussed novel technology capabilities, molecular electronics presents:

- scalability: far beyond the conventional CMOS
- speed: potentially faster than CMOS, still very limited
- power: potentially lesser than CMOS, still limited
- novel computing paradigms: molecular electronics open the way to several interesting new processing principles, e.g. in neuromorphic system [22], [37]. The main strong point in this direction seems to be the possibility of synthesizing *ad hoc* organic molecules, able to perform innovative basic operations, intended as the basis for the new computing paradigm [22], [37].

Currently molecular electronics seems to catch the attention as potential beyond CMOS technology at long-term, in the words of Mark A. Reed (2009) [22]:

"Even though in many ways I was responsible for the generation of this new incarnation of molecular electronics, I'm a critic. Because I don't see it going into electronic devices in the future. Yet. If I'm going to take my money today, I'm going to put it on silicon. But that doesn't mean that I don't think that there's potential. [...] My job description (unlike someone at Intel) is to explore the physics, explore the potential over a lifetime. So my time horizon is 20, 30, 40 years. [...] On top of that, there's an additional part of my job description, which is [...] is there any important physics to be learned from it?"

Indeed even if the initial goal of molecular electronics was to effectively substitutes the conventional CMOS devices, due to the above discussed drawbacks, it resulted clear that, at least at short-term, it would have been not possible to successfully accomplish this task [26]. This is the main reason why molecular electronics does not assume a central role in the last ITRS and IRDS reports (that are focused on short-term promising technologies) [17], [20]. Nevertheless nobody can today know what will happen at long-term, and if the molecular electronics technological limits and disadvantages will be overcome, leading to a much better promising technology w.r.t. CMOS one, or not.
Moreover, the research in this field led to important irrefutable achievements: it opened the way to a set of technologies aimed in the development of novel molecular devices that find place in several potential field of applications [26]. Molecular electronic sensors are examples of this. The developed technological processes and techniques for molecular electronics are exploitable for building a new class of sensing devices that became subject of research in the last few years: the "molecular electronic

sensors" [33], [34], [38], [39], [40], [41], [42]. They are also the subject of this entire work. A molecular electronic sensor is essentially a molecular wire (i.e. a molecule or a packet of molecules in between two contacts) or a molecular transistor, in which the conductance of the molecular channel, the device capacitance, etc.. are changed depending on the presence or not of a target substance or chemical species (e.g. gas sensor). In general they can also sense the temperature, the light (even in the far InfraRed -IR- gap) and other chemical-physical quantities [33], [38], [40], [42], [43].
Molecular electronic sensors will be presented more in detail in the next section, and then they will be widely considered in the rest of this work. Moreover they are a possible short-term answer to the "more than Moore" trend, allowing for hybrid integration of molecular sensors along with conventional CMOS electronics within the same IC, to the end of performing the sensing and the post-processing of several chemical-physical quantities in a unique low-cost ultra-scaled system [21].

In conclusion, the constant and incessant scaling pace during the last sixty years promoted the research not only in the silicon industry field but even in the beyond CMOS field. In a very rich, prosperous and ample landscape, molecular electronics represents only a possibility among the others for the long-term future of ultra-scaled electronics. At the same time the theoretical and experimental research in this field opened the way to the fabrication of sensors, that can be accessed at short-term end, for the advancements in the "more than Moore" trend. They are also a first potential market arrival point for molecular electronics, able to potentially become a driving force for the research in this fascinating field.

However, another important consideration must be made. Not only continual miniaturization should be enabled, but a new form of computational architecture must be accomplished, where logic and memory are combined in a single device. Magnetic materials can provide both features: a premanent magnetization state serving as memory and a stray eld, which can be exploited for interaction between magnetic, ultra small-scaled unities to perform logic operation. In the last 60 years, magnetic devices already approved in terms of data storage applications and gain an outstanding position in the microelec- tronic world. In terms of logic operation, research intensied investigations on magnetic materials, in the last decade. Spintronic is a branch of electronics that deals with spin-polarized currents. The discovery of spintronics was the breakthrough the lead to modern high-density hard drives and the newly developed magnetic RAMs. There are two main devices at the base of spintronics, GMRs (Giant Magneto-Resistances) and MTJs (Magneto-Tunnel Junctions). Both can be seen as variable resistances where the resistance value depends on the magnetic state of the device. While GMR and MTJ are commercially available devices, a big number of other devices has been investigated at research level. These new devices are interesting because they offer features not available in MOS-based circuits, like the ability to combine memory and logic on a single device, the intrinsic ability to fabricate 3D circuits and no standby power consumption. A recently developed branch of the study of on magnetic devices has led to the development of the Racetrack Logic, Logic-in-Memory devices based upon racetrack memories. A racetrack logic element can be seen as a memory element that is also able to perform logic operations on the stored information.

Together with molecular electronics devices, magnetic logic-combined-with-memory elements are interesting to be explored. Whether they will be the actual solution for the future it is not known. However, from the methodology point of view and for the enoumous amount of knowledge derived from the study of these new devices and their use in a possible future system for computation or sensing or computing-in-memory, it is worth diving with enthusiasm and curiosity in the next chapters.

Part I

Conduction: from microscale to nanoscale

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 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.

A backgound on quantum mechanics and further description and discussion on this subjects are given in Part VI

CREDITS: The chapter in this part is derived by the MSc thesis of Ing. Chiara Elfi Spano, elaborated on the basis on the lectures of Prof. Gianluca Piccinini and related literature.

IMPORTANT DISCLAIMER NOTE: as mentioned in the main introduction, the reference for the exam preparation are the SLIDES given during the lectures and the registration. The material given her is to be considered an integration, and not necessarily it is complete w.r.t. the SLIDES part.

CHAPTER 2

Toward the nanoscale

In this chapter a simplified treatment of quantum transport modeling, mainly inspired by [44], [45] 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.

  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.

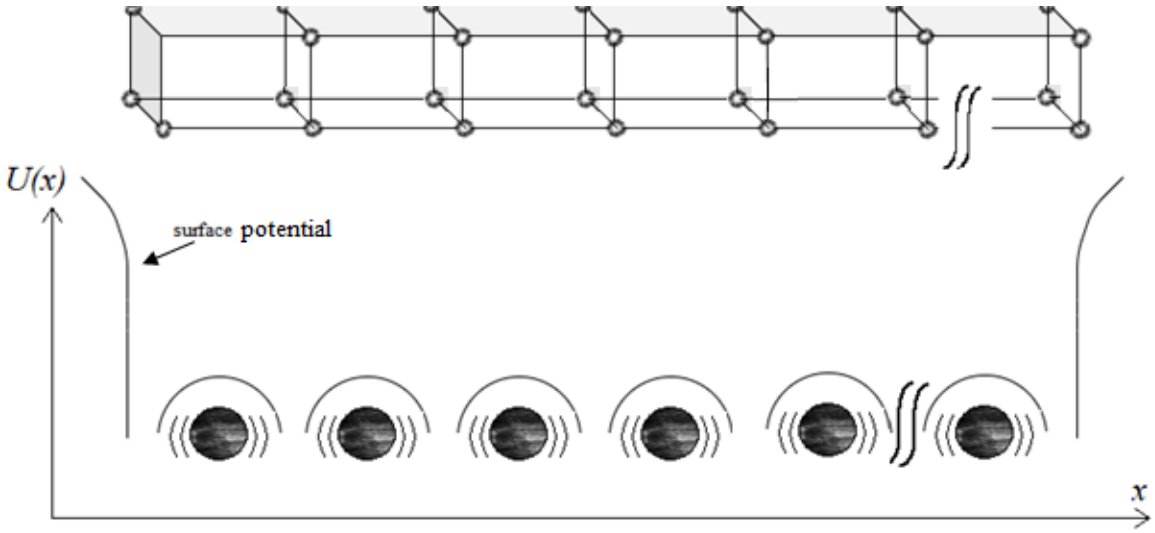


Figure 2.1: The 1D potential energy of the crystal lattice, periodic function of the space. It results from the periodic placement of the ions lattice within the crystal. The distance between one ion and the next is the lattice constant a.

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.

  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).

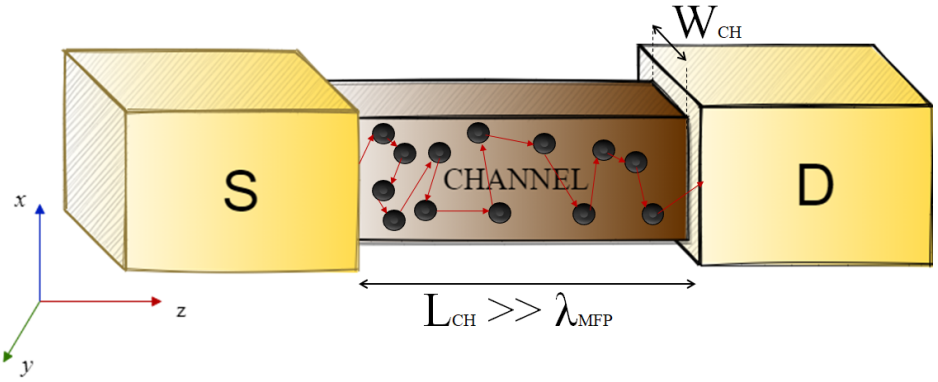


Figure 2.2: A sketch of diffusive charge transport in bulk systems, dominated by scattering events. The size of conductor L_{CH} is much greater than the electron free mean path.

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).

- **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

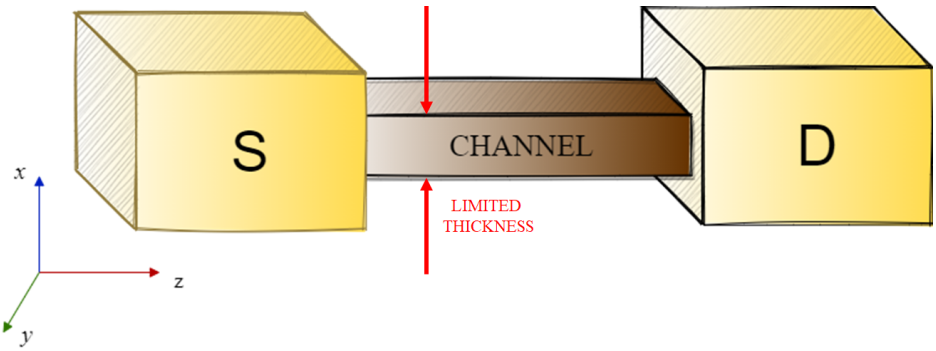


Figure 2.3: An example of 2D-system where the thickness is limited to few nanometers. The direction of confinement is x, a transversal one, orthogonal to the direction of transport (z).

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).

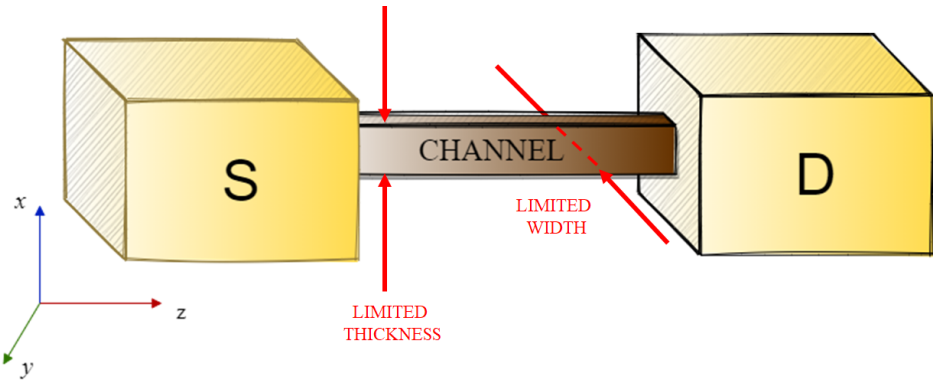


Figure 2.4: An example of 1D-system where the thickness and the width of the channel are limited to few nanometers. The direction of confinement are the transversal directions x and y, orthogonal to the direction of transport (z).

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, nano-crystals, 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 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

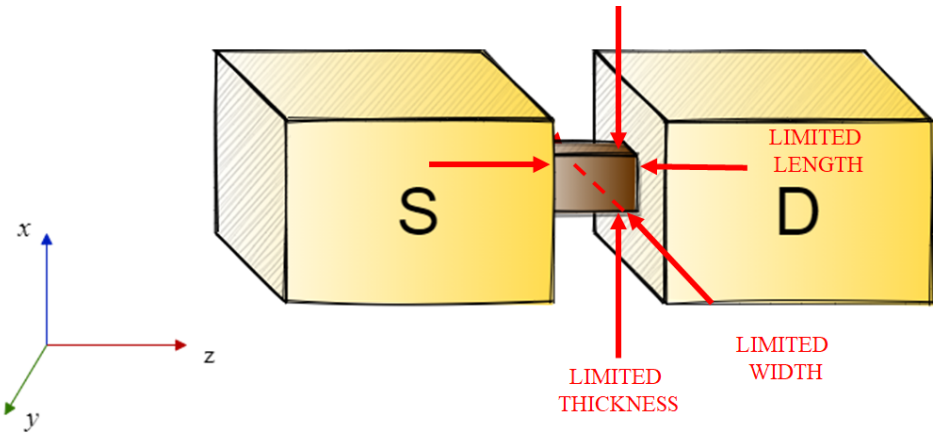


Figure 2.5: An example of 0D-system where the thickness, the width and the length of the channel are limited to few nanometers. The direction of confinement are x,y,z.

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.

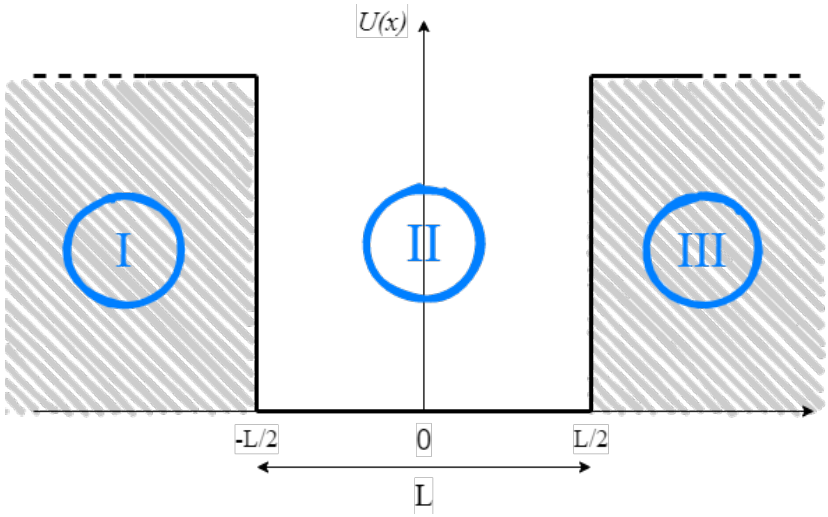


Figure 2.6: Potential energy distribution of 1-dimensional finite square well of length L. The finite heigth of the well is U_0.

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

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

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 $det(A) = 0$:

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

it yields: $\alpha = k tan(kL/2)$. 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:

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

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:

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

∗ 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 $det(A) = 0$:

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

it yields: $\alpha = -kcotan(kL/2)$. 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.

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:

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

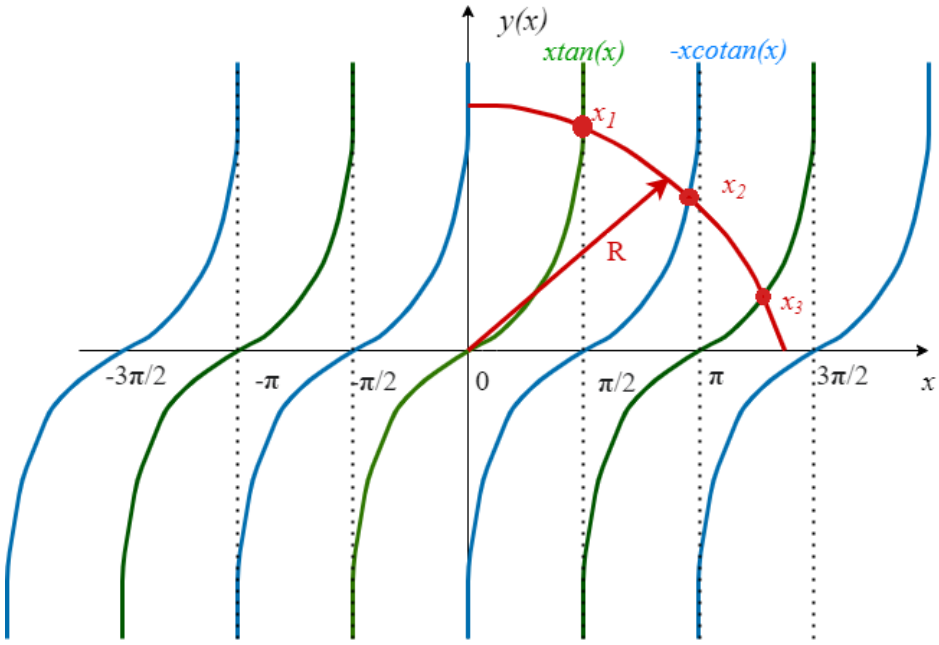


Figure 2.7: Graphical representation of the non trivial solutions of the considered 1D finite square well. The red curve is the circle of radius $R = \sqrt{\frac{2m^*}{\hbar^2}U_0}(L/2)$. Notice that in the figure there are only three allowed non trivial solutions (x_1, x_2, x_3)

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

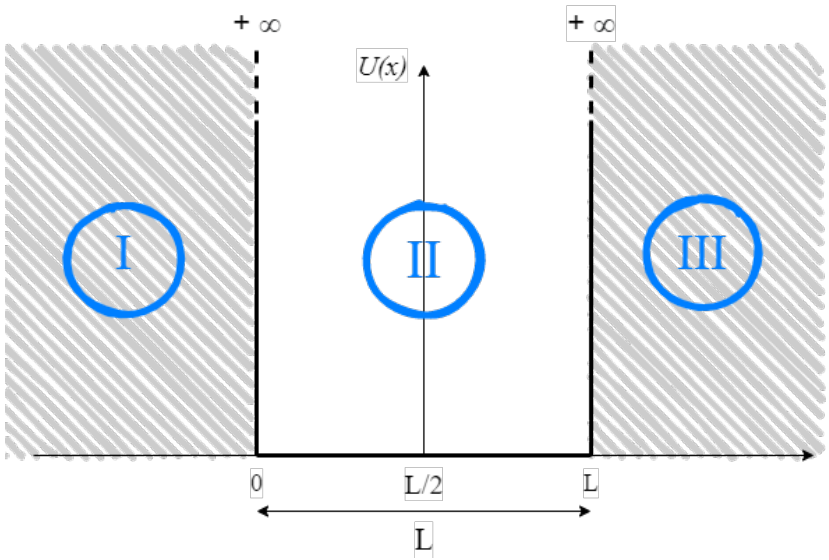


Figure 2.8: Potential energy distribution of 1-dimensional infinite square well of length L.

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

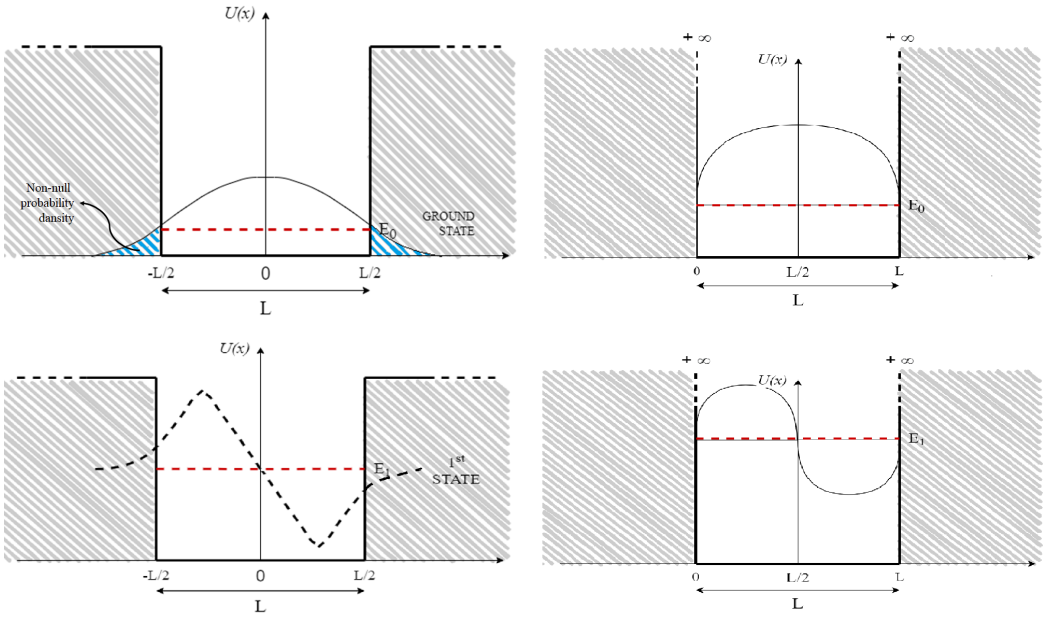


Figure 2.9: Differences in the first two energy levels between a 1D ideal finite square well and an infinite one.

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:

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

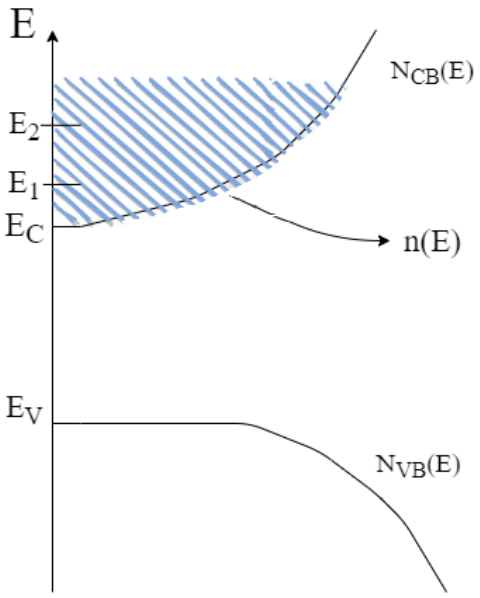


Figure 2.10: 3D-DOS of the Conduction Band (CB)

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

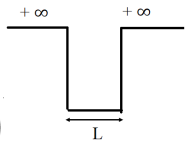
a bounding box inside which carriers are confined due to two high potential barriers. In the other directions they are free to move and their wavefunctions are modeled as *plane waves*.

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.

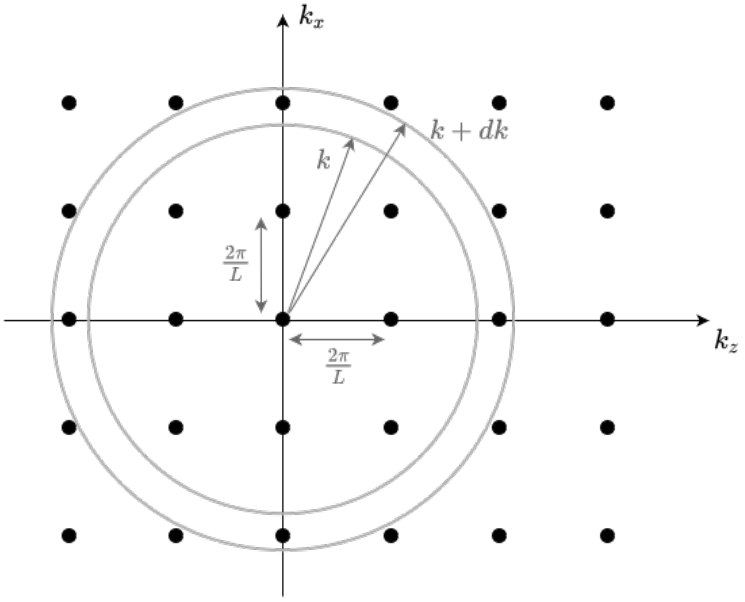


Figure 2.11: The (k_y, k_z) plane

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.

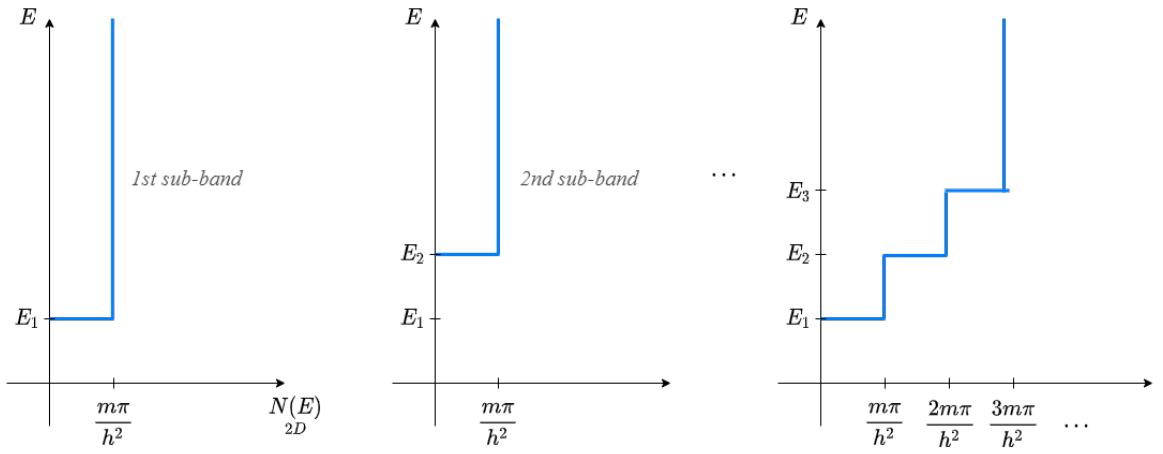


Figure 2.12: DOS of 2D system. On the left side the ones for single sub-bands. On the right side 2D-DOS for the entire system, resulted from the sum of DOS of all the subb-bands.

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:

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

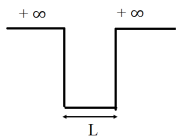
with $\mu(x)$ the step function centered in E_{n_x}

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

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



directions of confinement, where the potential 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.

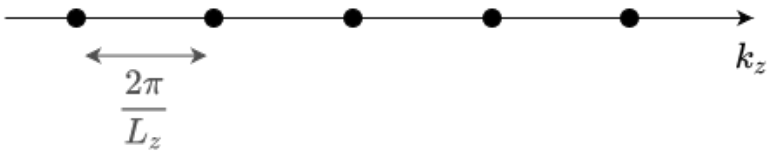


Figure 2.13: The k_z straight line.

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:

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

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 $E_{nx,ny}$

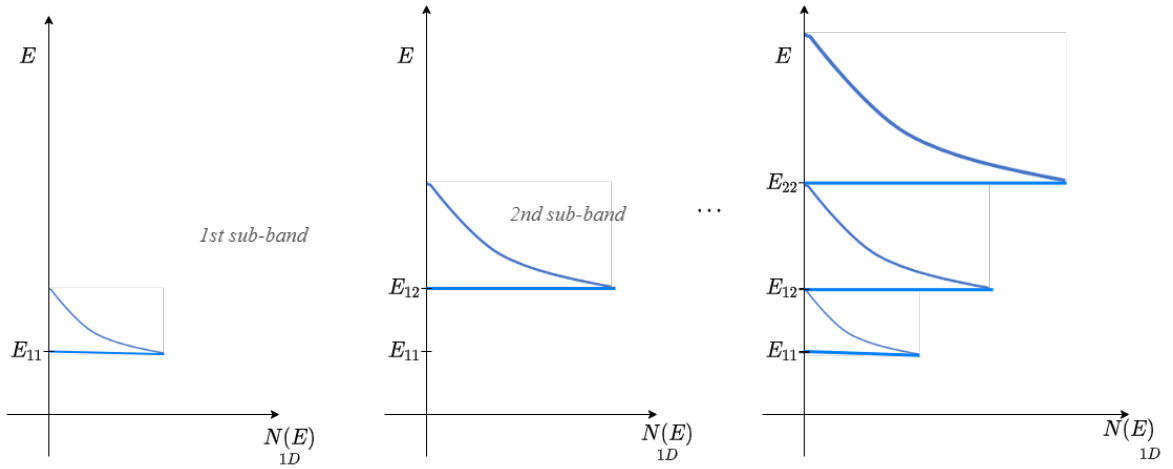
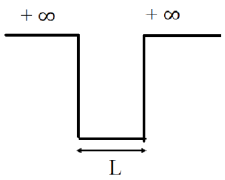


Figure 2.14: DOS of 1D system. On the left side the ones for single sub-bands. On the right side 1D-DOS for the entire system, resulted from the sum of DOS of all the subb-bands.

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

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



the 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} . Therefore, 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

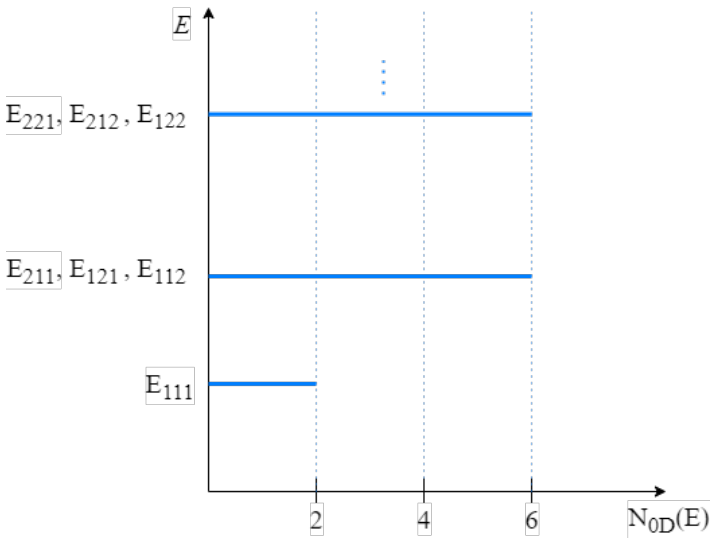


Figure 2.15: An example of 0D-DOS. Notice that is it possible to have more than 2 states per unit energy, since for three different combination of n_x, n_y, n_z the same energy values are obtained, simply because $L = L_x = L_y = L_z$ by hypothesis.

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.

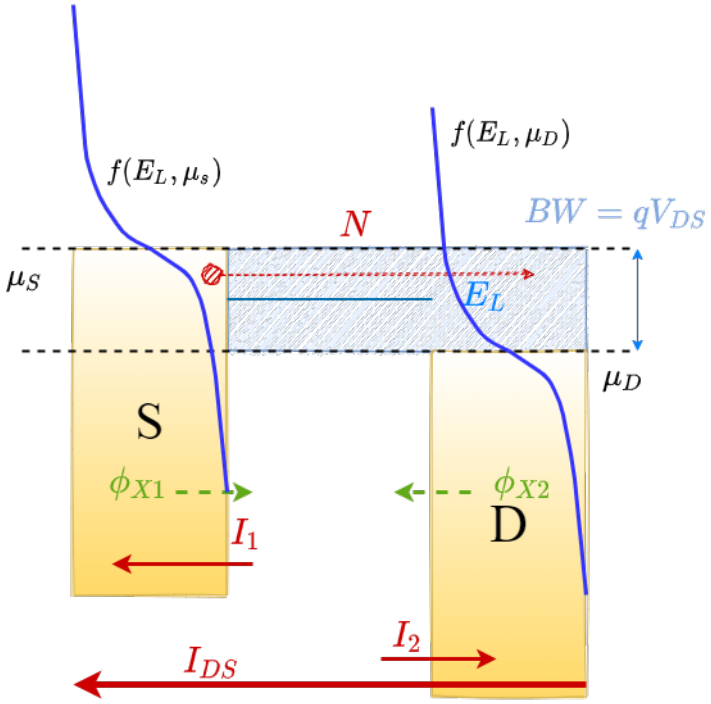


Figure 2.16: Case of a single energy level of the quantum dot E_L falling between μ_S and μ_D.

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: $\gamma_1 = \gamma_2 = \gamma$:

- 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, $I_{DS} \neq 0$.

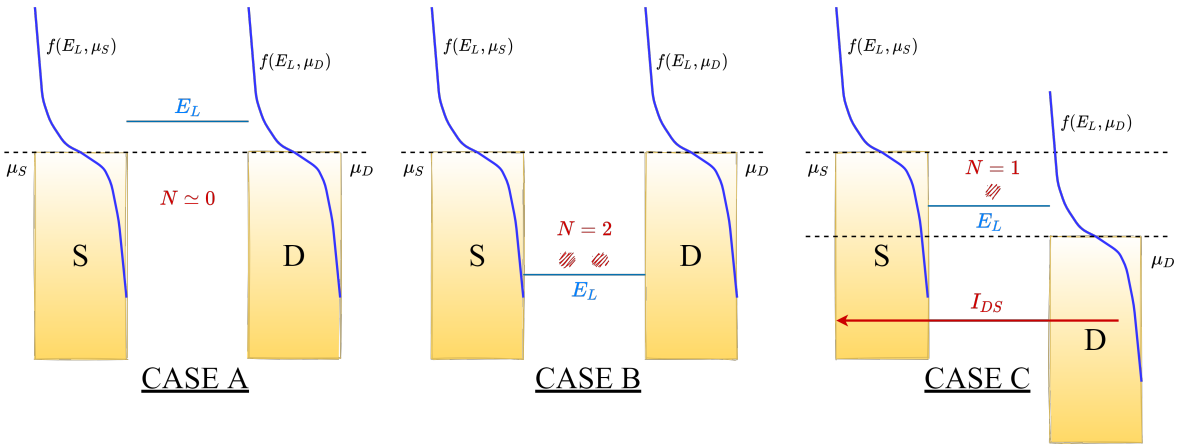


Figure 2.17: The previous cases described.

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:

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

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.

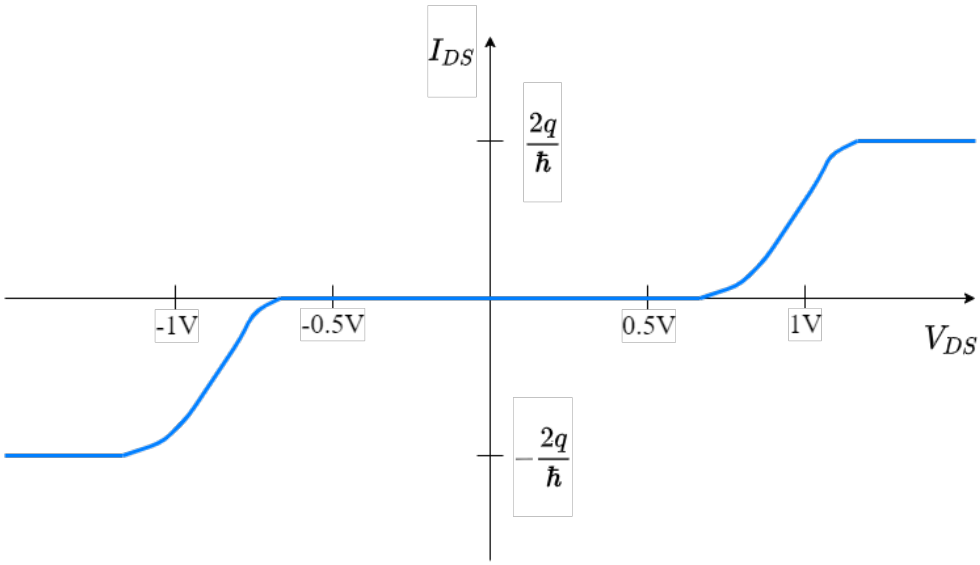


Figure 2.18: The resulting IV curve for LUMO conduction.

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: $f_{FD,S} = f_{FD,D} \simeq 1 \longrightarrow I_{DS} \simeq 0$;

- 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.

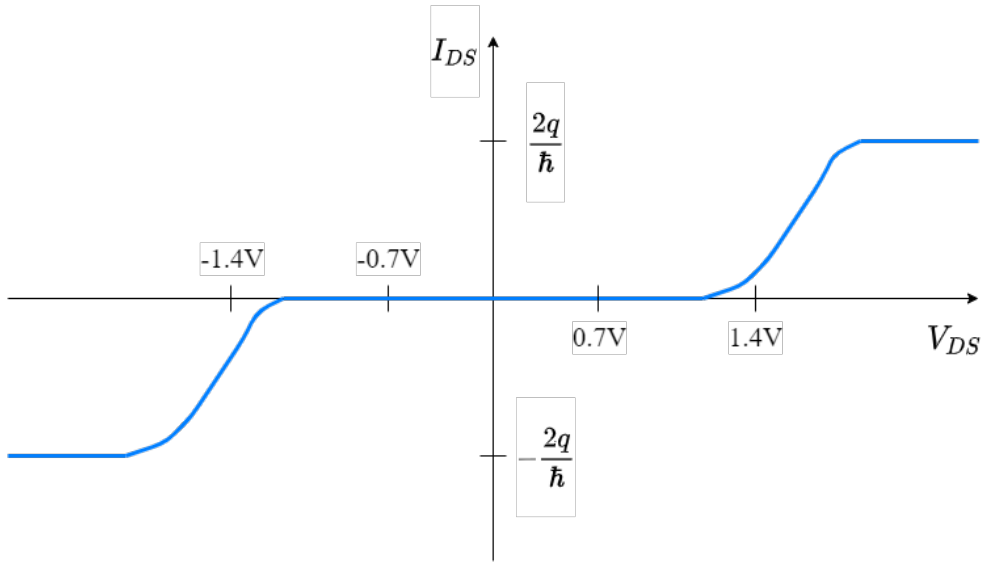


Figure 2.19: The resulting IV curve for HOMO conduction.

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.

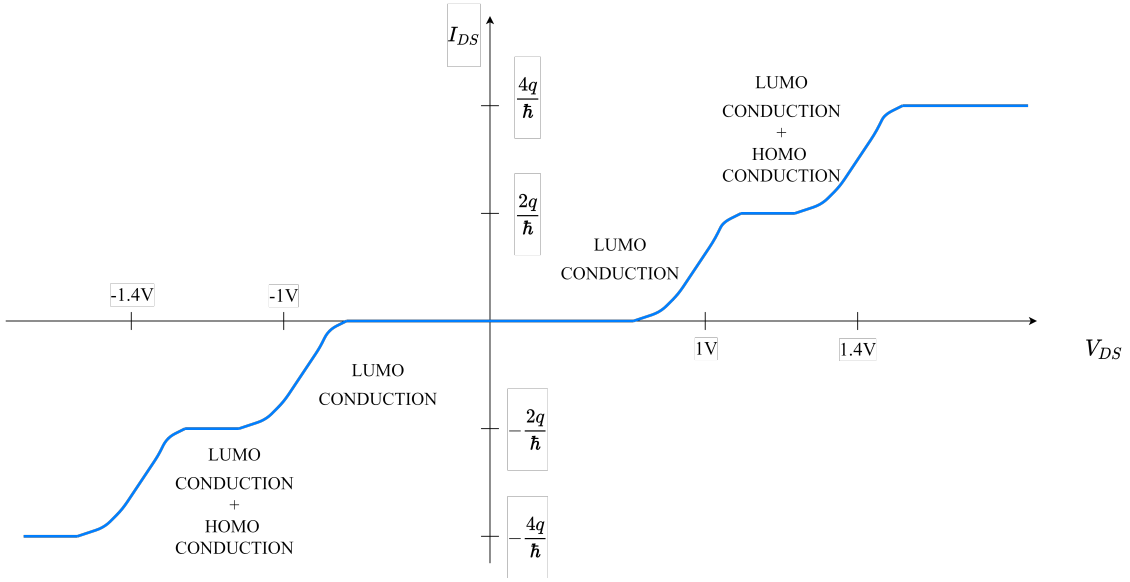


Figure 2.20: The resulting IV curve for LUMO and HOMO conduction.

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 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).

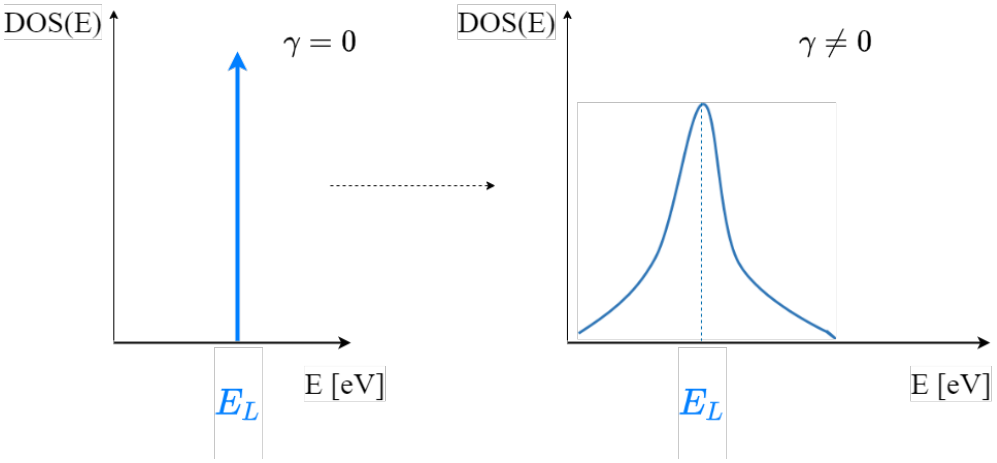


Figure 2.21: The phenomenon of level broadening.

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.

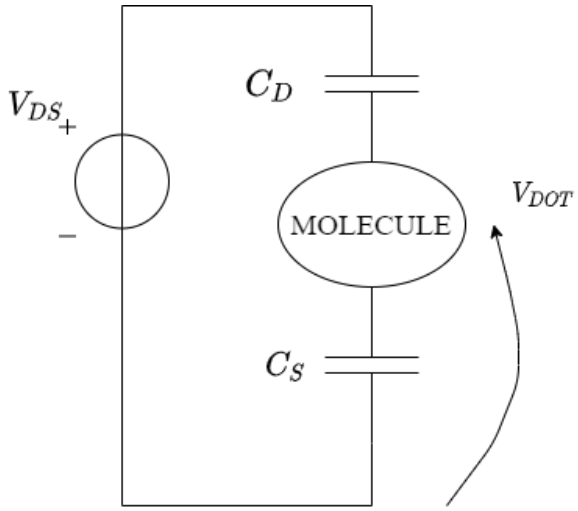


Figure 2.22: Equivalent circuital model of two-probe 0D-system.

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)

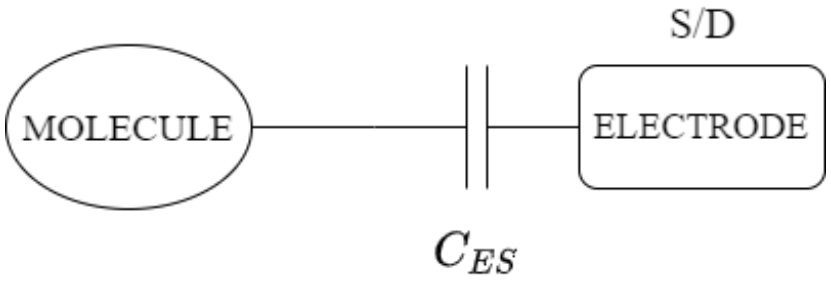


Figure 2.23: Equivalent circuital model of two-probe 0D-system.

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:

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

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:

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:

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

Meaning that if:

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

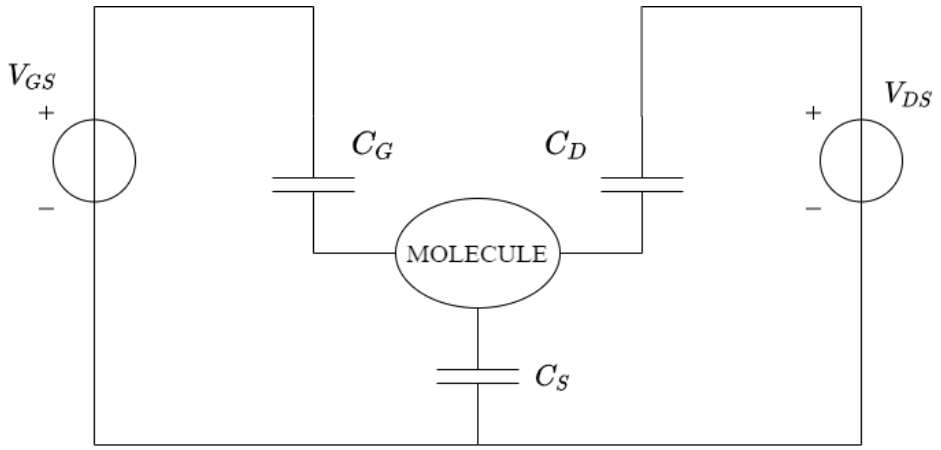


Figure 2.24: Equivalent circuital representation of MolFET.

- $V_{GS} < 0$, 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 [45]. 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 [44], that is again subtracted to Hamiltonian operator in eq. (2.73).

Part II

Molecular devices

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 the Lecture Notes is to investigate some of these features. Thus some two and three terminals molecular devices (molecular wire, molecular transistor, molecular sensors) 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.

CREDITS: The chapters in this part are derived mainly by the MSc thesis of Ing. Chiara Elfi Spano [1], the PhD thesis of Dr. Ali Zahir and the Msc thesis of Fabrizio Mo [2], and also from the Notes of some of the previous and current students of the course (Chiara Cannavò, Salvatore Vistato, Mario Simoni, Francesco Di Lorenzo). These chapters are elaborated on the basis of the lectures of prof. Mariagrazia Graziano and related literature.

IMPORTANT DISCLAIMER NOTE: as mentioned in the main introduction, the reference for the exam preparation are the SLIDES given during the lectures and the registration. The material given her is to be considered an integration, and not necessarily it is complete w.r.t. the SLIDES part.

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.

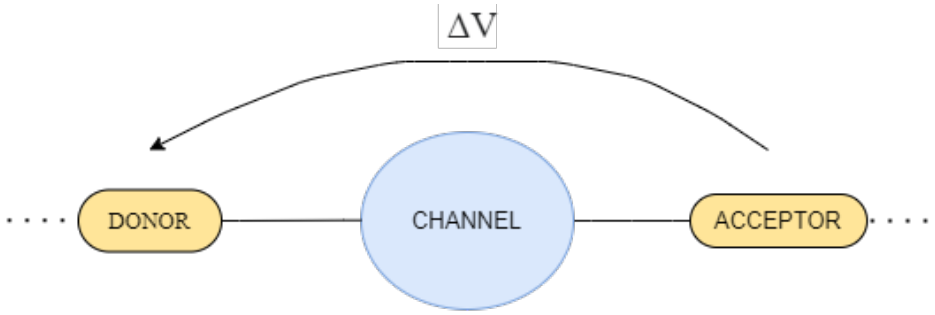


Figure 3.1: A depiction of a two probe system.

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 [46], or by the more common break junction techniques [47, 48].
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 [48], 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 [48]. 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) [49] to complex molecules (e.g., supramolecule, polymer)[50, 51, 52].
Memory, diode, switch, sensors for biomolecules, such as DNA [53], 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. [50, 51, 52]) 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).

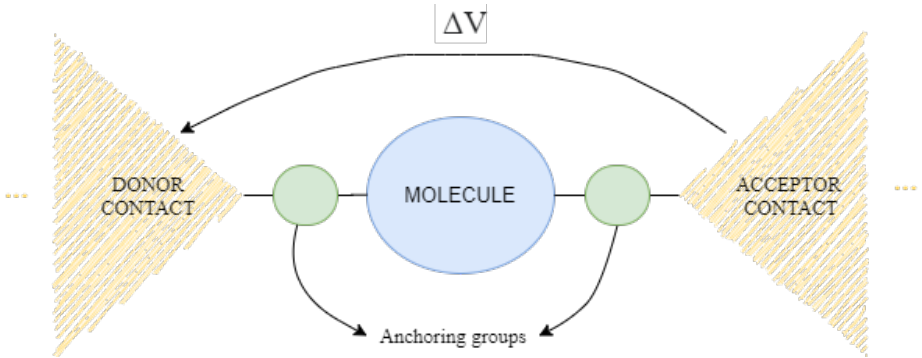


Figure 3.2: A depiction of a two probe molecular wire.

Despite its name, the molecular wire (also referred as *molecular juction* [54]) 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 [55, 56, 57].

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.

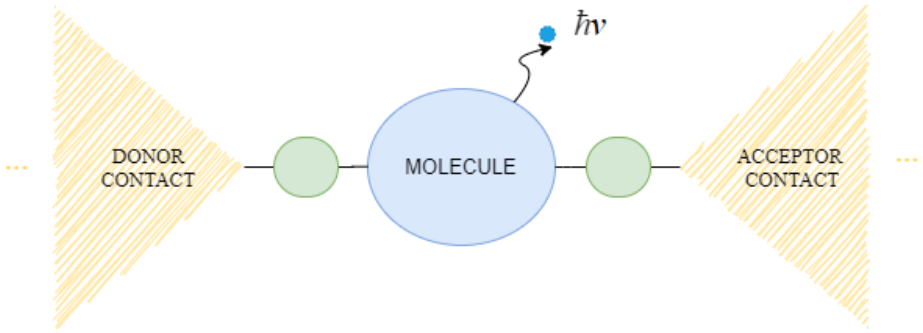


Figure 3.3: A depiction of a molecular system emetting photons.

  - 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 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.
  - 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

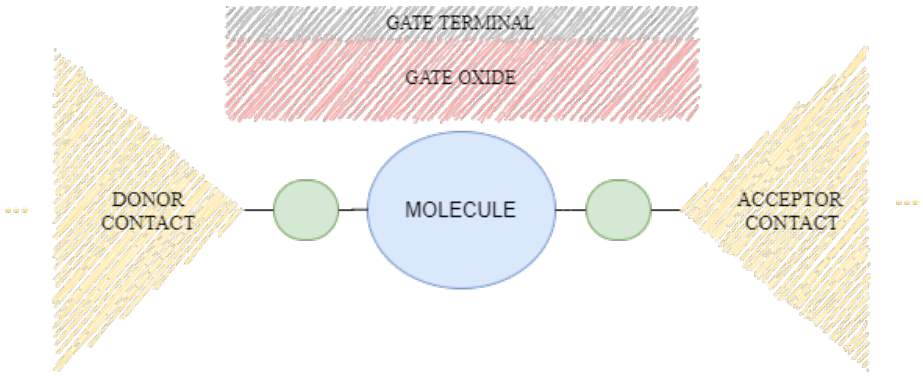


Figure 3.4: A depiction of a molecular FET with a single gate terminal.

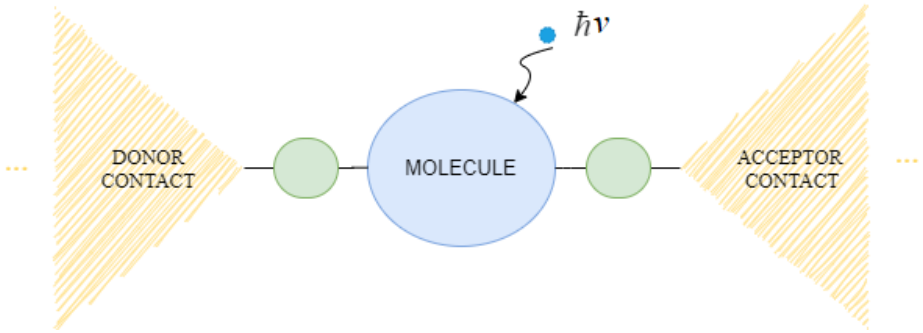


Figure 3.5: A depiction of a molecular photoswitch.

of current, making the system behaves like a receptor at nanoscale, thus giving rise to sensor/biosensors applications.

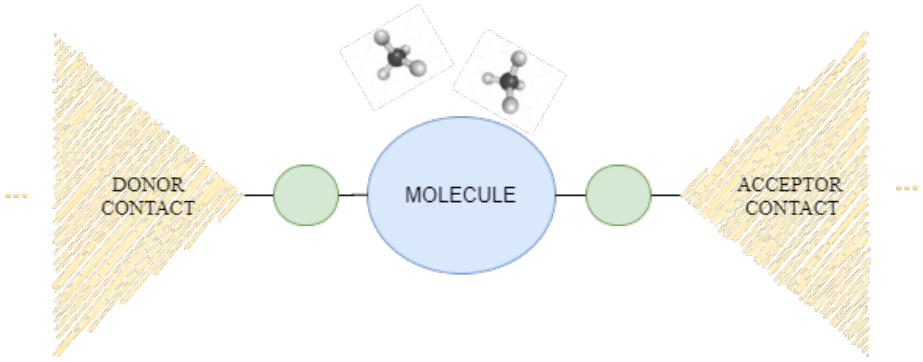


Figure 3.6: A depiction of a molecular biosensor of methane gas.

- 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* [58] 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 $0.05\,\mathrm{eV/\AA}$. 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 $1.71\,\mathrm{\AA}$ which corresponds to S-Au bond length of $2.38\,\mathrm{\AA}$, value validated experimentally in [59].

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

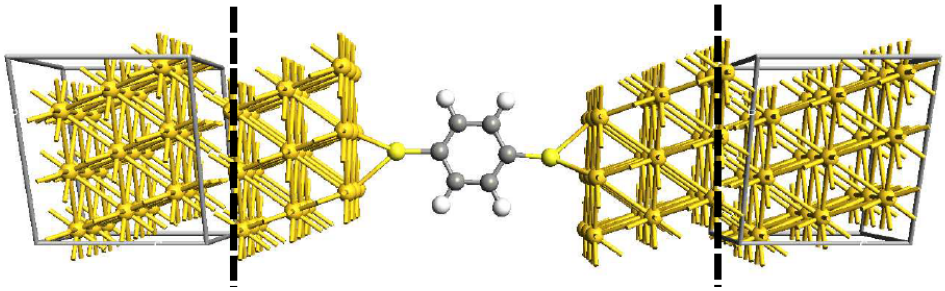


Figure 3.7: A typical structure of molecular junction for one ring of thiopene molecule (1TT). The central part is the active part of the device also called *scattering region*. The portion of electrodes between the two black dashed lines are called *leads*. They influences the broadening of the energy 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.

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)

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

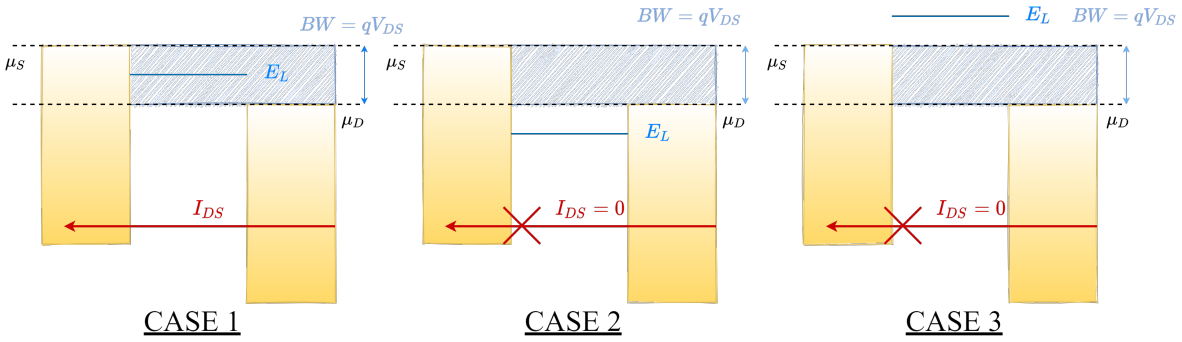


Figure 3.8: Case 1,2,3 desvribed above. The voltage division factor $\eta = 1/2$ shifts the Fermi level of the source $qV_{DS}/2$ above and the Fermi level of the drain below. In case 1: the enegy level of the dotE_L is within the BW therefore $I_{DS} > 0$. In case 2 and 3 instead is not included in 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.

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:

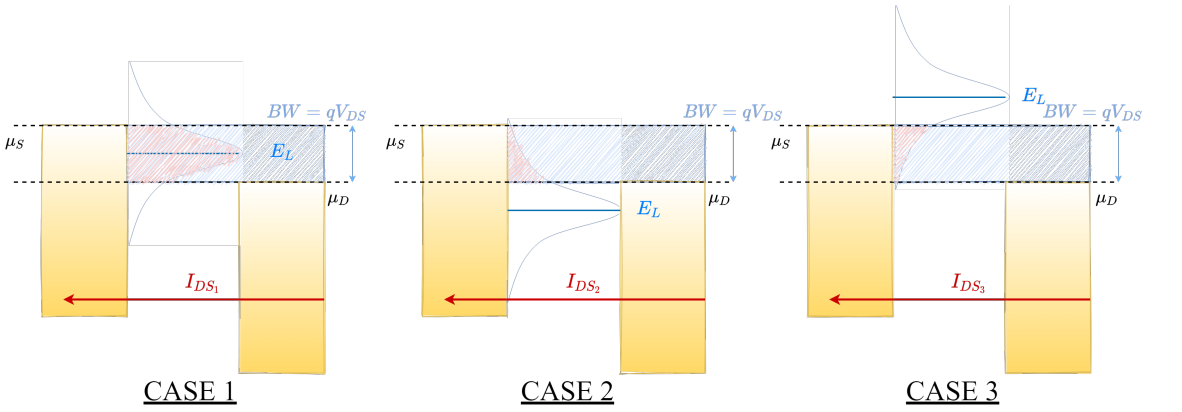


Figure 3.9: Case 1,2,3 described above in the case of a single broadened level. The voltage division factor $\eta = 1/2$ shifts the Fermi level of the source $qV_{DS}/2$ above and the Fermi level of the drain below. In this case $I_{DS} \neq 0$ in all the three cases, since always a portion of the transmission 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

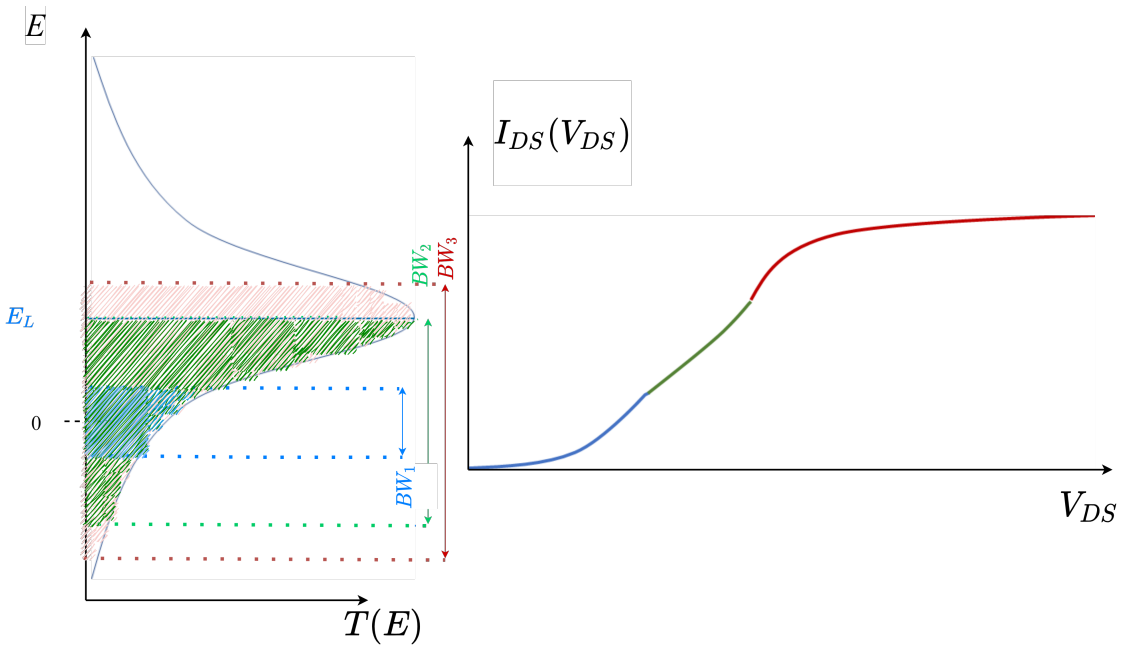


Figure 3.10: Transmission spectrum profile $T(E)$ on the left side and the corresponding drain current on the right side. The Fermi level of the system is placed at the refernce value 0.

obtained by enlarging more 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:

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

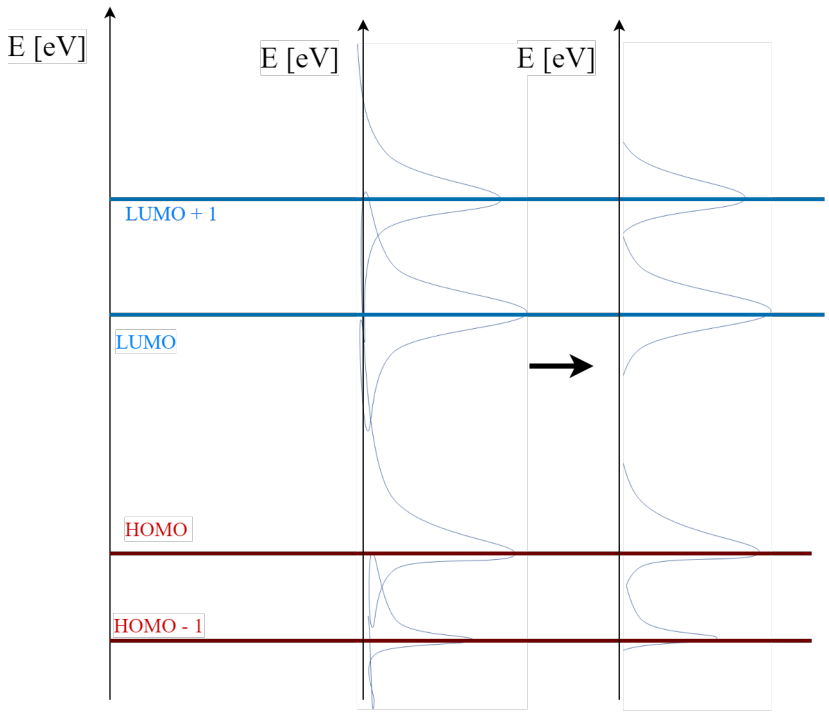


Figure 3.11: Discrete energy levels distribution on the left side, broadened levels on the left side result of the superposition of the single broadened levels.

- 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.

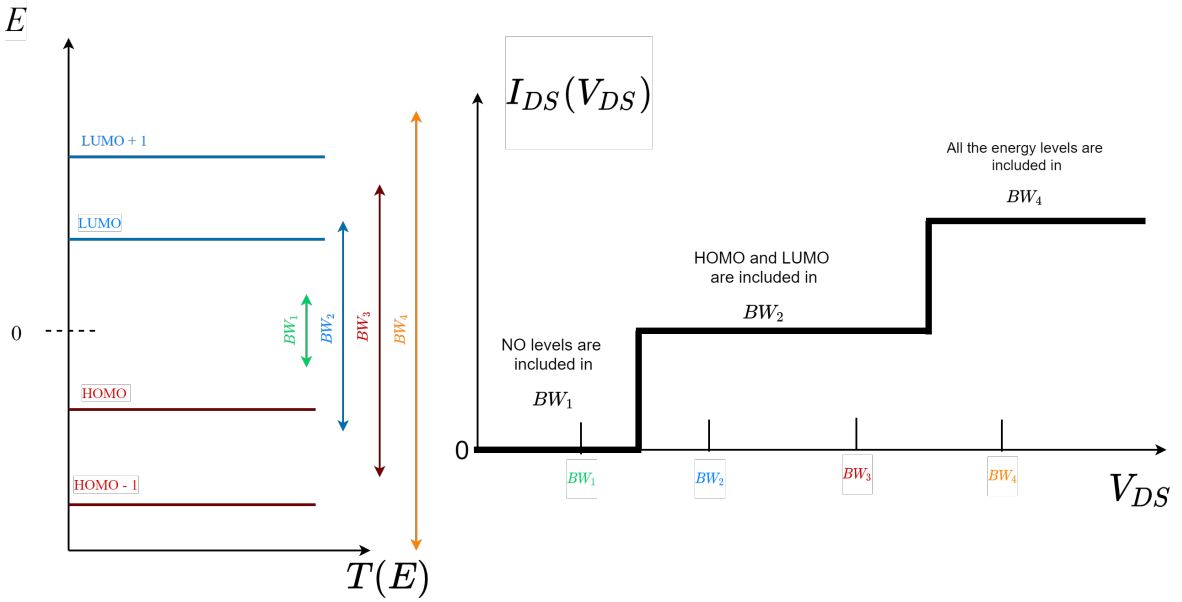


Figure 3.12: Discrete energy levels distribution on the left side, the resulting IV curve on the right side.

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 by an

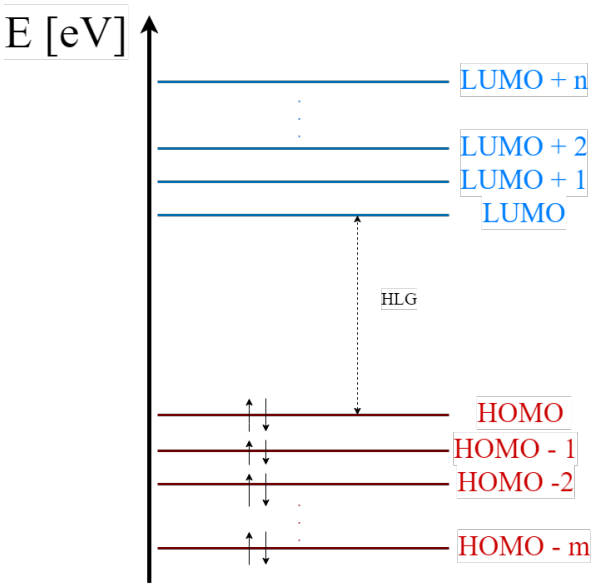


Figure 3.13: An example of energy levels distribution of MOs of a molecule. The energy level $HOMO-m$ is the ground states. All the HOMO levels are occupied at equilibrium, whereas the LUMO levels are unoccupied.

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 [60], 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 bonds. Among them,

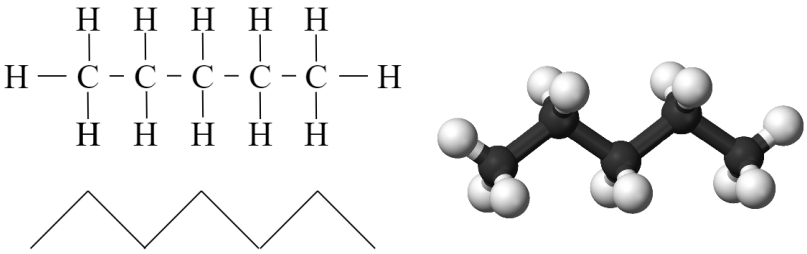


Figure 3.14: Chemical formula of Pentane C_5H_{12}. It belongs to alkanes, a type of acyclic saturated hydrocarbon.

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 effective in the

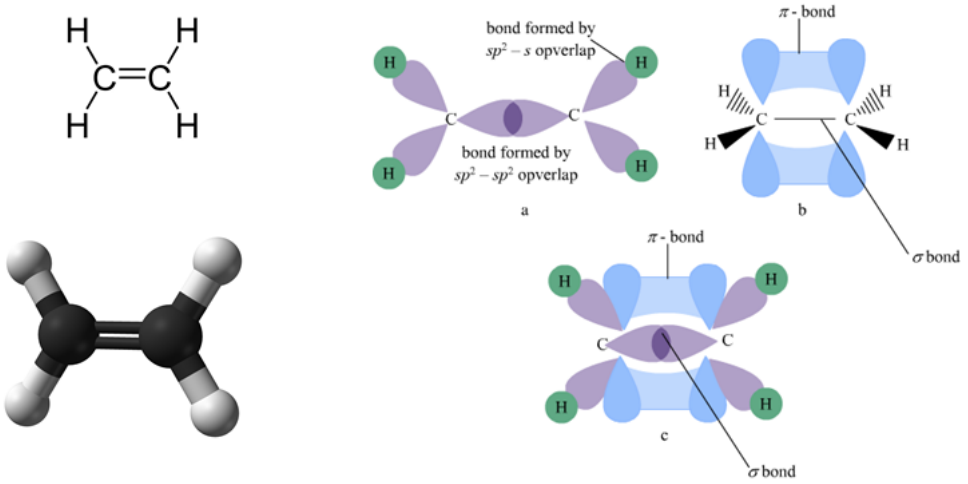


Figure 3.15: Chemical formula of Ethene C_2H_4. Representation of the molecule with balls and stick format, and on the rigth side te depiction of orbitals hybridization and formation of two pi-bonds.

electron transport [60].

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 [61].

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) [62, 63].

OPE derivatives are conjugated molecules with a rod-like shape. Its HLG is around 3.5 eV,

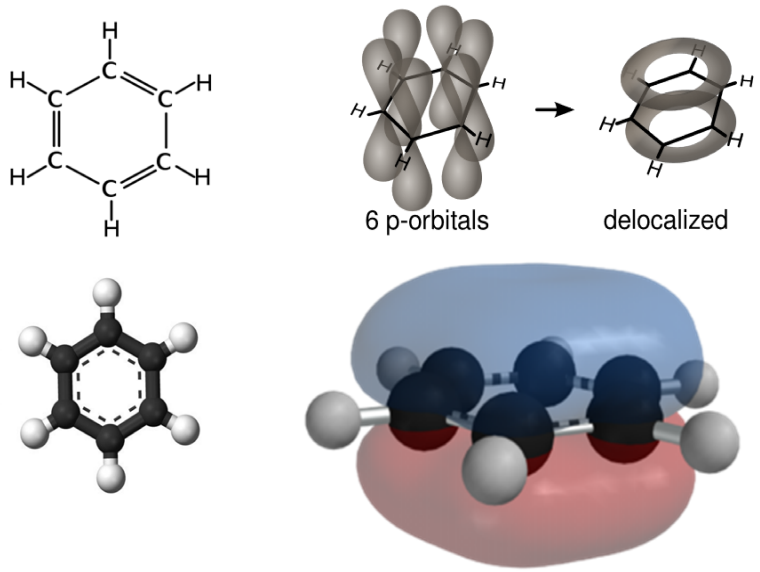


Figure 3.16: Chemical formula of benzene C_6H_6. Representation of the molecule with balls and stick format. On the right top the depiction of orbitals hybridization and formation of two pi-bonds, whereas in the right bottom the electron cloud.

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 [64], 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

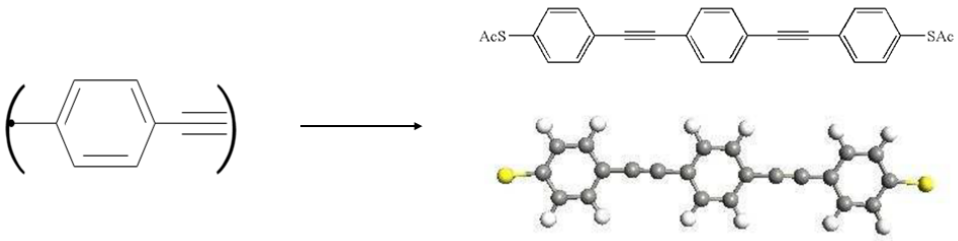


Figure 3.17: Elementary block of OPE, made of a phenyl- and ethynylene (-C≡C-) functional groups. The phenyl one is a benzene ring derivative without either 1 if the point in the figure is the beginning of the chain, or 2 atoms of H if it is in the middle of the chain. On the right side a chain of three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

electron shared able to contribute to 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.

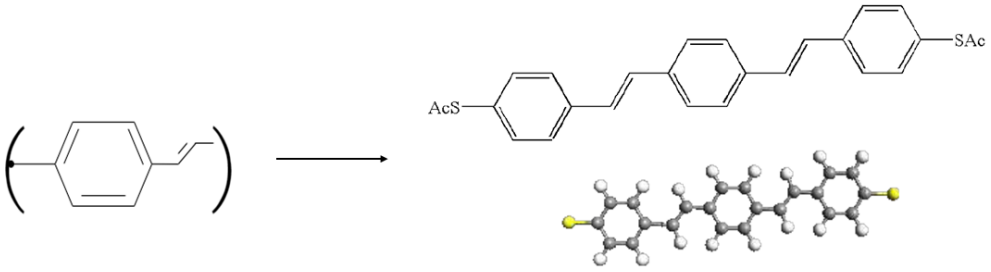


Figure 3.18: Elementary block of OPV, made of a phenyl- and vynilene functional groups. On the right side, a chain of three OPVs ended with thiol (OPV3): dithiol-3oligovynil.

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 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.

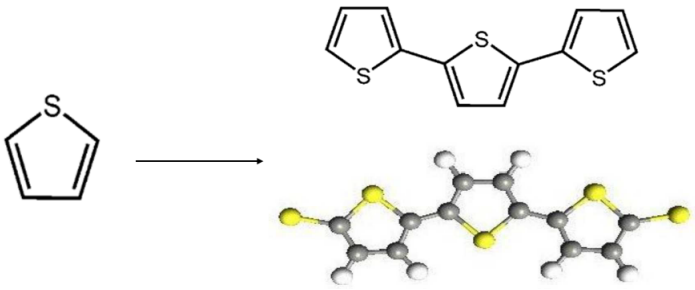


Figure 3.19: Elementary block of OT. On the right side, a chain of three OTs ended with thiol (3TT): dithiol-3thiophene.

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 wave-functions 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 trans-

mission 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 [65].

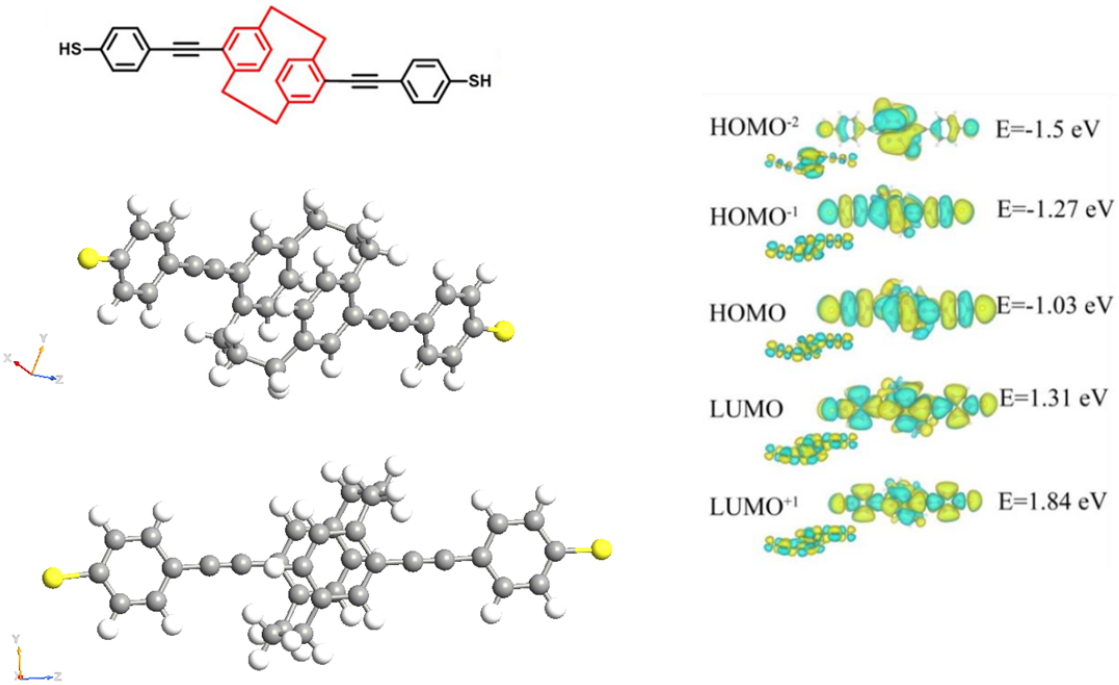


Figure 3.20: PCP molecule chemical structure on the left side. PCP molecule HOMO-2, HOMO-1, HOMO, LUMO, LUMO+1 orbitals representation o the right side.

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.

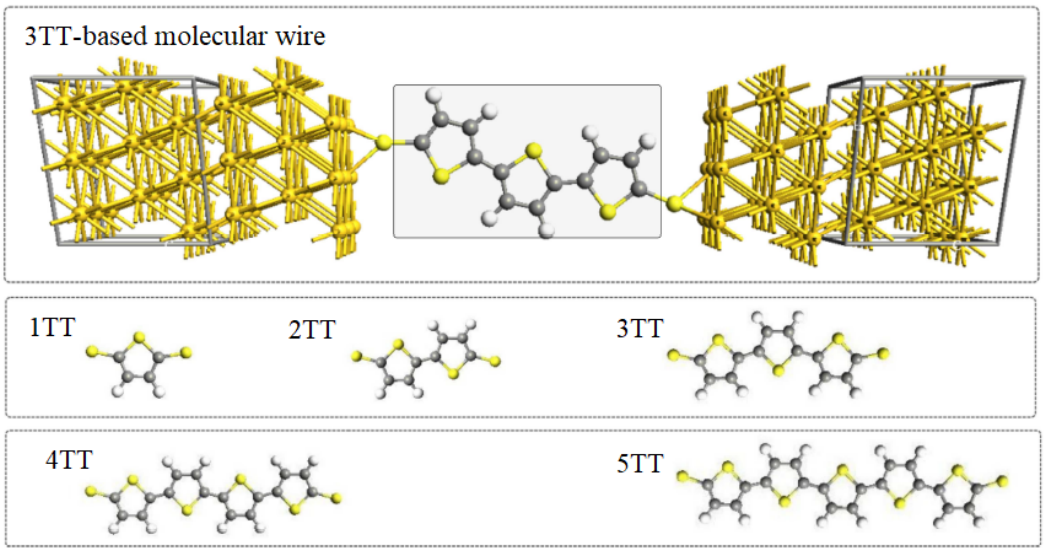


Figure 3.21: On top a typical structure of molecular junction. The molecule rappresented in the scattering region is the dithiol-trithiophenyl (3TT) Thiophene. On the bottom different OT molecules with increasing number of rings.

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, 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.

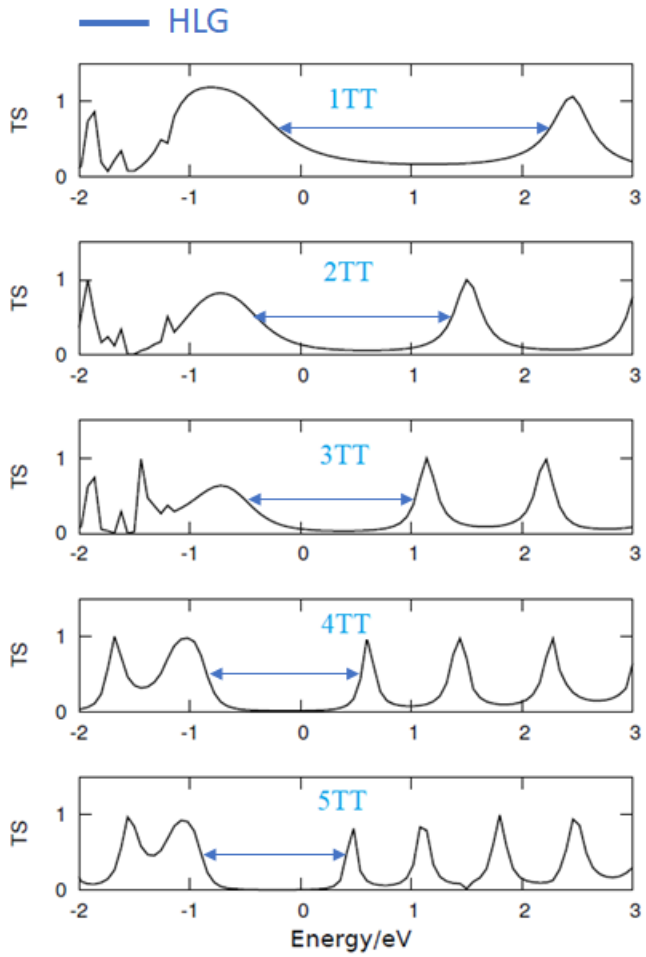


Figure 3.22: Equilibrium Transmission spectra of thiophene molecule with different number of rings. For all molecular systems the average Fermi-level is placed at 0 eV. The peaks on the left side of the Fermi level are the HOMOs and on the right are the LUMOs.

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 [66, 44]. 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.

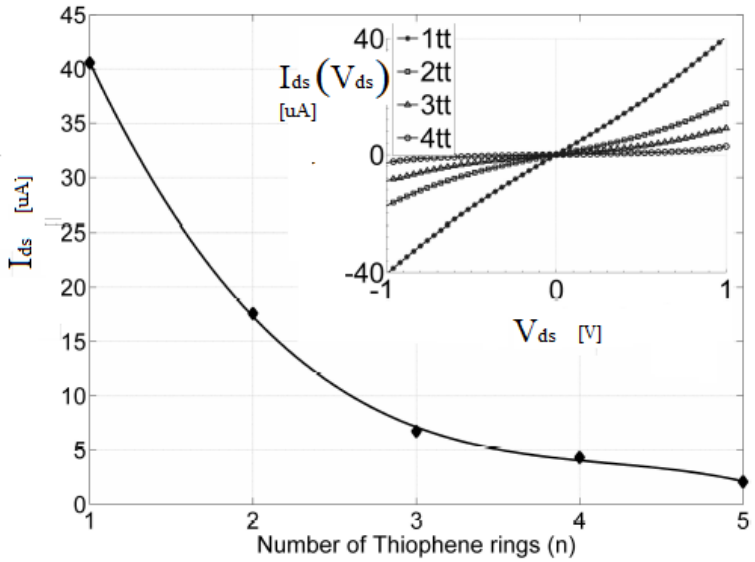


Figure 3.23: The resulting drain current I_{DS} for OT molecules as a function of increasing number of rings at $V_{DS} = 1V$. On the rigth top the IV curve of OT molecules with different lengths in the bias range of [-1V to 1V]

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 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 [67].

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

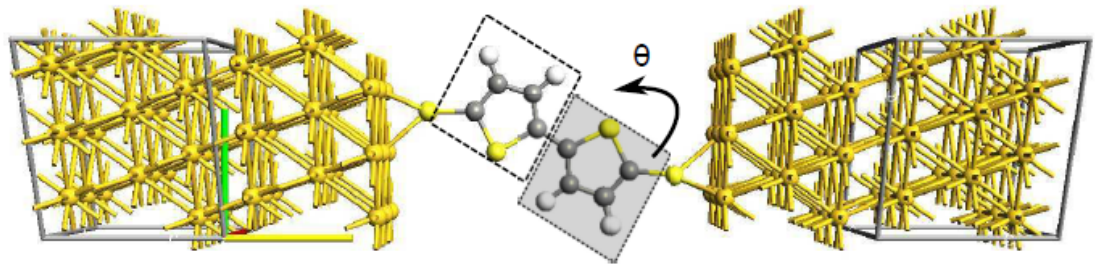


Figure 3.24: Dithiol-dithiophenyl (2TT) molecular wire. The plane of OT ring on the left side is fixed while the shaded plane is rotated with respect to the fixed one of an angle θ.

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.

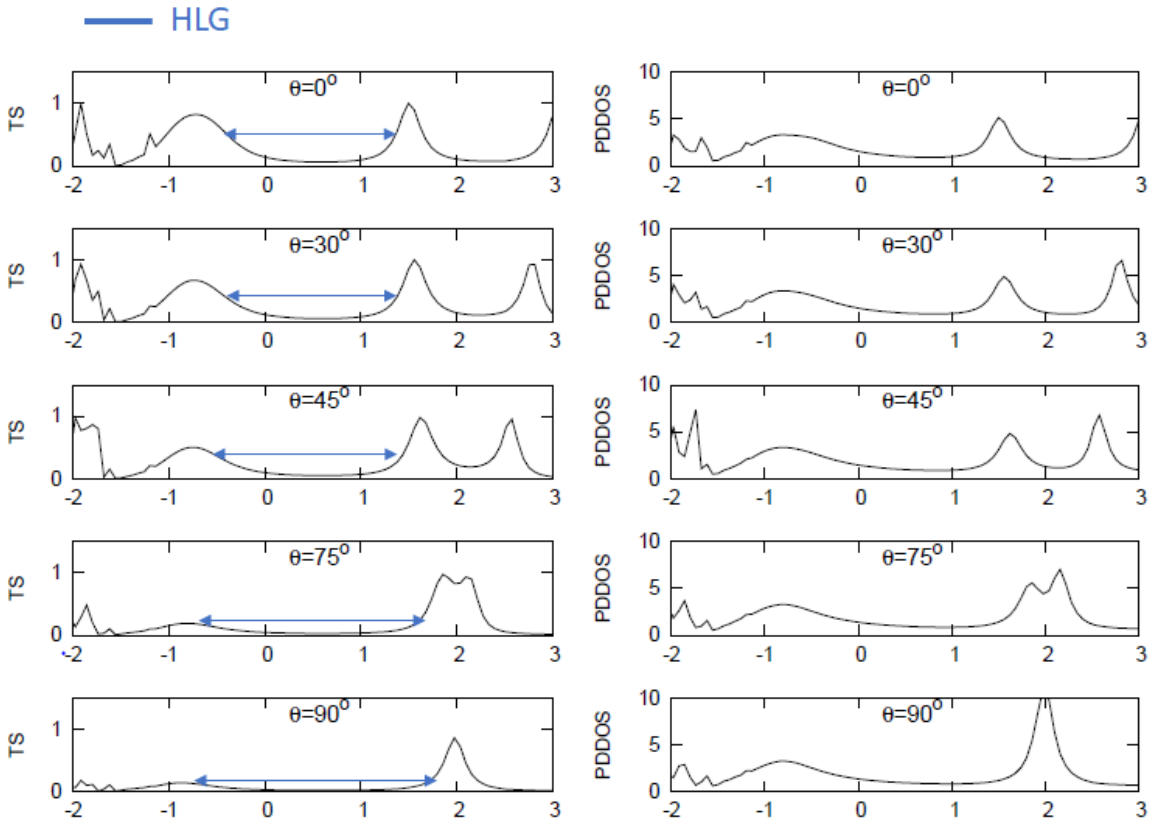


Figure 3.25: Equilibrium Transmission spectra and Projected density of states (PDDOS) of 2TT molecule with different degrees of inter-ring torsion. For all molecular systems the average Fermi-level is placed at 0 eV. The peaks on the left side of the Fermi level are the HOMOs and on the right are the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

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 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.

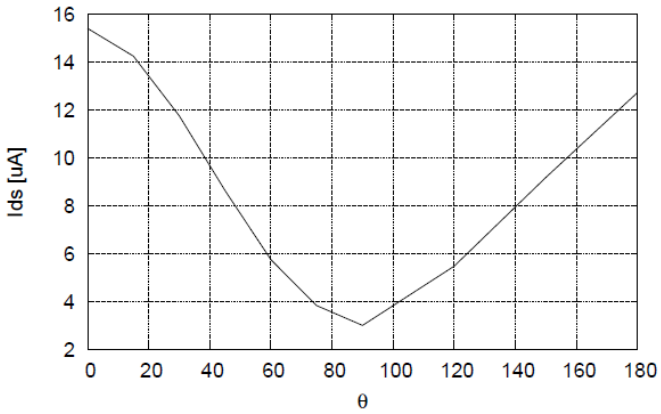


Figure 3.26: The resulting drain current I_{DS} for OT molecules as a function of increasing number of rings at $V_{DS} = 1V$.

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 [68, 69]. From [68], 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

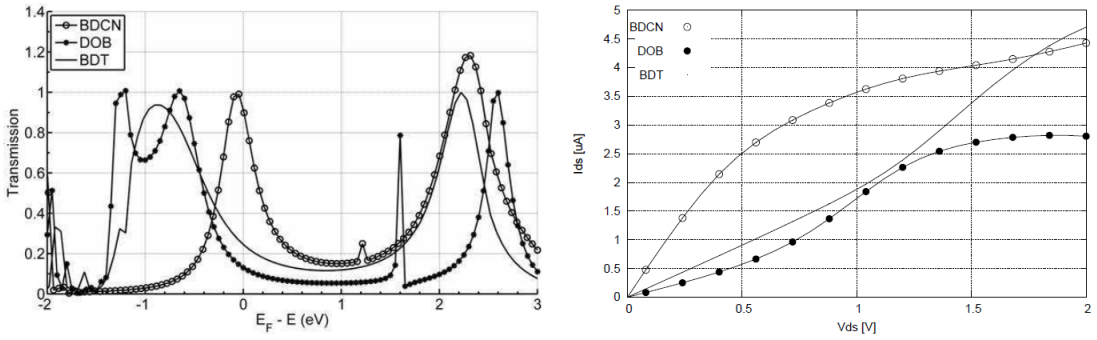


Figure 3.27: The equilibrium transmission spectrum on the left side and the resulting current on the rigth side, in the case of a benzene terminated with thiol (-SH), cyanide (-CN), oxigen (-O) groups.

choice of the type of linker. 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 equilibrium 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 (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.

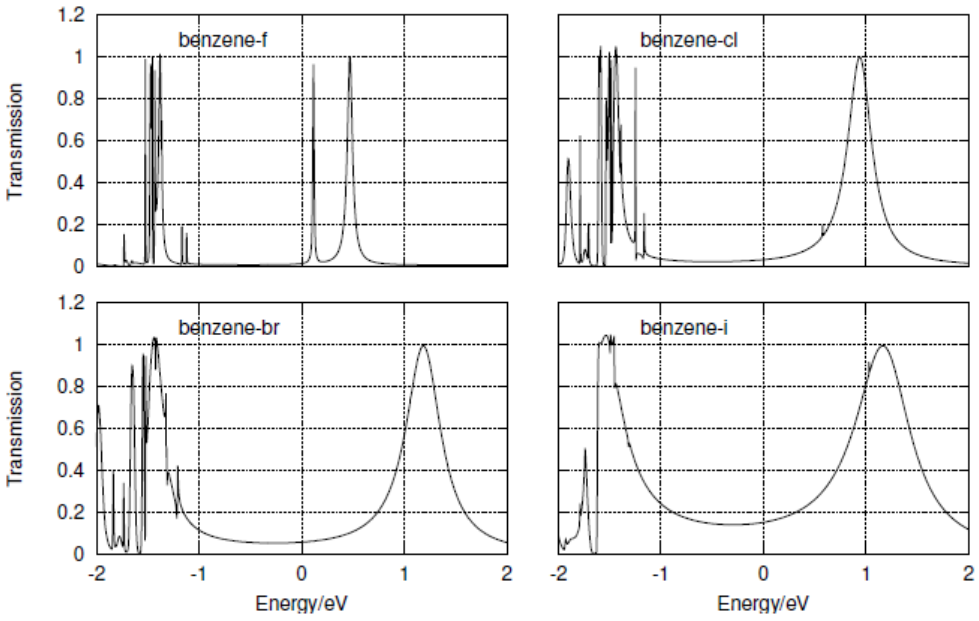


Figure 3.28: The equilibrium transmission spectrum in the case of a benzene terminated with Fluorine F (top left), clorine Cl (top right), bromine Br (bottom left) and iodine I (bottom right).

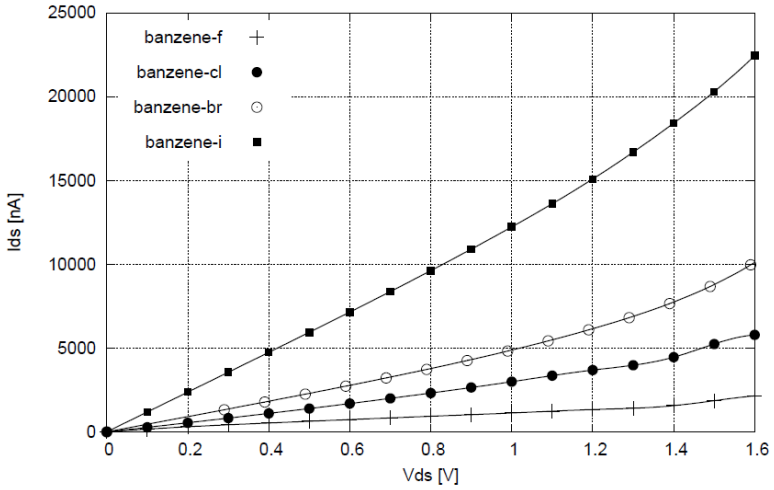


Figure 3.29: The drain current of the molecular wire in the case of a benzene terminated with Fluorine F (crosses line), clorine Cl (full dotted line), bromine Br (empty dotted line) and iodine I (square dotted line).

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.

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.

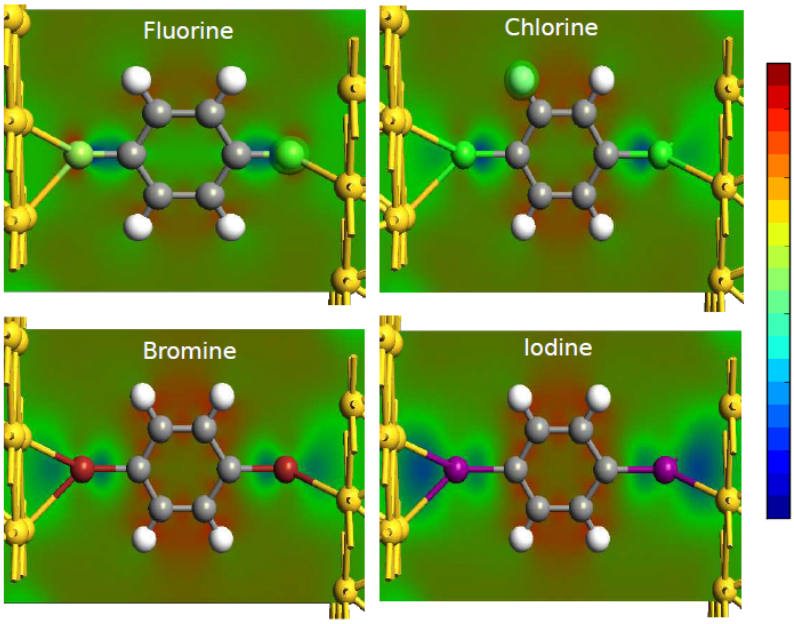


Figure 3.30: The drain current of the molecular wire in the case of a benzene terminated with thiol (-SH), cyanide (-CN), oxigen (-O) groups

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.

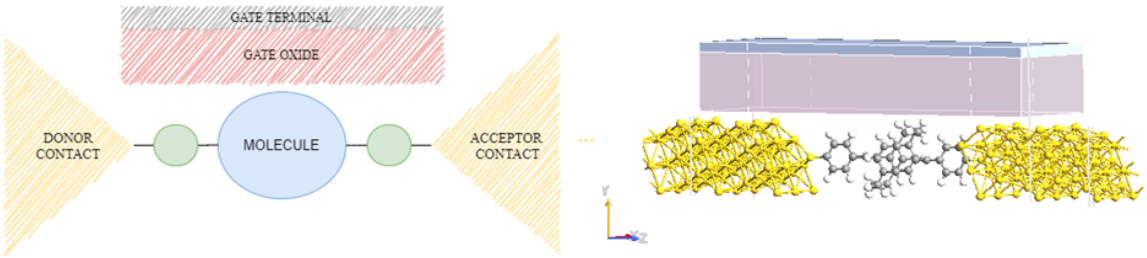


Figure 4.1: A depiction of a molecular FET with a single gate terminal on the right side, and a more realistic representation of the geometry of the device.

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 [70]. Two molecules are employed as quantum dot in the MolFET configuration: OPV7 and PCP.
The former has been widely investigated in literature [71, 72, 73], also for MolFET applications [74, 63]. The latter instead, has been only experimentally explored recently [65].

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 **??**, **??**,**??**.

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 [75], mechanical stress[76], heat [77], electrochemical stimuli such as for e.g. Van der Waals forces [78] or more commonly voltage [79, 80, 81, 82, 83, 84]. 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* [82], 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.

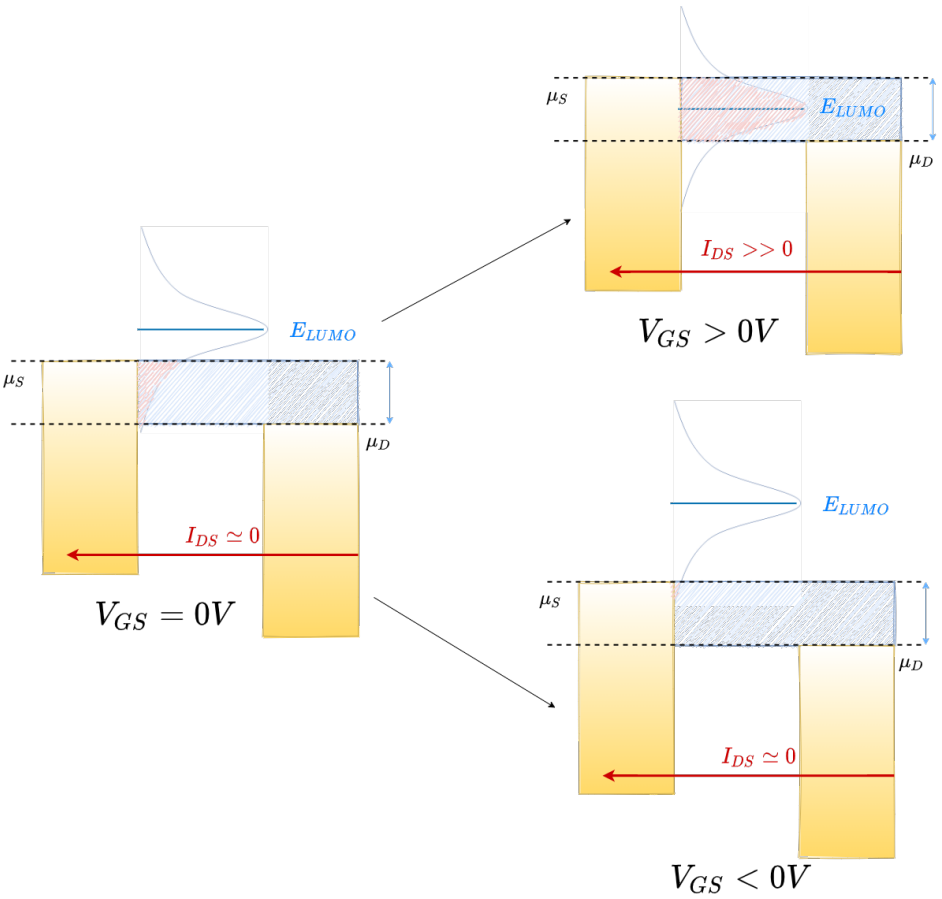


Figure 4.2: Shifting of the transmission spectrum of LUMO level according to the gate voltage applied.

- 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 [84], 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 $0.05\,\mathrm{eV/Å}$. 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 [59]. 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.

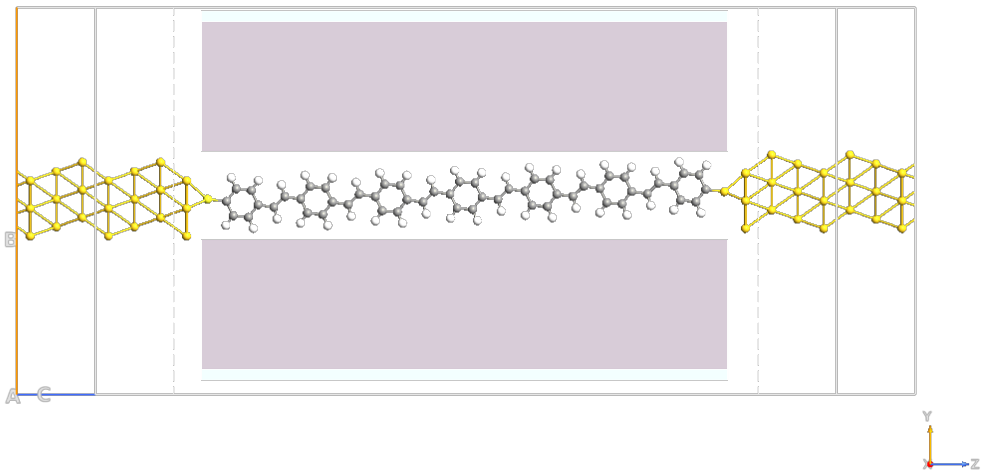


Figure 4.3: OPV7 double gate transistor geometry.

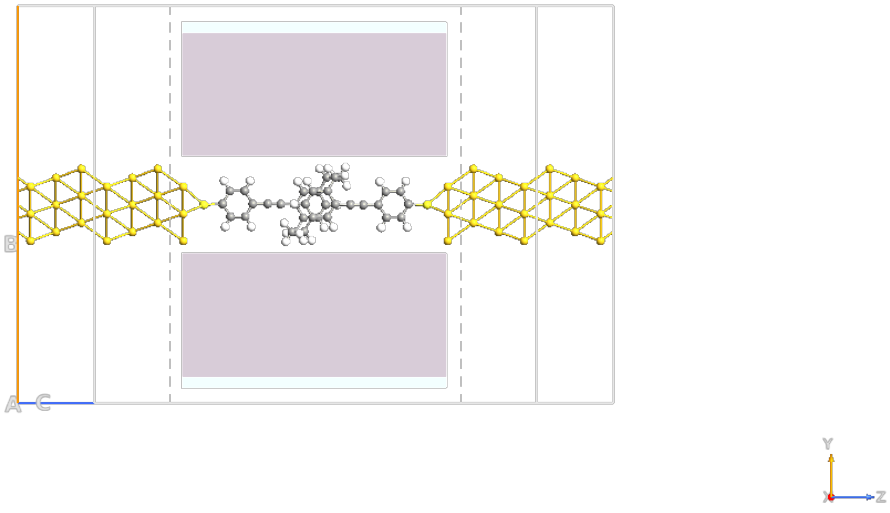


Figure 4.4: PCP double gate transistor geometry.

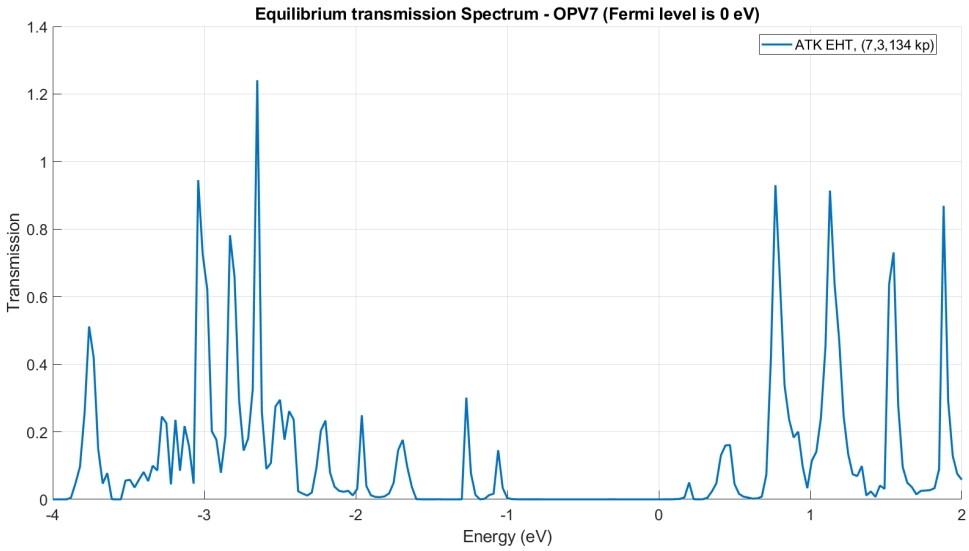


Figure 4.5: Equilibrium TS for OPV7 double gate transistor.

- **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 **??**. 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 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 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 **??**. 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

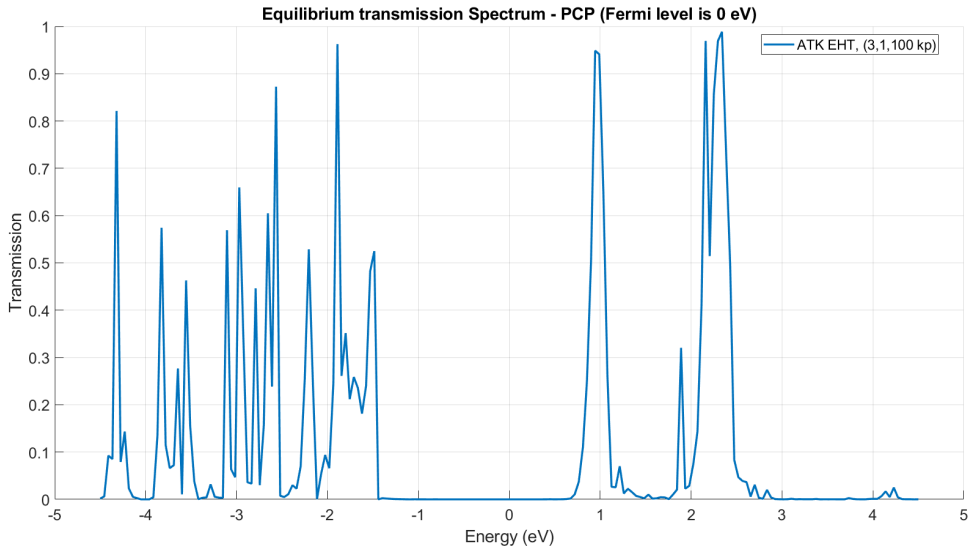


Figure 4.6: Equilibrium TS for PCP double gate transistor.

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 16.30 and comments, for sake of readibility, are directly provided in the captions.

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

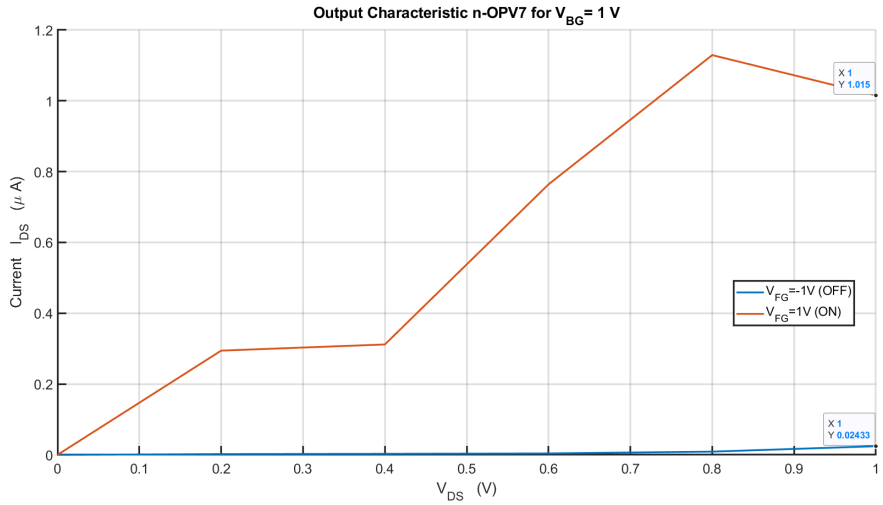


Figure 4.7: Output characteristic of n-OPV7 MolFET. A ratio $Ion/Ioff \simeq 42$ is achieved. There is a little Negative differential resistance (NDR) trend, that poses the question how this trend will impact at circuital level (for e.g. case of CMOS-like inverter in inversion region?).

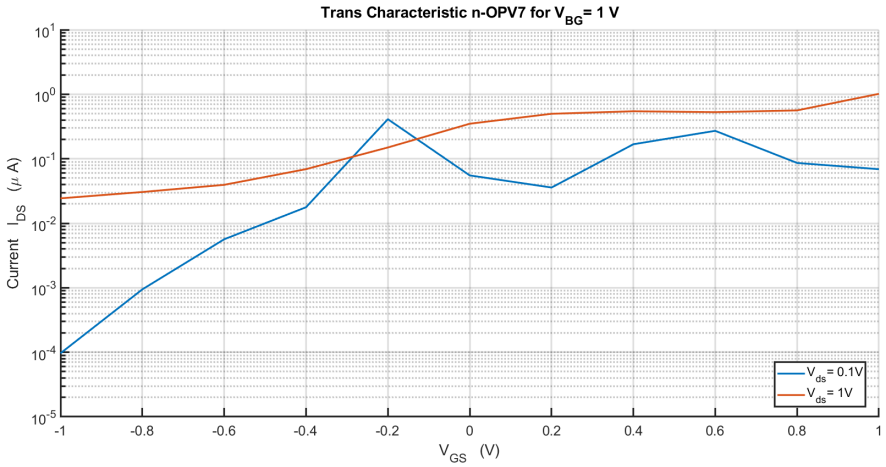


Figure 4.8: Trans-characteristic of n-OPV7 MolFET in log scale. Are evident the deviations wrt conventional MOSFET: there is not a well defined threshold voltage especially for the transchar at $V_{DS} = 0.1V$ and moreover a prominent DIBL is present. The SS that could be defined in a more or less linear 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**

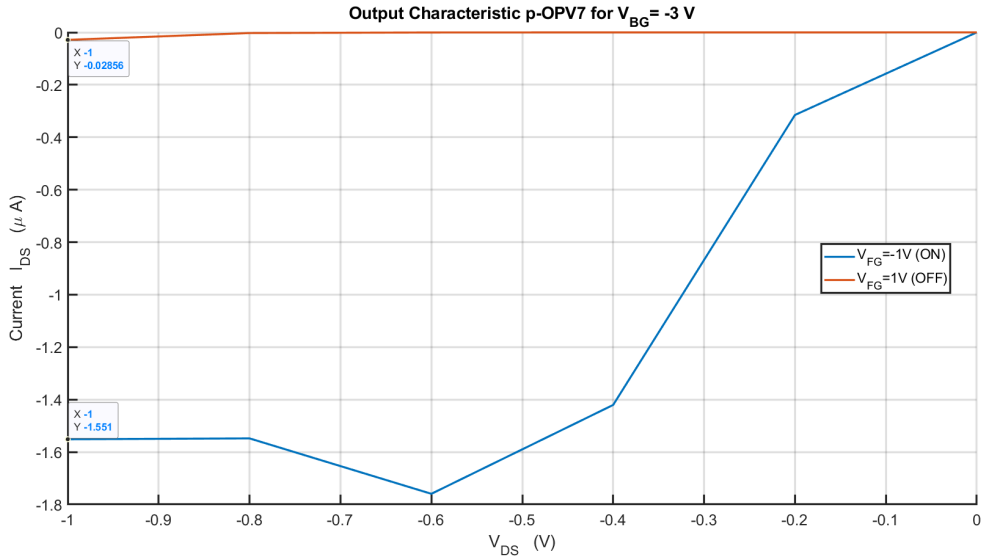


Figure 4.9: Output characteristic of p-OPV7 MolFET. A ratio $Ion/Ioff \simeq 54$ is achieved. Also here is present a little NDR trend.

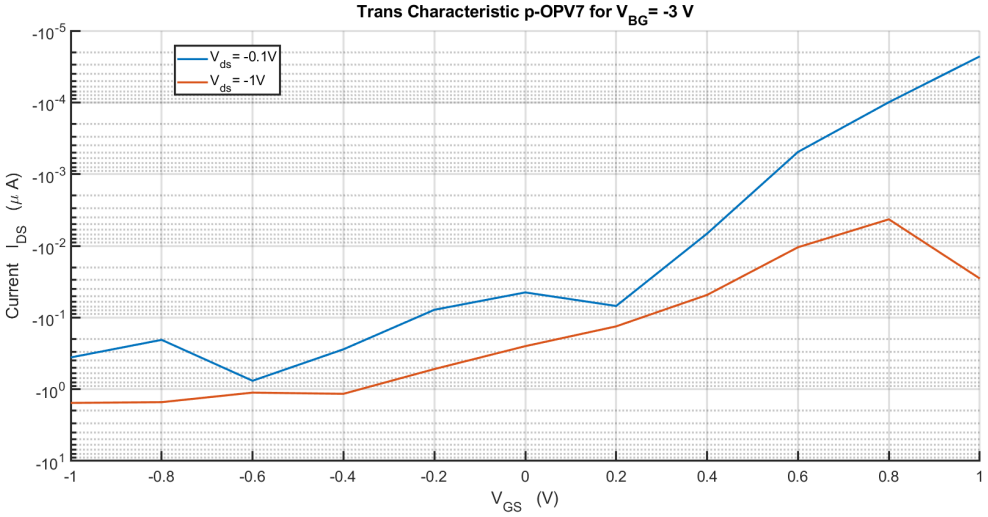


Figure 4.10: Trans-characteristic of p-OPV7 MolFET. Same comments done for n-opv7 transchar.

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

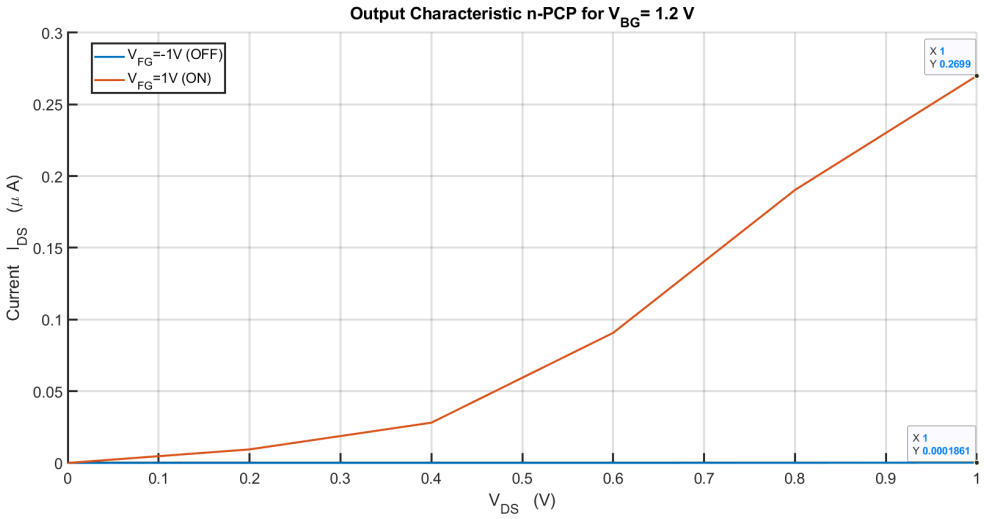


Figure 4.11: Output characteristic of n-PCP MolFET. A ratio $Ion/Ioff \simeq 1452$ is achieved. Here is not present a NDR trend, and the characteristic is almost linear in the range analyzed.

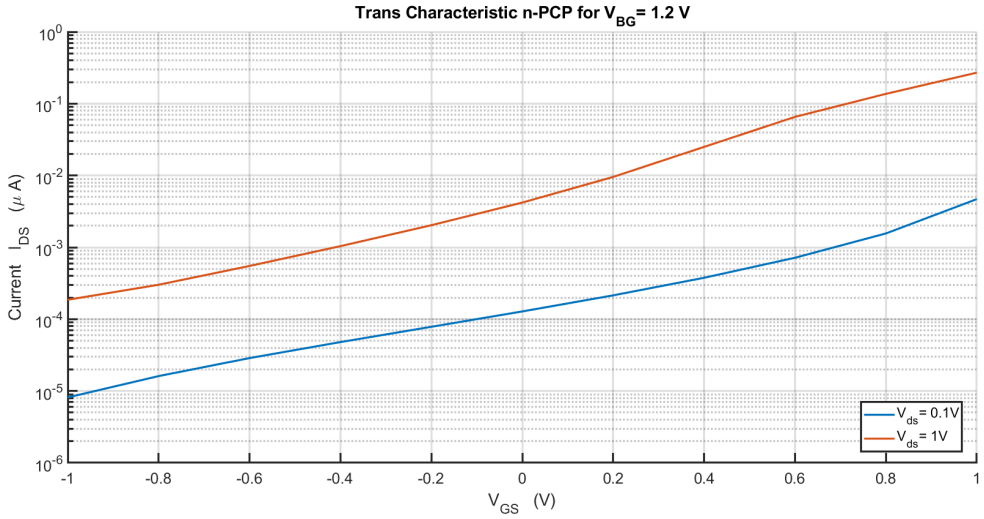


Figure 4.12: Trans-characteristic of n-PCP MolFET. Here the deviations wrt conventional MOSFET are very evident: the DIBL at 1V is almost of the order of 10^2. On the other hand, in this range is it possible to better define the threshold voltages (that are evidently subjected to roll off) and the SS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

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

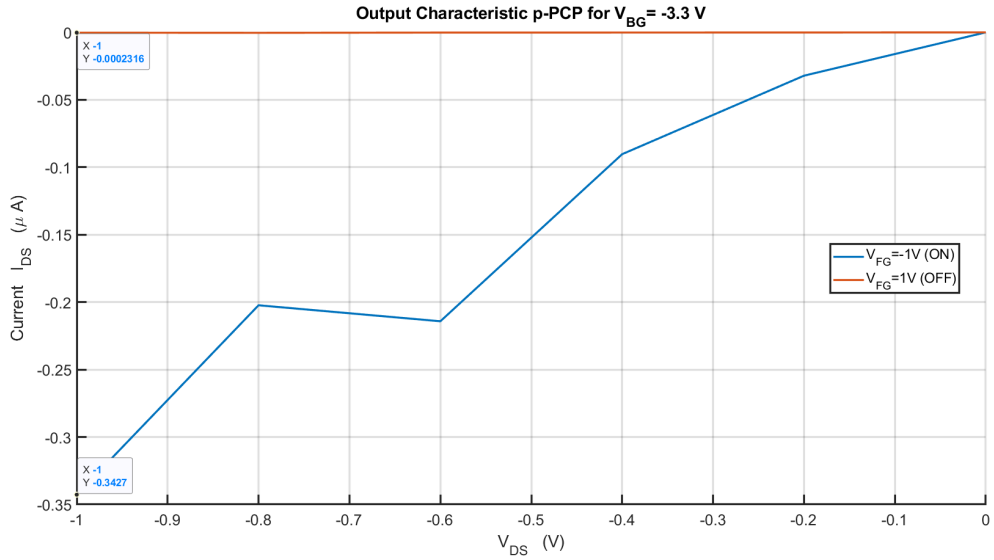


Figure 4.13: Output characteristic of p-PCP MolFET. A ratio $Ion/Ioff \simeq 1478$ is achieved. Here there is a little a NDR trend, and no saturation in the range analyzed.

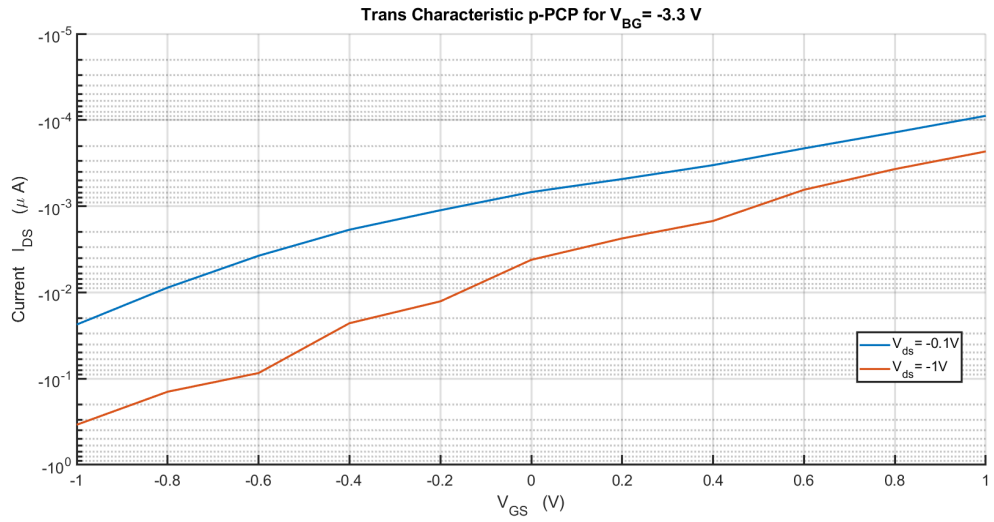


Figure 4.14: Trans-characteristic of p-PCP MolFET. Same comments done for n-PCP transchar.

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.

CHAPTER 5

Molecular electronic sensors

The molecular electronic sensors (MES) are introduced in section 5.1, and table 5.1 summarizes their main distinct features. At the presentcin literature there are only few molecular electronic sensor examples (a dozen of papers in total - see bibliography).
This is thus a pioneering field of electronics and sensing science and not much is already known, at least from the standpoint of practical implementations and MES performances. Nevertheless the physical phenomena and theory for transport that are behind the working principle of a MES are well known, and they are exactly the topics covered in the first part of this work.
In section 5.2 a brief review of the working principle of MES, with the gain and in terms of the rigorous theoretical models introduced in the part I and part IV of this LN. In section 5.3 I will introduce a possible methodology for proceeding in simulating and designing a MES gas sensor. Notice that the SETs, and sometimes also molecular channel SETs, are often used and demonstrated to be effective charge sensor. Also the elementary single electron charge can be detected and moreover by exploiting the Coulomb blockade and injecting a spin-polarized current it is possible to perform spin detection. This application is useful is quantum computing memory reading operations, since the information is encoded by means of the electron spin. Nevertheless such applications are outside the purposes of this work, that instead is focused on molecular sensors, and since at moment in literature such applications are barely considering molecular channels, but semiconducting (or metal) quantum dots instead, they are not considered here.

5.1 Molecular electronic sensors

The purpose of this section is to provide an overview about molecular electronic sensors and the context in which they find place today, in terms of currently used sensing devices. As mentioned at the end of the previous section, the recently developed both theoretical and experimental techniques and technologies for molecular electronics (beyond the photolithography limit) allowed for the fabrication of single molecule devices (or even small packets of molecules), to be used like conventional electronic devices (both digital and analog) or like sensing devices. The latter application is briefly introduced in this section and it is the subject of the rest of this work, widely investigated in the next chapters.
A molecular electronic sensor (sometimes shortened MES) is a molecular wire, i.e. a single molecule or a packet of molecules in between two contacts, or alternatively a molecular transistor, i.e. a molecular wire with one or more electrostatically coupled gate contacts, in which the channel conductance (or sometimes capacitance or other physical properties) is suitably and uniquely changed by the detection of chemical or physical events, such as the presence of a given chemical compound in proximity of the channel (gas sensor) or alternatively chemically bonded with the channel (chemical compound sensor for interacting gases, liquids etc...), or the absorption of a photon at a given wavelength by the device, or again the change in the relative

ambient humidity, temperature, and so on... Even single electron charge detection or more in general the proximity electrical charge detection is possible. Examples of MES, from which the above reported definition was derived, are reported in literature during recent years, e.g. [32], [33], [34], [43], [85], [86], [87], [88], [89], [90], [38], [39].
It is important that regardless of the specific chemical/physical phenomenon that produce the MES channel conductance variation, the sensor or device physical model for transport is always the same, and it coincides with the physical model for transport in mesoscopic systems (i.e. in that systems small enough to make classical physical out of its regime of validity but big enough to make a standard quantum mechanical approach unmanageable [91]). The physical model for transport in mesoscopic systems will be investigated more in details in the chapters of part I of this work. Thus once the final details of the molecular conducting channel are known, the transport calculations will be performed always in the same manner. The differences from an application to another are mainly two: the specific molecule used as conducting channel for a specific MES, and the physical or chemical event that causes the variation in the channel conductance (or device capacitance, etc...). Concerning the employed molecule for the MES, depending on its specific electronic and transport features different conductive behaviors may exist [23], [28], [26], [25], [92], [93], [35], [36], [94]. The specific molecule is not the only variable that influences the transport. Also the contact material (e.g. metal, graphene, etc...), its quality or orientation and the way through which the molecule is anchored (the anchoring group) to the contacts (and if the anchoring is of good quality or strong/weak) influence the transport features [23], [95], [96], [97], [85], [98]. Instead, concerning the physical or chemical phenomenon at the basis of the conductance variation, each specific case is different depending on the particular event that should be detected. In the case in which some physical parameters such as temperature or humidity are to be detected, the possible range of values should be considered during calculations (e.g. the temperature will affect the Fermi-Dirac statistics and thus the occupation of the electronic states within the channel) and the corresponding conductance sensitivity is easily found. In other cases the presence of an electrical charge in proximity of the channel can affect the conduction due to Coulomb interaction with the electrons within the channel, or the absorption of a photon at a specific wavelength can vary the number of carriers that take part to the conduction thus changing the resulting electrical current (exactly like in conventional optoelectronic devices). Again, the creation of a chemical bond between a target molecule and the channel molecule will obviously change the channel transport properties thus resulting in a current variation. From the standpoint of the physical/chemical event to be detected, the simplest case (along with the temperature or humidity variation) is the one in which the presence of a target chemical compound in proximity of the molecular channel of the sensor makes varying its conductance in an appreciable way, due to weak chemical/physical interactions such as the van der Waals ones. The latter is essentially the only case considered in this work. Notice that this could be the case in which a barely chemically interacting molecule is used as conducting channel, and it interacts through van der Waals forces with a target chemical substance, thus leading to a gas sensor [34], [43], [85], [86].
The MES functioning principle is quite similar to the conventional FETs-based sensor that constitute the current state of art in electronic sensors or bio-sensors [41]. For example in a GNR FET based gas sensor the detection of the target substance is possible because of the conductance modulation due to direct interaction between the graphene nano-ribbon (i.e. the FET channel) and the target molecule (through van der Waals interaction or through “chemisorbed” interaction - i.e. with the creation of a chemical bond between the surface and the adsorbate [99], that is the target molecule) [100], [101], [102]. Otherwise the detection of whatever chemical compound or substance in gas or liquid solution is possible by means of conventional CMOS/-MOS FETs, floating gate FETs and other FETs structures, again thanks to a modulation of the channel conductance, that can be enhanced by means of chemical interactions between the target molecules and specific receptors anchored to the gate [41].
Notice that, especially in biological applications, the biosensor target substance is often called “analyte”, where the analyte is a substance or chemical constituent that is of interest in an

analytical procedure [103], in this case the sensing or measuring process.
Currently the FETs used for sensing applications are mainly two-dimensional (2D) devices such as thin-film FETs or Organic FETs (OFETs), in which a sheet of organic material, e.g. graphene or other organic polymers, is placed between the source and drain contacts and its conductance is modulated by the event that should be detected [104]. Other possible devices are very miniaturized conventional FETs, CNT FETs, NW FETs (both silicon NW or other materials NW), etc... [40], [39], [38], [42], [105]. In all these devices the working principle is always the same: the conductance of the channel is varied depending if the target substance is present (and in what concentration) in proximity of the sensor or not; or alternatively depending if the target substance is chemically bonded (i.e. chemisorbed [99]) with the channel or not. Actually, the most general functioning principle should be expressed as follows: the sensor conductance is varied accordingly to the chemical or physical event to be detected [106], [41], [40]. Indeed, not only gas sensors are possible, but, exactly like in the case of MES, many other chemical or physical events and quantities can be detected. For example, a 2D FET can be mechanically strained thus leading to a Nano-Electro-Mechanical System (NEMS) or sensor, able to detect strain and mechanical deformations [107], [108].
Anyway, the point now is that the functioning principle of molecular electronic sensors and FET-based sensor is essentially the same [41], [38]. This should not be surprising since both are essentially FETs (or wires), even if in MES the channel is a single molecule (or a packets of molecules). Nevertheless, in general, 2D devices are at the present favoured in literature for two main reasons. Firstly, they can be fabricated with current technology without the molecular electronic fabrication issues mentioned in the previous section (such as the low yield, etc...) and thus many recent papers report the prototyping and demonstration of sensors [40], [104], [42], [109], [110], [111], [112], [113], [114], [115], [116]. Then, in a 2D device the interacting portion of the device is the 2D surface or sheet of organic material, thus the interaction region is bigger, leading to a greater sensor sensitivity e.g. to the target substance/gas or chemical-physical event in general [41], [105], [104], [40]. In this regard, for example in [110], the authors state:

"2D-layered materials (2DLM) have an exceptionally large surface-area-to-bulk ratio, and thus adsorbed gases can greatly modify their properties, making them promising candidates for gas detection".

The potentially lower sensitivity of MES w.r.t. current 2D devices is a significant drawback, at least for gas sensors. The problem can intuitively viewed as follows: a gas sensor should detect the presence of specific substances (pollutants, toxic particles, certain chemical compounds, etc...) that are present in the monitored environment at least in a given concentration, often expressed in parts per million (ppm) or parts per billion (ppb), e.g. accordingly to the legislative regulations [117]. The target is thus a minor portion of the total air in the environment, and in order to be detected it should be in proximity of the MES, i.e. nearby a tiny molecular device. This can constitute a limit in terms of sensitivity if a single device based on single is used instead of a 2D device for the reasons expressed above. Nevertheless a trivial solution could be to put several MES in parallel, thus increasing the interaction surface and maintaining the advantage of great miniaturization of the device [33], [106]. Notice that depending on the specific application and thus on the specific chemical/physical event that should be detected and measured by means of the MES, the sensitivity issue can be not present or completely solved [38], [39], [106]. Other issues, as already pointed out, are concerning the current fabrication processes of molecular devices (see the discussion in the previous section **??**), and the small currents involved in the sensing process (see again section **??**). The latter could be a limiting factor for detection due to the noise that unavoidably afflicts the system. Nevertheless also these issues can be often overcome (even if not always easily), for example by increasing the complexity of the front end electronics such as the signal conditioning circuit of the sensor [32], [118], [119], [120], [121], [122], [123], [124].
Besides these drawbacks the main advantages in exploiting molecular electronic sensors are gen-

erally their high selectivity, i.e. it is possible to choose *ad hoc* molecules to be used as conducting channel such that the device responds to the target event only (without the usage of intermediate recognition elements), high sensitivity, generally linked again to the possibility of choosing *ad hoc* molecules for the device channel (even if as mentioned this sometimes is not true at all for gas sensors if compared to 2D devices), real-time detection, indeed the current modulation occurs within the transit or relaxation time of the device, that is of the order of conventional electronic response times, and label-free detection. These advantages together with the above discussed drawbacks are summarized in table 5.1. The last mentioned advantage needs a bit of discussion, indeed the conventionally used lexicon must be introduced. When biosensors are taken into consideration (see below for a summary of possible application fields of molecular electronic sensors) one usually distinguishes between labeled detection processes (or sensing with label) and label-free detection processes. The distinction between the two is well explained in [106]:

"In the most general case a biosensor consists of at least two functional components: a molecular recognition element (receptor - it can be a natural, synthetic, bio-inspired molecule) that selectively interacts with its target analyte (e.g. ions, DNA, antibodies, cells, microorganisms, etc...) and a physicochemical transducer (e.g. the molecular wire/transistor). The latter converts the bio-recognition information into a measurable quantity, being in general an electrochemical, electrical, optical, magnetic, mass-sensitive, or thermal signal. Due to the fact that biological analytes are often hard to detect purely on basis of their intrinsic physical properties, biosensors often require labels such as enzymes and fluorescent or radioactive molecules attached to the targeted analyte. As a result, the final sensor signal corresponds to the amount of labels, representing the number of bound target molecules. As a drawback, label-based technologies are often labor- and cost-intensive as well as time-consuming. In addition, labeling of biomolecules can block active binding sites and alter the binding properties. Altogether, this may adversely affect the affinity-based interaction between the recognition elements and the target molecules. In contrast, label-free biosensing technologies, by definition, do not require the use of labels to facilitate measurements. Instead, they utilize intrinsic physical properties of the analytes, such as molecular weight, size, charge, electrical impedance, dielectric permittivity, or refractive index, to detect their presence in a sample."
(adapted from [106]).

From this definition it should be clear that molecular electronic sensors are always label-free sensors, in which the target analyte makes the molecular channel conductance (or capacitance, etc...) varying, making possible the detection. Notice that in general all FET-based biosensors are label-free. Indeed even in the case in which a biological entity is detected by means of a bio-macro-molecule (a protein, antibody, enzyme, etc...) chemically bond to the FET channel on one side and to the target analyte on the other side, from the above definition, it is a label-free sensor. For example a virus detection in a clinical solution (or even in air), that is made possible thanks to the FET conductance modulation due to the presence or not of the bond between the virus specific antibody that is anchored at the FET channel (e.g. a 2D graphene sheet) and the virus itself, is a label-free detection process [125], [126]. Label-free biosensing methods have made enormous progress in recent years due to their ability for rapid and inexpensive bio-detection even in small reaction volumes. Moreover, they lend themselves for integration into lab-on-chip platforms and allow monitoring the concentration of target analytes in real time [106]. In this regard molecular electronic sensors may be the next generation toward higher miniaturization, selectivity and sensitivity of label-free real-time sensors for biological applications and single molecule event detection [38].
The FET-based (standard FETs, 2D FETs, OFETs, CNT FETs, NW FETs, GNR FETs, molecular FETs or whatever kind of FET) sensors are not the only possible of biosensors. Another broad class of biosensors or chemosensors (i.e. chemical-sensors) are the so called "electrochemical sensors", that have similar important features in terms of sensors characterization (such

Table 5.1: Main features of molecular electronic sensors (MES).

Sensor feature	MES performance
selectivity: sensor response only in presence of the target	high
sensitivity: ample variation of conductance/capacity in response to target concentration variations	potentially high (not always true especially for gas sensors)
accuracy	potentially high
signal-to-noise ratio (SNR)	can be critical due to the low currents in molecular devices
response time	fast
real-time detection: sensor response speed	good/fast (comparable with current electronics)
linearity	good (depending on device structure)
repeatability	good (may deteriorate in long-scale times due to molecular stress e.g. variations of temperature/humidity or applied field)
reliability	good (can be worsened by process variations and device stress)
labeled / label-free	label-free
reversibility	good (may be in contrast with selectivity)
stability/drift over device life-time	good (depending on process variations and device stress)
cost	(potentially) low - see sec. **??**
size	small (potentially suitable for wearable/implantable bio-systems)

as sensitivity, selectivity, real-time, labeled/label-free, etc...) to the FET-based ones [105]. An electrochemical sensor is usually a label-free sensors often suitable for real-time detection of biological or chemical substances (often in solutions). In few words it is constituted by an electrode (usually an inert metal -e.g. gold- or graphene/polymeric electrode) immersed in the biological or clinical solution that contains the target molecule or chemical compound, above which suitable receptors (such as antibodies, enzymes, proteins, organic molecules/polymers, etc...) are anchored to it and suspended in the solution. The target analytes (molecules, bio-macro-molecules, ionic species, etc...) are then chemically bonded to the receptors producing a change in the resistance of the portion of solution in between the electrode and the top-side of bio-solution (or alternatively in between two contacts). This is then detected by means of amperometric methods or voltammetric methods (sometimes dynamic techniques are used: conductometric methods). Notice that if the electrical resistance is varied due to the presence of the target analyte the sensor material is often referred as chemoresistor [42], [105], [39], [40]. A comparison between FET-based sensors and chemoresistors (electrochemical sensors) is reported in table 5.2 (SSES -solid state electrochemical sensors- and QCM -quartz crystal microbalance-

are not considered here - see [40] and references herein for details). In this optics, MES have the advantage of having high selectivity and low cost (for the rest they are analogous to FETs).

Table 5.2: Comparison of different gas sensing technologies. SSES (solid state electrochemical sensors) and QCM (quartz crystal microbalance) are not considered in this work (see [40] and references herein for details). A comparison is provided in detail in [40].

Sensor technology	Sensitivity	Selectivity	Speed	Cost	Size
Chemoresistor	high	medium	high	low	small
FET	high	medium	high	medium	small
SSES	high	good	high	low	large
QCM	high	poor	medium	high	medium

The other ample class of above mentioned biosensors are the labeled ones, in which a label is exploited in order to make detectable a given event or bio-(macro)-molecule or substance in general. Labels have the role of converting the bio-recognition information into a measurable quantity, e.g. electrochemical, electrical, optical, magnetic, mass-sensitive, or thermal signal [106]. For example a widely spread labeled sensors are the one exploiting fluorescence (i.e. photon re-emission) depending if the target is present or not (with an intensity dependent on concentration). More in general, in optical molecular sensors a chemical reaction occurs between the target analytes and the labels such that photons at a given wavelength (typically in the ultra-violet or visible range) are emitted (or absorbed) and detected with conventional optical or optoelectronic sensors [106], [127]. These classes of sensors are beyond the purposes of this work and they will be no more considered. Nevertheless notice that they are called "molecular sensors" as well. In particular, in this regard, a "molecular sensor" or "chemosensor" is a molecular structure (organic or inorganic complexes) that is used for sensing of an analyte to produce a detectable change in a signal (often electromagnetic radiation but it is not said - see before). The chemosensor working principle is based on interactions occurring at the molecular level, usually involving the continuous monitoring of the activity of a chemical species in a given matrix such as solution, air, blood, tissue, waste effluents, drinking water, etc... and it is a labeled detection process [127]. As already said, since chemosensors are based on molecular recognition at molecular level, they are often referred as "molecular sensors" [127], even if they are completely different from the "molecular electronic sensors" that are the subject of this work. For this reason and in order to avoid confusion, so far I have always specified "electronic" together with "molecular sensors". In the next chapters if no misunderstanding is possible I will sometimes shorten the name in "molecular sensors", even if I will always refer to "molecular electronic sensors" (MES).

At this point it should be clear what molecular electronic sensors are, their basic working principles, and the ample background in which they find place. Moreover the basic lexicon was briefly introduced.

In conclusion to this introductory section, I would like to briefly summarize the possible fields of application of MES, on the basis of what is the current state-of-art in sensing science. The following list details some of the possible applications of molecular electronic sensors, and even if it is very likely incomplete, it gives the idea of the plenty fascinating applications of this emerging field of electronics and sensing:

- gas sensor: this application was already discussed above, the recent years demand is for on-line/*in situ* detection of gases for both outside and inside environment monitoring [40]. Possible markets are related to pollution and pollutants detection of particles, molecules, heavy metals, to chemical industrial processes monitoring (toxic gases can be a collateral product in case of fault in some industrial processes), or fault in public or private buildings (e.g. carbon monoxide or methane emission in dangerous concentrations) [40], [117], [105]. The main substances to be detected in these applications are usually formaldehyde (HCHO), benzene (C_6H_6), acetone (C_3H_6O), carbon monoxide (CO), carbon dioxide (CO_2), sulfur dioxide (SO_2), ammonia (NH_3), hydrogen sulfide (H_2S), ozone (O_3), nitric oxide (NO), nitrous oxide (NO_2), methane (CH_4), butane (C_4H_{10}), ethane (C_2H_6), and so on... [40], [117].
- environment monitoring: for air-quality tests and gas concentrations monitoring in outside and inside environments see above (gas sensor); for soil or water pollution monitoring the application is similar but the sensor is designed to be used in solutions in which the target analyte is present. Usually in this last application field the detection of the following chemical compounds is of importance: ammonia (NH_3), methane (CH_4), nitric oxide (NO), sulfur dioxide (SO_2) and heavy metals [40], [105].
  It is possible to include in the environment monitoring set of applications the outdoor/indoor air quality tests, factory emission detection, vehicle exhaust control of carbon monoxide (CO), carbon dioxide (CO_2), nitric oxide (NO), nitrous oxide (NO_2) and all the related applications [40].

- public security: the public security, especially anti-terrorism, is becoming extremely urgent [40]. The applications, in this field, mainly focus on the detection of the explosives and/or the nerve agents. Normally, the vapor pressure of explosives, e.g. 2-methyl-1,3,5-trinitrobenzene (TNT), 1,3,5-trinitroperhydro-1,3,5-triazine (RDX), octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (HMX) are 9 ppb, 4.9 ppt, and 0.25 ppt, respectively. Therefore, the detection of explosives needs ultra-sensitive sensors, exactly like MES potentially are [40].
- diagnostics/bio-medical: diagnosis of health conditions via detection and classification of volatile organic compounds (VOCs) or biomarkers into one or combination of body fluids released as a result of a disease evolvement in the human body [40]. The monitoring of biomarkes is in general performed by means of the analysis of exhaled breath, skin/sweat, feces, urine, saliva, breast milk, blood, interstitial fluid, cerebrospinal fluid, tears, seminal fluid and intestinal gas [42]. Even virus detection in clinical solution is currently performed by means of FET-based sensors [125], [126]; thus leaving the door opened for a possible replacement with molecular electronics sensors (to be still verified if and in what cases it would be possible).
- food industry: sensors are used for the detection of toxic/polluting substances in food items, or to monitor the presence of pesticides. Moreover they are also exploited to monitor the shelf life or fungi contamination in vegetables and fruits or meat/fish spoilage [40]. Again, the brand and/or the place of origin of tobacco can be recognize by means of e-nose (electronic nose) based sensing systems [40].
- agricultural production: sensors are used for monitoring production processes, disease detection, identify insect infestations in their infant stage, and soil/water pollution (environment or pesticides etc...) [40].
- e-nose or e-tongue: all the fields of application mentioned up to now can be also the applications of e-noses (electronic noses) or e-tongues (electronic tongues) [40]. An e-nose (e-tongue) is mainly constituted by three parts: a sensor, a conditioning circuit or front-end circuit, and a post-processing usually real-time on-line unit [40]. The latter very often employs algorithms (implemented both in hardware and in software) based on Artificial Neural Network (ANN) computing paradigms [40]. The sensor can be in general a molecular electronic sensor.
- wearable/implantable sensors: wearable and implantable electronics is gaining popularity and interest in recent years [128], [129], [130], [131]. It may be a relevant next-future market for biosensors and for implementing generic tasks with biological data collection and transmission to smartphones for more complex post-processing [130], [131], [132]. In terms of sensed quantities, mainly human bio-signals are monitored (especially in sport applications), such as heart beat. Moreover real-time wearable physiology monitoring, e.g. temperature, pH, sweat and metabolytes measuring or multi-ion monitoring (potassium, sodium, etc...), is becoming relevant and important in various applications even already on market [132], [133]. In this field molecular electronic sensors may become valid substitutes to conventional FET-based sensors due to their high selectivity and sensitivity. Furthermore, due to their huge scalability they can play a central role for implantable systems, especially considering the recent advancements in biodegradable and biocompatible materials and molecules [134], that can potentially open the way to implantable biosensors that are nontoxic or spontaneously ejected or disintegrated/decomposed from the human body.
- single-molecule detection: this is an emerging field in sensing science, in which the presence of a single molecule is detected [38]. Usually the detection happens in real-time allowing for the detection also in the case in which the target molecule has a limited life-time of few milli- or micro-seconds (or even less), for example because it is an unstable product of a more complex chemical reaction [38], [135]. The process can be labeled or label-free [135], if a molecular electronic sensor is used to this end, it is label-free [38], [39]. The purposes of single-molecule detection can be e.g. the experimental proof of predicted chemical reactions (or chains of reactions) through the measurement of the type and life-time of

intermediate compounds, the better understanding of biological and bio-chemical dynamic processes, useful both in biology and medical researches, the early detection of analytes for disease diagnostics, etc... [38], [39], [135]. Other interesting applications are in real-time DNA sequencing. Often single electron transistors (SETs) that are used to this end are comprehensive of a nanogap (that can be natural, such as specific enzymes, or artificial like graphene nanogaps) to which an applied electric field allows for the scrolling of the DNA nucleotides through it at a constant velocity. The conduction through source and drain contacts changes depending on what nucleotide is present in that moment within the nanogap, allowing for DNA real-time sequencing [136], [87], [88], [137], [138]. In this field the single-molecule recognition is fundamental, and MES may become and alternative to conventional techniques or integrate and ameliorate them [38]. The single-molecule detection is sometimes referred as the ultimate limit in analytical chemistry, allowing for visualizing the dynamic processes of chemical/biological interactions with single molecules at single-event sensitivity [38], enabling the study of stochastic fluctuations under equilibrium conditions and the observation of time trajectories and reaction pathways of individual species in non-equilibrated systems, and the direct observation of novel microscopic quantum effects and fundamental discoveries of underlying molecular mechanisms in organic reactions and biological processes that are not accessible in ensemble experiments [38]. All this is becoming reality thanks to molecular electronic sensors.

5.2 Working principle

A molecular electronic gas sensor is a molecular wire or transistor in which the molecular channel conductivity is modulated such that it is possible to detect a given physical phenomenon or a specific chemical interaction aimed in identifying the presence, in the vicinity of the sensor, of a given chemical compound. In the rest of this work I will refer to gas sensors, nevertheless other sensors are in principle possible. In the following subsections I will considered them one by one, starting from the gas sensor one. In order to understand the principle behind a gas sensor, the concepts of physisorption and chemisorption are needed and for this reason they are preliminary introduced.

5.2.1 Chemical adsorption: physisorption and chemisorption

In surface chemistry the "adsorption" is the process by means the binding of a molecule to a surface is performed [99], [139]. More generally the adhesion of atomic, ionic or molecular species to a surface (both in gases and liquids) is referred as adsorption. Notice that it is spelled different from "absorption", that has the conventional meaning (e.g. the absorption of a photon/phonon by an electron), and in macroscopic chemistry is instead referred to dissolution of a chemical species (the absorbate) by a liquid or solid (the absorbent). In the case of adsorption one talks about the adsorbate, that is the chemical species, adsorbed on the surface of the adsorbent. As aforementioned the adsorption is a surface phenomenon, while absorption involves the whole volume of the material [139]. In general the term sorption includes both the adsorption and the absorption, while the term desorption is used to indicate their reverse processes. Thus in the case of chemical adsorption, for example, a given molecule can be adsorbed on a metal or graphene surface, indicating that it is somehow bond to it (see below). After that, if an enough energy is provided to such molecule (i.e. the to the adsorbate), then a desorption chemical process can take place and make again free the molecule. The adsorption of a molecule to a surface can occur in two fundamentally different ways: by physisorption or by chemisorption. The first process is governed by van der Waals interactions, while in the latter a chemical bond (covalent or ionic) occurs [99], [139].
Physisorption, also called physical adsorption [99], is a process in which the attraction of a molecule to a surface occurs [139]. This might be due to an electric dipole in the adsorbate or to

a van der Waals attraction. The interaction between the surface and the adsorbate is weak and typically is non-directional, i.e. little energy difference is expected between different binding sites and diffusion across the surface can occur and be rapid [139]. An important feature of physisorption is that no change to the chemical bonding structure occurs. In general there is also no (or small) overlap between the adsorbate and the adsorbent molecular orbitals. In a classical view (in the sense of Lewis theory for chemical bond) no chemical bond is created with physisorption between the adsorbate and the adsorbent. Since the van der Waals interactions are attractive, as the adsorbate moves closer to the surface, its electron wave-functions start to overlap with that of the surface atoms, leading, in turn, to a repulsive interaction (this is one of the so called steric effects [139]). Therefore in general (in system allowed to relax - e.g. because there is enough geometrical space) an equilibrium position or distance of the adsorbate from the surface is present. Such an equilibrium position is obviously the one that minimizes the total potential energy of the system.
Chemisorption, contrarily to physisorption, involves the creation of new chemical bindings between the adsorbate and the surface atoms. Consequently there can be a significant orbital interaction between the two, leading to high binding energies [139]. This means that a larger amount of energy should be provided to break the chemical bond and free the adsorbate w.r.t. the one needed in physisorption. The chemical bond between the adsorbate and adsorbent in chemisorption can be either ionic or covalent [99]. Notice also that chemisorption is a very specific process, i.e. the nature and the intensity of chemisorption can be greatly different depending on the considered chemical species. It is interesting to notice that corrosion is for example due to chemisorption. Another very interesting practical example of chemisorption is the formation of Self-Assembled Monolayers (SAMs), that are indeed formed by chemisorbing reactive reagents with metal surfaces. For example thiols (sulfur functional group S-H) adsorbing onto the surface of gold, allow the formation of strong chemical bonds with the release of H_2 (such that at the end a molecule can be connected by means of a sulfur atom, called the anchoring group) [139]. This is a widely spread technique at basis of fabrication of several molecular devices [29], [26]. Notice that the chemisorption is also dependent on the initial kinetic energy of the adsorbate [139]. The adsorbate needs to be trapped onto the surface by not possessing enough energy to leave the gas-surface potential well. If it elastically collides with the surface, then it would return to the bulk gas. If instead it loses enough momentum through an inelastic collision, then it "sticks" onto the surface, forming a precursor state bonded to the surface by weak forces, similar to physisorption. Then the particle diffuses on the surface, indeed the physisorption interaction is not enough to keep it in a given position, until it finds a deep chemisorption potential well (i.e. a position onto the surface that is an equilibrium position by minimizing the total potential energy). Then it can react with the surface or simply desorb after enough energy and time.
The main differences between physisorption and chemisorption are:

- Physisorption is a general phenomenon and occurs in any solid/fluid or solid/gas system. Chemisorption is characterized by chemical specificity.
- In physisorption the forces at the basis of the interactions are the van der Waals forces (see section 18.7.2) such as London forces, dipole-dipole attractions, dipole-induced attraction and eventually hydrogen bonding. In chemisorption the creation of chemical bonds occur.
- In chemisorption the electronic configuration is changed by the creation of a chemical bond. In physisorption only small perturbation occurs (for example obitals deformations due to steric repulsion of electron clouds).
- In practice, the categorisation of a particular adsorption as physisorption or chemisorption depends principally on the binding energy of the adsorbate to the substrate, with physisorption being far weaker on a per-atom basis than any type of connection involving a chemical bond. Typical binding energies for physisorption are indeed of the order of $10 \div 300$ meV and non-localized. Chemisorption usually forms bonding with energy of $1 \div 10$ eV and localized.

Finally notice that despite the fact that there is an equilibrium position/distance in both physisorption and chemisorption (for chemisorption this is the bond length), there is no an abrupt

transition from the two phenomena. Indeed it was experimentally demonstrated that it is possible a gradual transition between the two [140]. Namely the chemisorption can be obtained from physisorption if the distance between the adsorbate and the surface is forced to be reduced, with a gradual creation of the chemical bond (that it is not created abruptly). In general, if a system is allowed to relax, depending on the nature of the considered chemical species and also on the initial kinetic energy of the adsorbate (see above), it is possible to have physisorption or chemisorption, and they will occur at their equilibrium positions. This means that if a system is allowed to relax (there are no particular geometry constraints), the adsorbate will be find at a distance from the surface corresponding to the equilibrium distance of the phyisisorption/chemisorption process. Otherwise if constraints are enforced, all the intermediate cases can occur, leading to a gradual transition between physisorption and chemisorption and more generally to a gradual creation of a chemical bond [140]. For example, in [140], the enforced constraint is exactly the distance between a CO molecule, attached to the tip of an atomic force microscope, by means of which such a distance is controlled. Then the interaction between the CO molecule and the adsorbent surface is measured and monitored, proving what mentioned above. The fact of having a "smooth" transition and creation of a chemical bond is in accordance with the quantum mechanical treatment of chemical bonds [139], that indeed justify the creation of chemical bonds by overlap of electron orbitals (due to electron attraction from both/all the involved positive nuclei). This is in principle slightly different from the classical Lewis theory of chemical bond, even if it is, in practice, a more satisfying explanation of it.

Metal-molecule interactions and molecular devices

In [21] is provided a good review of the possible situations characterizing a molecule-metal interface. In molecule-metal systems (such as molecular wires with metal contacts) the quality of the interface is an important factor influencing the transport through the entire system. Accordingly with the previous discussion, the molecule can generally be physisorbed or chemisorbed onto the metal contact, nevertheless also intermediate cases are possible. In general the following cases can be identified (with increasing interaction strength) [21]:

- Physisorption with no charge transfer between the metal and the molecule. This can be the case of saturated molecule, that are thus non-interacting (like noble gases are). In this case there is essentially no interaction.
- Physisorption with possible charge transfer between the metal and the molecule. In this case the molecule is still weakly coupled with the metal surface and an integer electron charge can be transferred. This of course corresponds to high charging energy case.
- Physisorption with rearrangement of electron density, i.e. molecule and metal surface polarization. This is possible for neutral and polarizable molecules or for molecules with an intrinsic dipole moment.
- Weak chemisorption with a possible charge transfer. This is the case of molecules with free electron pairs chemisorbed on metal surfaces. A small amount of charge is exchanged between the molecule and the metal to reach Fermi level alignment and thermodynamic equilibrium.
- Strong chemisorption with strong covalent bonds. In this case there is an high level of hybridization of the orbitals and a charge transfer is possible to reach the equilibrium. The presence of surface dipoles can arise due to the formation of covalent bonds between atoms with (in general) different electronegativity (the molecule anchoring group and the metal atoms). This case typically is the one in which a specific anchoring group is exploited to guarantee the creation of such chemical bonds during the fabrication process (see above the SAMs example).

In general, in the physisorption case, a weak coupling between the molecule and the contacts occurs. This leads to localization of electron states, small or absent broadening, high charging energy and Coulomb blockade. Usually no charge transfer is observed in physisorption case

[21], and by biasing the structure an integer electron charge can be exchanged between the molecule and the metal. In the case of chemisorption a strong (covalent) bond is instead created between the molecule and the contact, and usually partial charge transfer occurs. This implies a strong interaction, with an high level of orbital delocalization and small charging energy with no Coulomb blockade. Notice that as mentioned above all intermediate cases are possible and there is no a net distinction between the two regimes, instead there is a gradual transition between the two [140]. In particular, to this purpose, notice that there will be two equilibrium distances between the contacts and the molecule, corresponding to the equilibrium distance for physisorption and for chemisorption respectively. Nevertheless if the system has constraints, it may happen that the geometry cannot relax, an intermediate situation occurs. For example, there can be a too short space in between the two electrodes to ensure physisorption relaxation to equilibrium, but at the same such distance between the contacts can be too large to ensure a creation of a strong chemical bond with wave-functions hybridization. In such a case, if the technological process is designed to have that the molecule has no enough kinetic energy to ensure being "caught" by the electrodes, then it will stay in between the two contacts, but with no optimal geometry, i.e. the most stable configuration will not be achieved. In that case the potential energy of the system will be higher than the equilibrium potential energy for physisorption, but lower than the same for chemisorption, and an external perturbation might easily (more easily than if equilibrium would reached) be enough to distort the geometry and cause variations or desorption.

5.2.2 Gas molecular electronic sensors

In gas sensing applications the molecular channel is exploited as sensor of specific chemical compounds present in air or solution, and its conductivity is uniquely modulated by the presence of the target chemical species in proximity of the molecular channel. Examples in literature, based on single molecule conductive channels, are: [85], [86], [90], [141], [142]. In general the target molecule (sometimes referred as the analyte especially in biological applications [106]), will interact via van der Waals intermolecular interactions (see section 18.7.2) with the molecular channel, thus changing the electron states and the transmission function (thus the conductance - see Landauer sec. 19.3 and **??**) of the channel. As a final effect the electrical current will change depending on the presence or not of the analyte. By detecting such a specific variation of current, it is possible to detect the presence of the target analyte and likely also its concentration. Accordingly with the treatment of section 5.2.1, the target molecule can be in general physisorbed or chemisorbed by the molecular channel. Even if, precisely, this would be an abuse of notation, since the two terms "physisorbed" and "chemisorbed" are strictly referred to an adsorbent that is a surface, while a single molecule is not. In general, anyway, if the channel molecule is very large w.r.t. the adsorbate then the process is quite similar. Even if it is not (i.e. the sensor molecule and the adsorbate one have similar dimensions), one can use the terms "physisorption" and "chemisorption" to highlight the nature of the chemical interaction [90], having in mind that they are not referred to a surface phenomenon. In addition, such (not precise) lexicon descend from a class of gas sensors that is very demonstrated in literature both by means of theoretical studies/simulations and from prototypes experimental data: namely the 2D FET-based gas sensors. In such sensors the working principle is analogous to the one of gas-MES, that indeed can be seen as the "novel generation", at moment at embryonic level, of the previous ones. For example a GNR-gas-sensor is a sensor in which a 2D sheet of graphene (i.e. the graphene nano-ribbon) is used adsconducting channel. Then the target molecule can be physisorbed or chemisorbed onto the GNR resulting in a modulated conductance. This class of devices, the 2D-FET-gas sensors, are well demonstrated in literature [101], [102], [102], [110], [109], [112], [115], [113], [114], [104], [41], (1D nano-wire example with the same principle: [116]).
In [85], a C_{60} fullerene is used as molecular channel for a complete molecular transistor, and by means of simulations with EHT + NEGF methods, both at equilibrium and in non-equilibrium, the authors support its application as ammonia (NH_3) gas sensor, providing also good results

in terms of selectivity and response to different ammonia concentrations.
In [86], the authors prove the potential application of a modified/doped C_{70} fullerene as gas sensor for sulfur dioxide (SO_2) and nitrogen dioxide (NO_2), that are very toxic for humans [117]. In particular the study is focused only at equilibrium with no electrodes (thus the entire device is not studied), and the adsorption properties of SO_2 and NO_2 by the metal ion (Cr^{2+} and Co^{2+})-prophyrin-induced C_{70} is proved to be effective, thus making the latter promising candidate for gas nanosensors.
Similarly to the previous paper, in [90], the author prove the effective adsorption properties of C_{20} fullerene of carbon monoxide (CO), carbon dioxide (CO_2) and methane (CH_4). Again the study is carried out at equilibrium only and no complete device is proposed.
The two papers [141], [142] present other two gas sensors with analogous principles of the above mentioned ones.

5.2.3 Single molecule event detection with molecular electronic sensors

Single molecule detection is a novel frontier of sensing science, in which the presence of a single molecule can be dynamically detected, making also possible the detection of intermediate (unstable) products of a complex chemical or bio-chemical reaction [38]. There are many ways of achieving such a task [38], [39], among which the novel molecular electronic sensor approach seems to be promising for a label-free real-time detection. For example in [34] a molecular wire is exploited for this task. The working principle is essentially identical to one presented previously for the gas sensors, with the exception that here the channel molecule is engineered to allow high specificity and thus high selectivity such that the sensor responds exclusively to the target analyte. In particular, for the first time, it was experimentally proved that a given theoretically predicted chemical reaction chain occurred, indeed the current flowing in the molecular wire resulted dynamically modulated by the presence of an unstable intermediate chemical species with a short life-time (of the order of ms) that results to be an intermediate product of the chemical reaction. This is in my knowledge the first prototype of its type, fabricated and tested. The presence of the target molecule is detected by means of a specific chemical interaction with the molecular channel, causing a modulation of the molecular wire current. The theoretical modeling is based on a DFT + NEGF approach: in particular it was used DFT with the hybrid functional B3LYP (*Gaussian* software) and equilibrium Green's function (*Quantum-Wise ATK* software) for evaluating equilibrium transmission spectra. The theoretical calculations matched the experimental data, thus being a literature prove of the effective modeling techniques introduced so far in this work.

5.2.4 Light molecular electronic sensors

In principle light sensors are possible, even if at present in my knowledge there are no practical or complete device studies in literature. The basic principle is very easy: an incoming photon can be absorbed by the molecule thus exciting an electron to an higher energy level and modulating the current flowing through it. The principle is thus analogous to the working principle of conventional opto-electronic devices, also organic ones. The correct way for modeling the photon-channel interaction is by means of incoherent transport, with the addition of a suitable self-energy within the NEGF framework in order to account the photon-electron interaction, as already explained in section **??**. The potential advantage of molecular device used as light sensors is that the molecular channel and the device can be optimized such that only specific transitions, involving only well defined energy levels and thus photon energy, are allowed and therefore detected. This allows to obtain great freedom at design level on the energy or wavelength of absorbed photons. Molecules can also emit photons, in such a case they are no more sensors or detectors but sources instead (analogously to Organic Light Emitting Diodes -OLED-).

5.2.5 Humidity and temperature molecular electronic sensors

Humidity and temperature sensors can be also in principle realized. An example of trivial humidity sensor is provided in [43]. In this case a SET is used to detect the presence of H_2O in air. In particular an empty device, with ideal (PEC) contacts is supposed to have like a window in which air can circulating. If an H_2O molecule happens to be in between the source and the drain contacts, a variation of the (noise) current occurs, since it has in principle specific conduction properties, and the detection is possible. In principle it is also possible to have a molecular device (such as a C_{60}-based wire) and to study if (and in which way) the conductance changes if water molecule is present nearby the channel (in this case the principle becomes exactly the one presented previously for gas sensors).
In the case of temperature sensor it is obvious that it somehow affect the transport, since it influence the Fermi-Dirac's functions and also the number and the energy of phonons; thus it can be in principle be detected a temperature variation by detecting a current variation.

5.2.6 DNA sequencing with molecular SETs

A thrilling application of SETs and molecular electronic sensors is the possibility of performing real-time DNA sequencing [87], [88], [137], [138], [136]. The working principle is the following. The DNA molecular chain is forced to flow through a nanopore to which a potential difference is applied. The nanopore can be an artificial graphene nanopore (e.g. [137], [138]) or a suitable enzyme and it works as contact (an all-around contact that surrounds the DNA nucleotides). By means of the applied electric field it is possible to force the flow of the DNA through the nanopore, and at the same time by measuring the electrical current it is possible to detect in real-time what nucleotide is in each moment present in the nanopore. Indeed DNA nucleotides are molecules, and by changing the molecule of the channel of the SET, it changes the electrical current flowing in it, thus making possible the recognition. The simulations techniques [87], [88] are generally in agreement with experimental data and prototypes were demonstrated to be functioning and effective as a fast and reliable method for DNA sequencing [137], [136].

5.3 Methodological approach: gas sensors example

The purpose of this section is to summarize the logical steps that are suggested to discuss the structure, design and performances of a C_{60} gas sensor. In this optics, it is a brief introduction to set the very practical part of the work [2], in order to avoid getting lost in the following.
In order to keep the treatment as much systematic as possible the key points are:

- First of all a molecule that should be used as channel and the target molecule should be selected. In general there is no unique way of doing it, and the the channel can also be engineered to posses some specific physical-chemical properties such that the interaction with the target is enhanced and maximized, both from the standpoint of the sensibility/detectability and from the standpoint of selectivity/specificity. In the following I will consider a C_{60} molecule as channel, and I will initially investigate its performances as ammonia (NH_3) gas sensor, by following the approach of [85]. As mentioned in section 5.2.2, the choice of the C_{60} has multiple reasons: it is a well known molecule stable at room temperature [143], and since made of carbon only it is pretty interacting [143], indeed some valence electrons are not involved in chemical bonds and they are highly interacting, thus promising good interaction properties with target analytes [143]. Moreover its sensing properties were already investigated in recent years, especially in standard chemo-sensor architectures as summarized in the review paper [89], besides that in [85]. By following what did in the same paper I will firstly investigate such device as gas sensor for ammonia.
- Once the channel molecule is chosen it should be chosen the device structure. In general mainly two architectures are possible: a molecular wire (two-terminal device) and a molecular transistor (with one or more gates). Since the fabrication process is more complex and thus expensive for molecular transistors than for the wires, I have chosen to start with studying and testing a molecular wire. Then, once the wire is characterized it is possible to add a gate in order to verify if a substantial performance improvement can be obtained. If not so, it has no sense using a transistor.
- At this point the contact material and eventually the presence of anchoring groups (i.e. functional groups aimed in anchoring the molecule to the contacts) should be selected. This is a crucial point since the contact-molecule interface strongly determines the quality and the regime of transport. In general in real cases it should be evaluated considering also the technological process and the production necessities. I have chosen to employ no anchoring group and metal electrodes of gold (Au), ideally FCC (111), i.e. Face-Centered-Cube crystalline cell, with orientation (111) - please refer to the widely used Miller notation, see e.g. [144], [145]. The reason behind this choice is that the gold contact technology is already widely used and known for molecular junctions [29], [26]; and at the same time it is demonstrated in literature that carbon is capable of forming strong covalent bonds with gold, thus making an anchoring group redundant.
- The main parameter to be fixed in a molecular wire is the gap length between the two electrodes, in which the molecular channel will be host. In other words the distance between the contacts and the C_{60} molecule should be fixed. In order to do that, it is advantageous choosing a distance that maximizes the sensitivity to the target, if there exists one. In this optics I have studied the sensitivity dependence in function of the contacts distance, resulting in higher sensitivity when the distance is smaller. The chosen distance was thus the one corresponding to the covalent gold-carbon bond length. This corresponds to the equilibrium distance in case of chemisorption, indeed notice that the gold electrode is essentially a gold surface with large dimensions w.r.t. C_{60} fullerene one, thus leading effectively to chemisorption. In order to choose and fix the geometry it is also reasonable to perform geometry relaxation simulations, to verify if the chosen geometry is actually a stable configuration, since it should be robust in time (and resist to energy/phonon absorption in order to avoid dynamical changes in the device geometry that would be highly undesired).

- In parallel to the geometry choice it should be also performed another kind of important analysis. Indeed it is convenient to preliminary check if the molecular channel is somehow sensitive to the target. This can be accomplished at equilibrium by verifying a change in the DOS or in the transmission spectrum. Moreover this analysis can be performed on the isolated C_{60}, i.e. without the contacts, like for example done in [86], [90]. Nevertheless the disadvantage of doing so, especially in single molecules channels, is that the influence of contacts on the final DOS and transmission features can be relevant and can considerably change the attended results. For this reason I have chosen to perform a preliminary study on the equilibrium DOS and transmission spectrum of the entire device, i.e. with contacts. Such a study was performed with DFT calculations (considering the Grimme-D2 and Grimme-D3 correction for van der Waals interactions - see section 18.7.3). A significant variation, in the case in which the target is present nearby the channel w.r.t. to the case in which it does not, either in the DOS or in the equilibrium transmission spectrum, is enough to verify the potentiality as sensor for that target.
- In general one may have time/resources requirements/constraints, such that it is not always possible to use the method known to be the most accurate for representing the physics/-chemistry of interest. For example during my work I could not perform non-equilibrium simulations with DFT methods, because the server in which all computations were performed did not support it (the RAM was not enough to sustain such calculations, even with simulation settings to minimize memory usage). In that cases one may use other methods. For example I decided to use the EHT one, conscious of its limits. Nevertheless as mentioned several times EHT method should be validated. For this reason I have tried to reasonably validate it, for the non-equilibrium case, starting from the equilibrium more accurate DFT method simulations. The details are provided in the dedicated sections, nevertheless here the point is on the general methodology: when a method, generally expected to be not much accurate for the specific application, is used, then the obtained results should be somehow validated.
- At this point the device geometry should be fixed and the device is expected to be sensitive to some target. A crucial performance verification should be performed: the selectivity of the sensor. Indeed the sensor in order to be able to detect the presence of a given target should respond in a unique way to the presence of such target, like it does not happen in response to other substances that can be present in the usage environment, such as air components (oxygen, nitrogen, humidity/water, etc...). This means that many simulations should be performed in order to verify such sensor selectivity. Moreover the sensor should provide a response to the various orientations of the target (especially if polar), and in this regard also a study of the equilibrium physisorption/chemisorption of the target onto the channel molecule should be performed.
- Once all the aforementioned tests are performed then a suitable bias point should be selected. The optimal one should maximize the sensitivity and selectivity, if those are the most important feature for a given application. Otherwise it may be chosen to maximize linearity and so no... Another criterion of choice can be the bias point for which there is maximum correspondence between the EHT and the DFT simulations (of course only if a method not at all suitable is used, as explained above).
- Sensor performances should also be checked, in order to highlight possible critical issues. To this purpose the response to different target concentrations should be studied, and for example the bias point with maximum linear response can be chosen if linearity is important in the sensor specifications. Moreover also the maximum distance above which no significant sensor response is obtained is a useful data for proceeding in the design. Another useful check can be on response time of the sensor and eventually dissipated power by it.
- Depending on the specifications several properties may be checked and verified. If not satisfied a study on the reason why it is so is always helpful in understanding if and how the device should be modified to satisfy the requirements. The actual device structure of

MES is quite easy if compared to conventional nowadays electronic devices. Thus in order to keep the structure simple (since already in this way the technology is a big issue [29]), few parameters can be changed, unless the channel molecule is modified or changed. The latter solution is very effective and opens the way to an incredible variety of heterogeneous solutions. Nevertheless another solution can be to modify somehow the device; this could require a long time before converging to the desired solution.

– In order to improve performances the introduction of a gate electrode can be considered. This is a trivial solution that nevertheless can be potentially very effective in ameliorating the overall performances of the sensor.

By following more or less the above reported points it is possible to predict several features of the sensor and have thus a complete model, more or less refined, for the sensor, providing an idea on its behavior and response in different more or less realistic situations. In a real design process, depending also on the specifications and requirements, it is possible to have jumps among the above introduced points, and also it is possible to have a sort of feedback among them in order to optimize the device with the gain in knowledge of its performances, etc...

5.4 C60 gas sensor

..... coming in the next version of the LN....

5.5 PCP based chrome sensor

..... coming in the next version of the LN....

5.6 Bio sensor for tyamine

..... coming in the next version of the LN....

CHAPTER 6

Molecular Transistors fabrication

The process of fabricating molecular transistors involves several intricate steps to create a nanoscale structure capable of hosting and connecting individual molecules. Here, we outline the key stages:

1. **Structure Preparation:** The initial step is to prepare the groundwork for the creation of a nanogap, which serves as the fundamental building block for molecular elements.
2. **Nanogap Creation:** Creating the nanogap is a critical phase, and various techniques are employed to achieve this, including:
   - **Direct Creation:** This approach typically utilizes photolithography techniques, such as Focused Ion Beam (FIB), to carve out nanoscale gaps.
   - **Electromigration (EM):** EM involves inducing controlled movement of atoms within a metal structure to form the nanogap.
   - **Mechanical Stress:** Applying mechanical stress to the structure can also generate nanogaps, although the specific technique used depends on the desired outcome.
3. **Molecule Attachment:** After the nanogap is established, the next critical step is to facilitate the attachment of molecules within this gap.

The size of the nanogap is of paramount importance in molecular transistor fabrication. To accommodate molecules, these nanogaps must be exceptionally small, typically on the order of a few nanometers. The exact gap size depends on factors such as the size of the molecule itself or the length of a molecular chain. Our research on Molecular Transistors has demonstrated that the behavior of these transistors is intricately linked to the length of the molecular chain, emphasizing the need for precise fabrication.
It's noteworthy that while nanogaps of around 100 nanometers find application in various nanotechnology fields, they prove too large for molecular applications. Additionally, real-world nanogaps often exhibit non-ideal lateral characteristics, impacting the proper attachment of molecules. Consequently, during the fabrication process, variations and non-idealities must be considered as factors influencing the performance of molecular transistors.
Another critical consideration is whether the goal is to create a single nanogap and an associated molecular transistor or to establish multiple nanogaps within a single circuit. The latter scenario enables several molecular transistors to function concurrently, contributing to the implementation of specific functions at the system level. These two scenarios necessitate different fabrication approaches. For laboratory-level investigations, creating a single nanogap and molecular transistor suffices. However, for practical applications, the ability to fabricate multiple elements in parallel becomes essential.
The choice of fabrication technique significantly influences resolution and scalability. For instance, the direct creation approach using FIB offers exceptional resolution but is best suited for creating one nanogap at a time. Conversely, alternative techniques allow for the simultaneous creation of multiple devices but with reduced resolution. Electromigration is highly effective

for generating a single nanogap, although extending it to multiple nanogaps can be challenging. Mechanical stress-based methods offer versatility and can be adapted to create either a single nanogap or multiple nanogaps, often by combining mechanical stress with EM.

Fabrication Method	Minimum Gap Size
Direct FIB	1 nm
Physical fabrication methods	Sub-nanometer
EBL	Sub-nanometer
Metal deposition	3 nm
Nanosphere lithography	2 nm
2D material	1 nm
Electromigration	1 nm
Mechanical	3 nm
Chemical fabrication methods	Sub-nanometer
Electrochemical	1 nm

Table 6.1: Minimum Gap Sizes for Fabrication Methods

Key Features	Applications
Single-gap fabrication, high precision	Research and development
Large-area fabrication, high precision	high-end manufacturing
Large-area fabrication, high resolution	semiconductor devices
Large-area fabrication, controllable metal deposition	Electronics, sensors, actuators
Large-area fabrication, low cost	Optics, sensors, catalysts
Large-area fabrication, unique properties	emerging applications
Large-area fabrication, self-assembly	nanowires, nanostructures
Large-area fabrication, high throughput	MEMS devices, sensors, actuators
Large-area fabrication, high precision	nanofluidic devices, biosensors
Large-area fabrication, low cost	Batteries, fuel cells, electrochemical devices

Table 6.2: Key Features and Applications of Fabrication Methods

6.1 Fabrication of Molecular Nanowires

The fabrication of molecular nanowires and molecular transistors represents a dynamic field of ongoing research. It involves the development of structures capable of detecting one or more molecules at the nanoscale and connecting these molecules to an external system at the microscale. This research area shares techniques with several related applications, including plasmonic antennas, subwavelength waveguides, nano-optics, and medical treatments.
A central concept in this field is the "nanogap," a critical element for hosting molecules and enabling molecular-level interactions. The nanogap, illustrated in Figure 6.1, is a nanoscale void between two structures. What sets it apart is that it is not typically created through conventional lithographic processes used in constructing molecular vessels. Instead, nanogaps are often formed as a byproduct of a process, such as the rupture or milling of a metal wire. This process can be seen in the diagram, progressing from the left scheme to the right one in Figure 6.1.

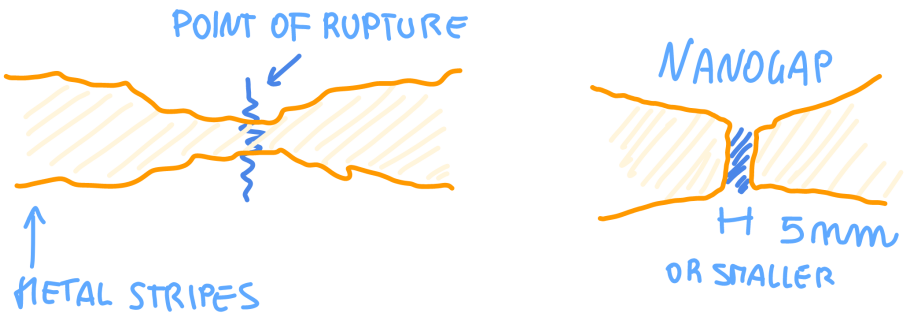


Figure 6.1: A sketch illustrating the general concept of a Nanogap.

6.2 Direct nanogapcreation

In the field of nanotechnology, the fabrication of nanoscale structures is a critical aspect. Several techniques are employed for this purpose, each with its own advantages and limitations. One of the primary techniques utilized is lithography, specifically "high-end lithography," which is extensively employed in the creation of integrated circuits. High-end lithography offers excellent resolution capabilities, achieving as low as 20 nm resolution. However, as the demand for even finer features arises, we are moving towards resolutions of 10 to 7 nm. These advancements come at the cost of increasingly expensive processes, but they are still considered reliable, especially for technologies where achieving nanoscale dimensions, ranging from 3 to 5 nm, is essential.
Another valuable technique in this realm is "electron beam lithography" (EBL). EBL can achieve resolutions of 10 nm or even slightly less, and it is particularly well-suited for parallel production. While EBL is used for various phases, such as connecting electrodes and preparing the structure, it is not typically employed for the final creation of nanogaps due to certain limitations.
A more precise approach is the "Focus ion beam" (FIB)6.2, which utilizes an array of ions, typically Ga+ ions, directed onto the substrate. FIB offers exceptional precision but necessitates careful tuning of the ion beam's strength.

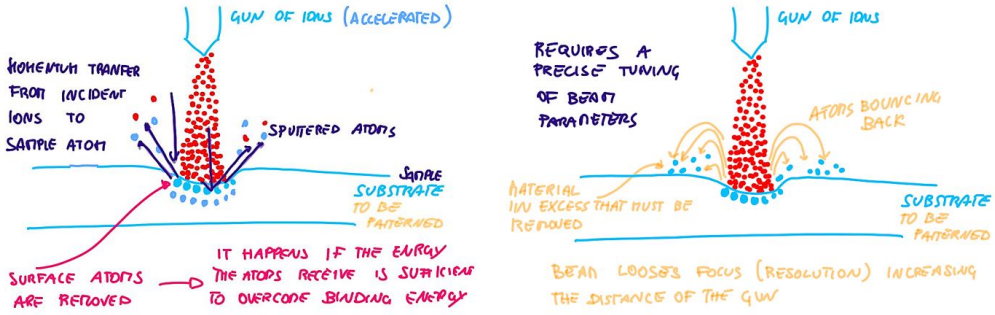


Figure 6.2: Fabrication of a molecular transistor using focused ion beam (FIB)

However, one challenge in FIB is that the ions do not travel in a perfectly straight line toward the substrate. As they move farther from the ion gun, the ion beam widens, resulting in the scattering of atoms from the substrate material. These scattered atoms bounce around on the material's surface. Despite this drawback, FIB is advantageous because it allows for the creation of precisely defined holes or gaps in the substrate material. Nevertheless, it generates excess material around the gap, which must be taken into consideration. Interestingly, this excess material can have advantages in certain applications.
For the creation of nanogaps, the initial structure typically involves a silicon (Si) substrate,

followed by a layer of silicon dioxide (SiO2), around 300 nm thick, obtained through oxidation. Next, a thin layer of titanium (Ti), about 3 nm thick, serves as an adhesion layer. Finally, the actual metal, usually gold (Au), is deposited for gap creation and electrode formation. This metal layer typically ranges from 50 to 60 nm in thickness and functions as the nanowire where the gap will be created, molecules will be placed, and electrodes will be connected to the Source and the Drain.
In the top view of the structure, it is evident that the nanowire is typically patterned using EBL, resulting in a narrower section in the middle where the actual nanowire and gap creation will occur, with a width of 200 to 250 nm. This varying width is necessary to accommodate connections with the external components, known as "pads."

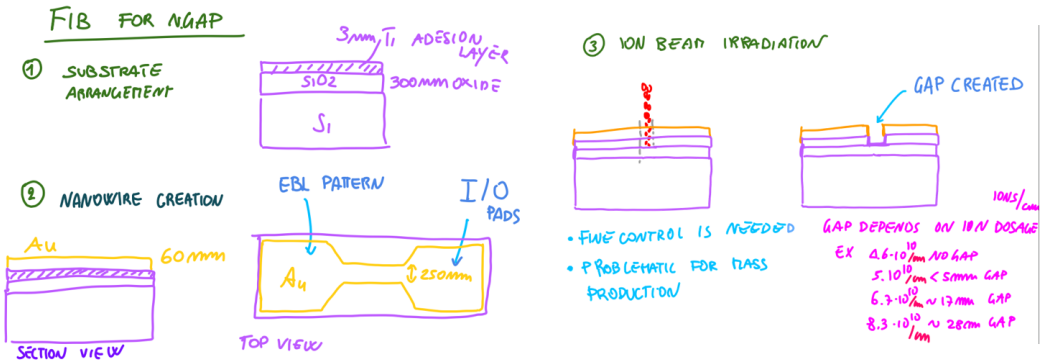


Figure 6.3: Cross-section of a nanowire with a nanogap created using FIB

Moving forward, the process involves ion beam generation. In the section view, ions are directed onto the substrate to create the desired gap. Over time, the highlighted zone should theoretically contain nothing, resulting in a structure on the right where the gap is successfully created.
The quality of nanogap creation in FIB depends on the ion dosage, typically measured as ions per square centimeter (ions/cm). For example, using a typical FIB instrument with a dosage of 4.6*1010 ions/cm, In the absence of a gap, no gap would be created. With a dosage of 5×10^{10} ions/cm, a gap around 5 nm or smaller would form. However, increasing the dosage to 6.7×10^{10} ions/cm would result in a gap with a width of 17 nm. This example illustrates the significant impact of ion dosage on gap width, emphasizing the need for precise control6.3. Achieving the desired gap width depends on various factors, including the thickness of the gold layer, temperature, and the limitation of creating one gap at a time. While FIB is well-suited for laboratory testing, it may not be suitable for mass production.
Following FIB, a cleaning procedure is typically required due to the scattering of material atoms. Interestingly, in the context of creating nanogaps, this cleaning process can be advantageous. The scattering of atoms during the uncleaning process can further narrow the nanogap compared to the initially intended width. For example, if one aimed for a 5 nm gap, the cleaning process might result in a 2.5 nm gap. While this effect can be desirable, achieving precise control over it is a key consideration in the process6.4.

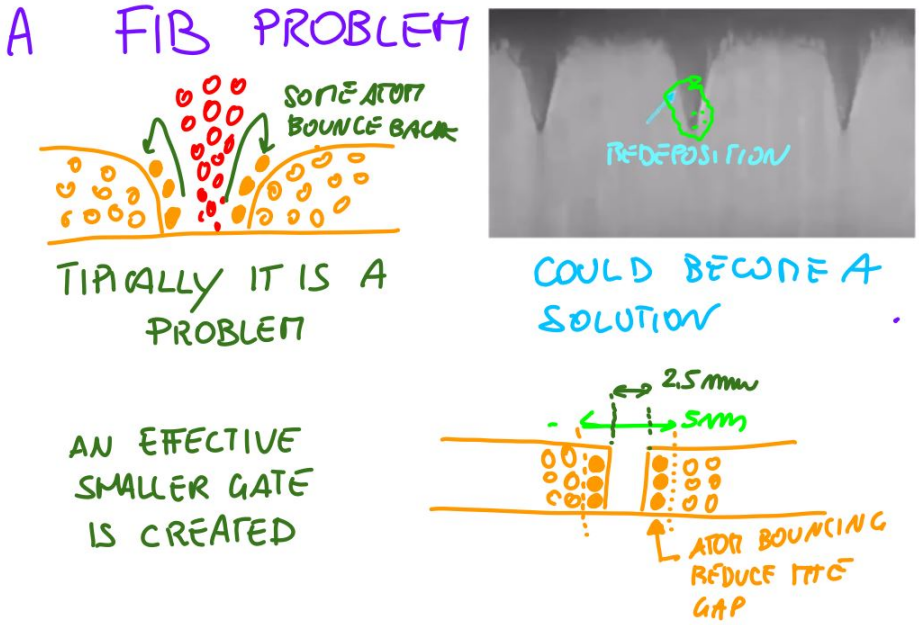


Figure 6.4: Problem of backscattering in FIB milling of nanogaps or De-position

Re-positioning poses a notable challenge in the fabrication process. However, as previously mentioned, it can offer distinct advantages. One noteworthy application of re-positioning is the ability to finely adjust the width of the gap. This adjustment is achieved by strategically utilizing the molecules that have been deposited and are firmly attached to the inner wall of the gap itself. This precise control over gap width can be crucial in tailoring the nanoscale structure to meet specific requirements or to enhance its performance characteristics.

6.3 Electromigration

6.3.1 General Concepts on Electromigration (EM)

Electromigration, often abbreviated as EM, is a critical phenomenon in integrated circuits that can lead to the degradation and failure of electronic components. This section explores the fundamental concepts of electromigration and its implications in semiconductor devices.

Electromigration Overview

Electromigration refers to the process in which the movement of electrons in a conductor results in the transport of metal ions within the conductor material. This movement can lead to gradual physical changes, such as void formation and hillock growth, which can ultimately affect the functionality of the conductor. The electric wind, generated by the motion of electrons, imparts momentum to atoms within a metal. When this process occurs continuously, it can lead to the displacement of metal atoms from their original positions. As atoms move, they create vacancies, leaving empty spaces within the lattice structure of the metal. This movement of atoms within the lattice, driven by the transfer of momentum from the flowing electrons, can eventually result in the creation of voids. These voids or vacancies can lead to a weakening of the metal structure and, in the case of interconnects, can potentially result in the formation of open circuits. The process can be likened to a cascade effect, where the movement of individual atoms triggers a chain reaction, eventually leading to a breakdown in the integrity of the material.

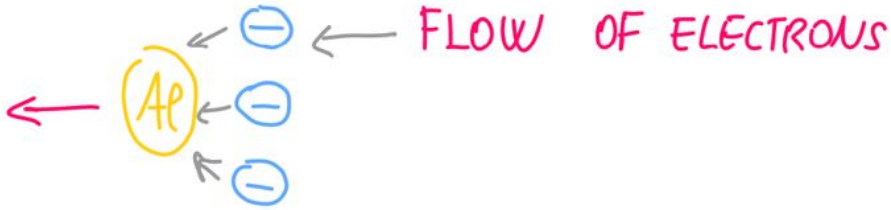


Figure 6.5: The movement of the Metal atom after the while

Creation of Voids and Hillocks

Under the influence of high current density, metal atoms within a conductor can experience electromigration. Electrons flowing through the wire transfer momentum to metal atoms. Two significant outcomes of electromigration are:

- **Voids Formation:** Electromigration can create voids within the conductor material. As metal atoms migrate from their lattice positions, voids or gaps are formed, disrupting the continuity of the conductor and potentially leading to open circuits6.6.
- **Hillocks Formation:** In addition to voids, electromigration can lead to the formation of hillocks. As metal atoms migrate and accumulate, they create raised regions or hillocks on the conductor's surface. These hillocks can cause short circuits when they come into contact with neighbouring conductive paths6.6.

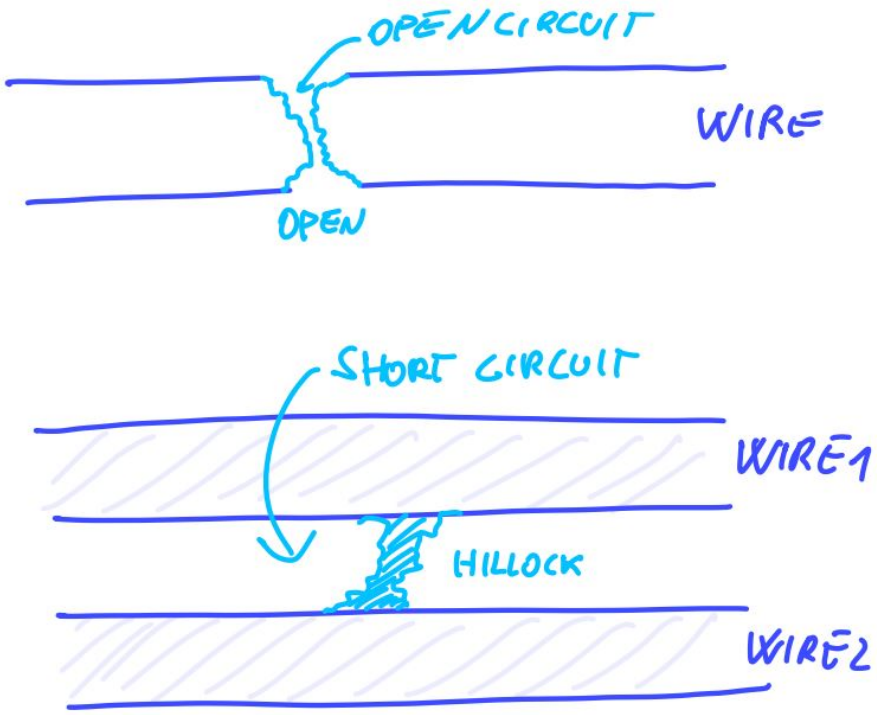


Figure 6.6: Voids and Hillocks

In the context of electromigration, feedback control systems are often employed to monitor and mitigate its effects in integrated circuits. Feedback control involves actively adjusting operational

parameters to minimize electromigration-related issues. Various techniques and algorithms are used to achieve this goal.

Formulas and Equations

While feedback control involves complex algorithms and models, some fundamental equations and relationships help in understanding and addressing electromigration:

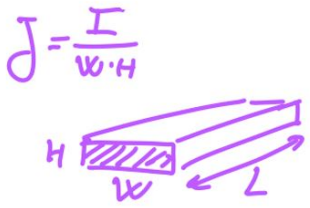


Figure 6.7: Current density

1. **Mean Time to Failure ($M.T.F.$):** $M.T.F.$ is an empirical measure representing the average time expected until electromigration-induced failure occurs. It is proportional to the inverse of the square of the current density (J) and depends exponentially on several parameters. The formula for $M.T.F.$ is:

$$M.T.F. \propto \frac{1}{J^2} \cdot e^{\frac{Q}{k \cdot Tm}} \tag{1}$$

   Where:
   - Q is the activation energy for metal atom movement.
   - k is the Boltzmann constant.
   - Tm is the actual temperature of the metal.
   - J is the current density.

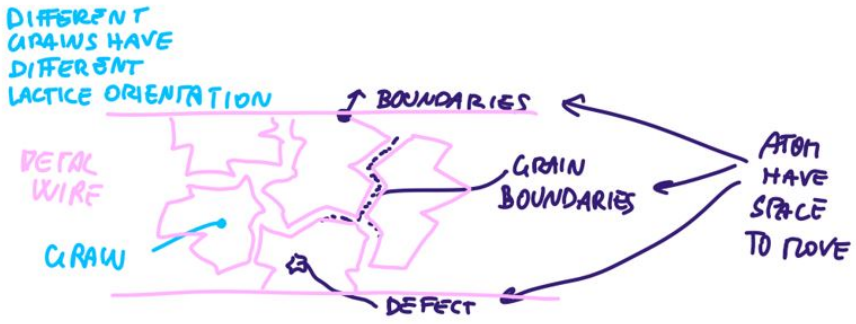


Figure 6.8: Different grains have lattice orientations.

2. **Joule Effect:** The Joule effect relates the resistivity (ρ) of the metal to temperature (T_m) and current (I). The resistivity can be approximated as:

$$\rho = \rho_0 \cdot [1 + \alpha \cdot (T_m - T_0)] \tag{2}$$

   Where:

- ρ_0 is the resistivity at room temperature.
- α is a technology-specific parameter.
- T_0 is the reference temperature.
- T_m is the actual temperature of the metal.

The actual resistivity of the metal (ρ_m) is given by:

$$\rho_m = \rho_0 \left(1 + \alpha(T_m - T_0)\right)$$

Where:

- ρ_m is the actual resistivity of the metal.
- ρ_0 is the reference resistivity at room temperature.
- α is a technology-specific parameter.
- T_m is the actual temperature of the metal.
- T_0 is the reference temperature of the metal.

The actual temperature (T_m) of the metal is determined by the sum of the environmental temperature (T_{ENV}) and the power dissipated (P_m) in the wire, multiplied by the thermal resistance (θ):

$$T_m = T_{\text{ENV}} + P_m \cdot \theta$$

Where:

- T_m is the actual temperature of the metal.
- T_{ENV} is the environmental temperature.
- P_m is the power dissipated in the wire.
- θ is the thermal resistance of the wire.

The power dissipated (P_m) is given by:

$$P_m = R_m \cdot I_{\text{RMS}}^2$$

Where:

- P_m is the power dissipated in the wire.
- R_m is the actual resistance of the metal.
- I_{RMS} is the root mean square of the current flowing in the wire.

In this equation, R_m depends on the actual temperature (T_m) of the metal. As the temperature increases, the resistance (R_m) also increases. This relationship is crucial to understanding the Joule effect in electromigration.

So, it can be summarized as:

$$\rho_m = \rho_0 \left(1 + \alpha(T_m - T_0)\right) \tag{5}$$
$$T_m = T_{\text{ENV}} + P_m \cdot \theta \tag{6}$$
$$P_m = R_m \cdot I_{\text{RMS}}^2 \tag{7}$$
$$R_m \leftarrow T_m \leftarrow P_m \leftarrow R_m \tag{8}$$

The correct mathematical expression for I_{rms} in terms of the integral of the square of the current ($I^2(t)$) over time (t) is:

$$I_{\text{rms}} = \sqrt{\frac{1}{t} \int_0^t I^2(t)\, dt} \tag{9}$$

This formula calculates the RMS value of a time-varying current signal $I(t)$ over a specified time interval from 0 to t.

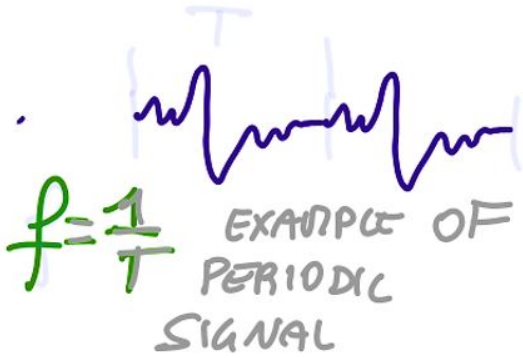


Figure 6.9: using the frequency instead of the time

3. **Thermal Resistance (θ):** In a first-order model, it does not include the second parameter in the parentheses. Thermal resistance is influenced by the thickness of the oxide layer beneath the conductor. This layer's role becomes crucial when the heat sink is positioned beneath the substrate, and the conductor line has specific dimensions (W, L, and thickness) and is located at a certain distance from the heat sink. The greater this distance, the higher the thermal resistance becomes. Thermal resistance essentially quantifies the difficulty in dissipating power effectively.

   Thermal resistance (θ) is a critical parameter that determines the heat dissipation capabilities of a conductor. It depends on factors such as wire dimensions (H, W, L), substrate proximity (tox), and material properties (kox). The formula for thermal resistance is:

$$\theta = \frac{L}{kox \cdot W \cdot tox} \tag{10}$$

   Where:

   - L is the length of the conductor.
   - kox is the thermal conductivity of the oxide material.
   - W is the width of the conductor.
   - tox is the thickness of the oxide layer.

6.3.2 Nano-Electronics and Electromigration

Micro-Electronics Perspective: From a micro-electronics perspective, we recognize the importance of managing various parameters, including factors like thermal conductivity to control temperature. This management is essential for reducing the electromigration effect.

Transition to Nano-Electronics: Now, in the context of nano-electronics, our goal is to induce and control electromigration to create the desired gap. However, it's crucial to note that temperature control is challenging due to its inherent dependency.

Controlling Atom Movement: Atoms are in motion, and there is a possibility that they may move into the gap we are designing. We must consider this factor because the created gap can potentially trigger the thermal runaway effect in the metal.

Creating a Nano Gap Without Runaway Effect: Our challenge is to create a nano gap while harnessing the benefits of electromigration's exponential behavior without experiencing the runaway effect.

Our Approach:

1. Organizing the Wire: Initially, we organize the wire to minimize atom movement by creating a narrow point conducive to controlled electromigration.
2. Controlled Electron Migration: We aim to promote controlled electron migration and determine where it will occur precisely.
3. Gradual Current Application: We apply the current incrementally, employing a sequence of on-off cycles. When we approach the desired break point, we pause to reduce temperature and prevent the runaway effect.
4. Controlled Consumption: We carefully increase the current as we stop the cycles. This gradual process eventually leads to the creation of the desired break point.

This approach is reversible and allows us to create the specific nano gap we need while effectively managing temperature and avoiding the runaway effect.

6.4 Electromigration in case of nanogap creation

In this case, in order to create a good nanogap, we first need a very narrow structure with precise size control, typically in the range of 1 to 5 nm. This is essential for making it suitable for molecular electronics. Another requirement is the presence of sharp edges. In the previous image, we can observe that the metal atoms are highly ordered within the lattice at the edges. This order is crucial because we intend to place a molecule inside the gap and establish a connection with the sides. If the edges are not sharp, as shown in the figure on the right, we won't have a clear understanding of what will happen to the molecule, whether it will anchor or not, and so forth. Now, we have to use Electromigration not as a problem but as a tool to exploit. We can induce Electromigration to create the nano-gap. In fact, we have an electron flow depending on temperature, creating some voids. However, we have inadvertently created a loop, where resistance depends on power, which in turn depends on resistance. This loop is not easily controlled, leading to atoms being moved and placed in a less organized manner. Nanogaps are essential not only for molecules but also for antennas. Therefore, the issue of this loop must be taken into account since the breakpoints can lead to thermal runaway, rendering the system unusable. Thermal runaway describes a process accelerated by increased temperature, which, in turn, intensifies current flow and power dissipation, releasing energy that further raises the temperature. It's a form of uncontrolled positive feedback.
More specifically, we desire a very narrow gap (a), but with sharp edges (b).

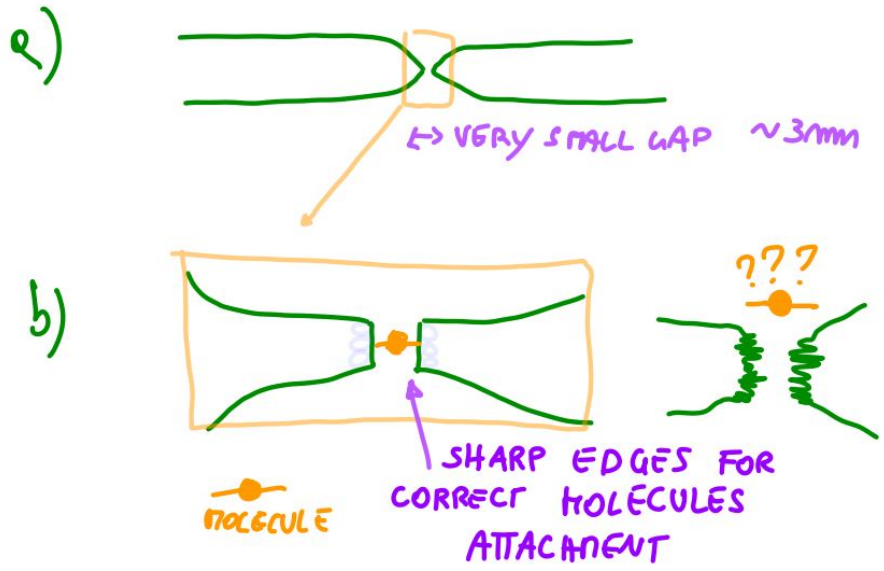


Figure 6.10: very narrow gap (a), but with sharp edges (b)

So, how can we create a nano-gap while harnessing the exponential law?
To detect gaps, we use scanning tunneling microscopy (STM) or atomic force microscopy (AFM) when the width (W) of the gap is less than 10 nm. Therefore, the resistance ranges for gap detection are as follows:

- If the resistance (R) is approximately in the range of 10^2 to 10^4 ohms, there is likely no gap due to the conductive behavior.
- If the resistance is approximately around 10^{14} ohms, the gap is wide enough, and the electrodes are isolated.
- If the resistance falls within the range of 10^5 to 10^{14} ohms, the width of the gap will be around 3 mm.

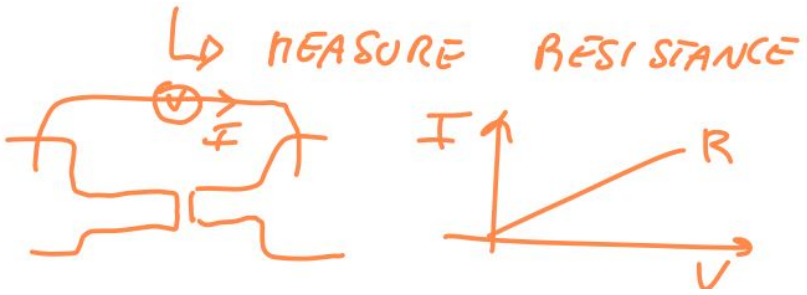


Figure 6.11: the measurement of the Resistor

Therefore, the process of creating nanogaps for molecular electronics using Electromigration involves two main steps:

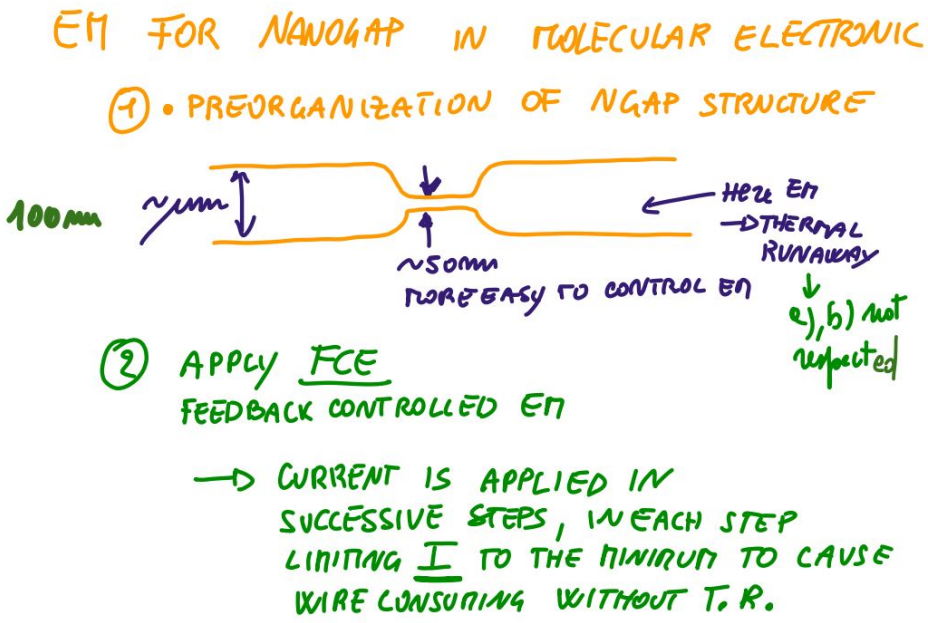


Figure 6.12: a) Preorganization of the Nanogap Structure and b) Feedback Controlled Electromigration

1. Pre-organization of the nanogap structure: This entails constructing a structure consisting of two larger components (on the order of $>$ 100 nm) and a small one (on the order of nm) in the middle, where the nanogap will be formed through the use of Electromigration. These nanometric dimensions are essential to achieve the desired nanogap. Attempting Electromigration on larger dimensions can be very challenging to control and may lead to

thermal runaway. This can result in gap formation, but without meeting the previously mentioned requirements, such as a very narrow structure with precise size control. This lack of control can introduce defects and make it impossible to determine if the molecule will successfully connect.

2. Apply Electromigration with feedback control, referred to as "Feedback Controlled Electromigration" (FCE).

6.4.1 Pre-organization of nanogap structure

We need to prepare the structure so that we can later apply the current for Electromigration. Here, our substrate is typically made of silicon (**Si**), on which we have an insulation layer, often made of silicon dioxide ($\mathbf{SiO_2}$), forming what's known as an SOI wafer. On top of this, we have the actual structure, featuring electrodes that are set up to connect to the external world.

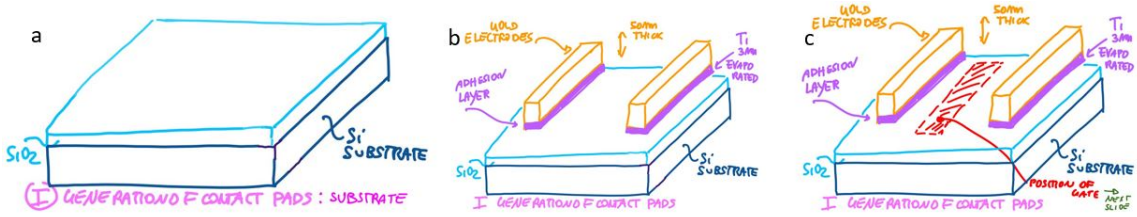


Figure 6.13: Preorganization of the Nanogap Structure

For these electrodes, we first have very thin adhesion layers, approximately 3 nm thick, usually made of titanium (**Ti**). On top of these adhesion layers, we find the gold electrodes. The actual gap is positioned in the middle of these electrodes. In the following picture, we'll zoom in on this specific area.

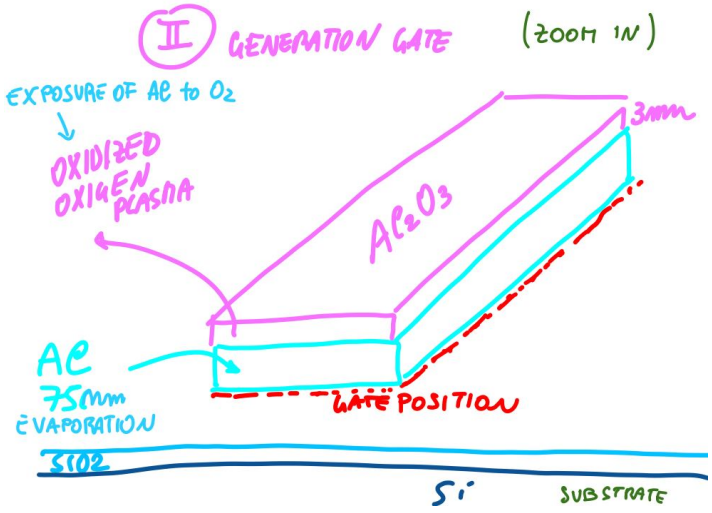
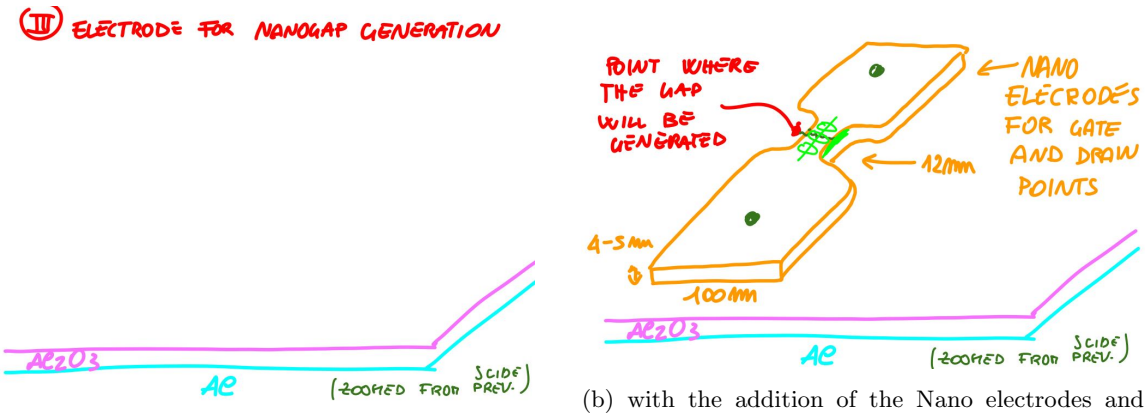


Figure 6.14: Generation of the Gate

Between the electrodes, we begin by depositing a layer of aluminum (around 75 nm), followed by a layer of alumina (approximately 3 nm), which is generated by exposing the aluminum to oxygen plasma, causing it to oxidize. At this point, we'll shift our attention to what we're creating on top of the alumina. The combination of aluminum and alumina essentially forms a tower-like structure that sets the stage for positioning the metal line where Electromigration will take place.
On top of the alumina, we're forming a metal element with a structure as described earlier. This element is made of gold and has the dimensions we mentioned. It's crucial that the central part



(a) Fundamental structure

(b) with the addition of the Nano electrodes and points for the gaps

Figure 6.15: Electrode for Nanogap Generation

is as small as possible for precise control of the Electromigration process. The thickness should also be around 5 nm. If it's thicker,
we'll encounter the same issue we explained for the horizontal direction. However, controlling the thickness is more manageable because we have various techniques at our disposal.

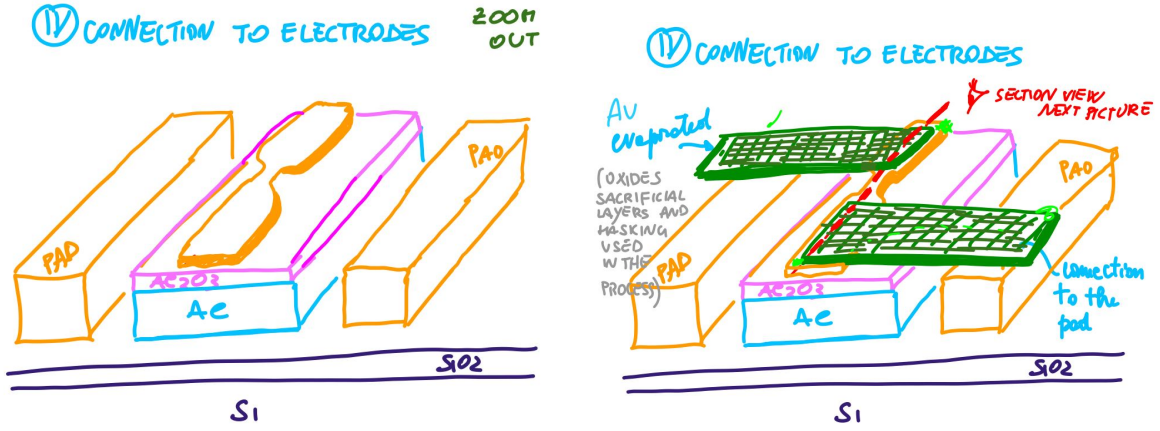


Figure 6.16: Connection to electrodes

Now, let's look at the entire structure. To inject current from the external electrode into the gold line, we need to use another metal electrode. This means we'll use two additional electrodes to connect the gold line to the external world, as indicated above. Now, let's focus on the section visible in the image above because this is where the break will occur. In the image representing the central section of the previous slide, we've highlighted the region where the break will happen, and where we intend to place the molecule. Consequently, we can imagine that the source and the drain can be placed where we've indicated with the black arrows, especially if this structure is going to function as a transistor.

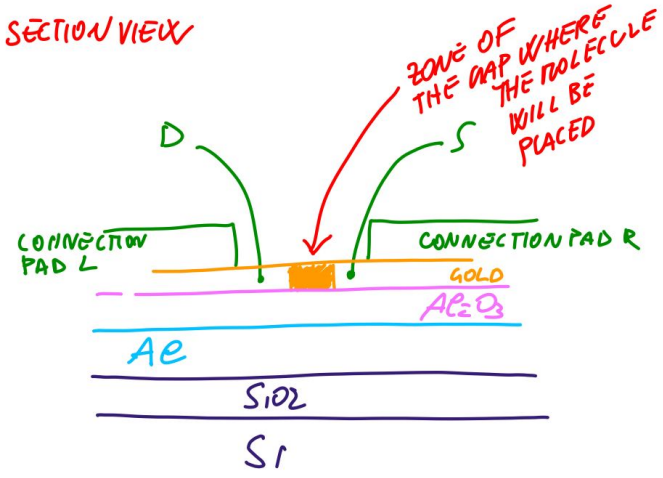


Figure 6.17: section view

In such a case, the gate could be derived from the Al-Al_2O_3 layers (we'll see that by patterning the encircled part, we can create the gate and connect it to the external world through another connection point) or it could come from the top using the air as an oxide layer.

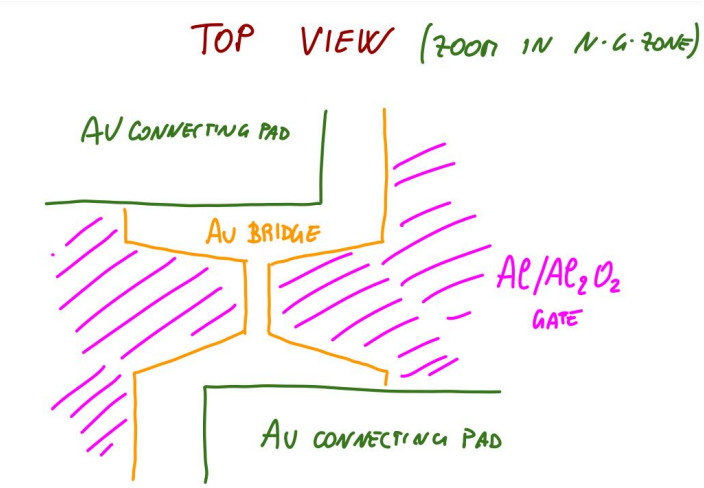


Figure 6.18: top view of the pad

6.4.2 Feedback controlled Electromigration

After pre-organizing the structure, we apply Electromigration with controlled feedback. The current is applied in successive steps, with each step limiting the current to a minimum value that can cause wire consumption without leading to Thermal runaway.

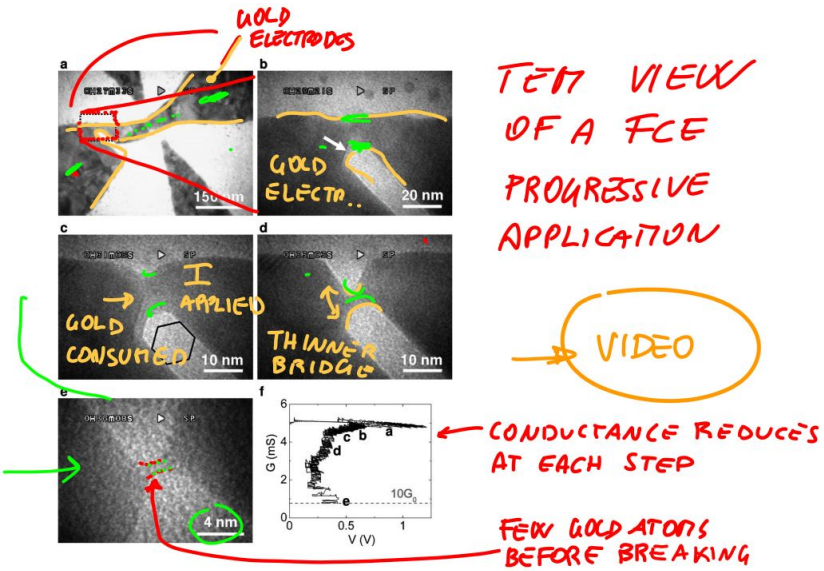


Figure 6.19: FCE progressive application

In the image labeled (a) on the right, we can see the Feedback Controlled Electromigration (FCE) cycle applied to a pre-organized structure. The zoomed-in point in image (b) is where the gap will be created. During the current flow, the atoms in the narrow section gradually move, ultimately leading to a break. Images (c) to (e) illustrate how this section becomes progressively narrower until it consists of only a few gold atoms before breaking, resulting in the gap. At this point, we need to understand what happens during this process. While it's not entirely clear, there are some theories and observations that can help us control the process.
For Electromigration in nanogaps to occur, two requirements must be met:

1. The current density (J) needs to be sufficiently high.
2. The mobility of the atoms must be significant enough to allow mass flow, which can be achieved by raising the temperature. This means we need Joule heating (caused by the high current density) to facilitate this.

The temperature must reach a critical value (T_{CRIT}, critical temperature), and the current density must reach a certain minimum value (J_{MIN}, minimum current density). Simultaneously, we must avoid Thermal runaway, where the temperature could reach the melting point of gold, which is undesirable. If J_{MIN} and T_{CRIT} are reached, the process begins. Metal ions start moving, and the narrow section of the structure becomes thinner. Maintaining the same current from the external pads causes J to increase due to the reduced wire area. This, in turn, leads to higher Joule heating and local temperature, potentially causing problems.

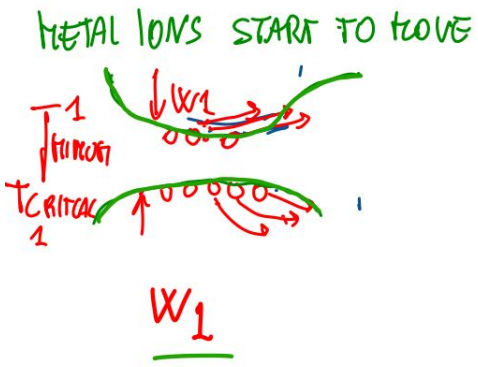


Figure 6.20: Starting the FCE process

For example, if we start with a wire width (W_1), we ensure $J_{\mathrm{MIN},1}$ and $T_{\mathrm{CRIT},1}$ to initiate the process. Ions at the wire boundaries start moving towards the edges, making the wire thinner until reaching a width (W_2). If we maintain the same current, J increases as the wire becomes smaller, leading to elevated temperature. To prevent Thermal runaway or melting, we must halt the process. From the new condition (W_2), we set a new external current value to satisfy $J_{\mathrm{MIN},2}$ and $T_{\mathrm{CRIT},2}$, resulting in a thinner width (W_3).

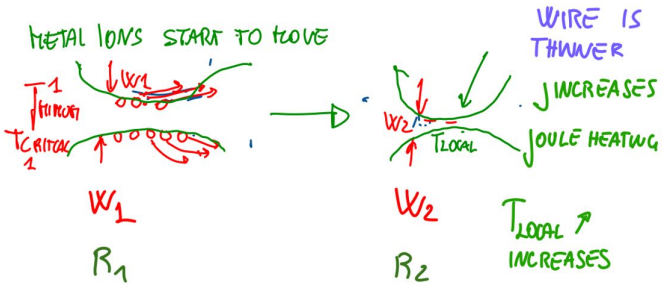


Figure 6.21: The loop of the FCE process

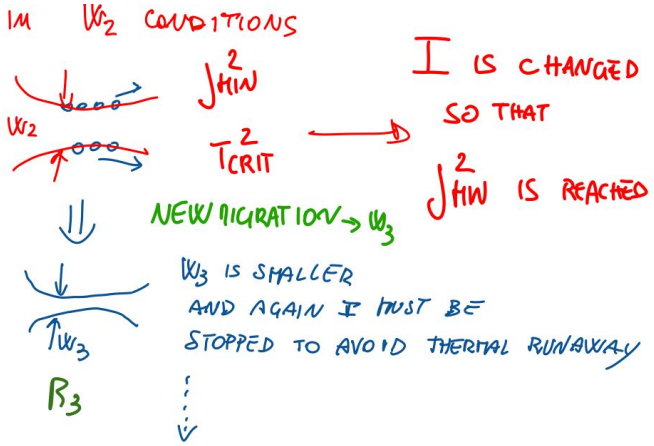


Figure 6.22: Starting the FCE process

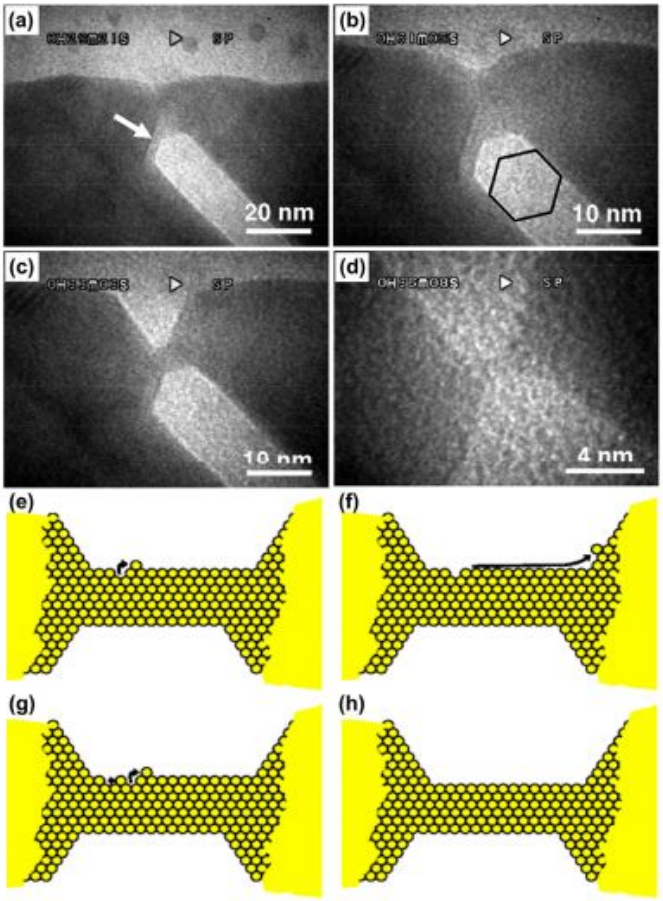


Figure 6.23: Formation of a crystalline gold nanogap during Feedback Controlled Electromigration (FCE). (a−d) Crystal faceting and thinning of the nanogap along the ⟨110⟩ direction, where electromigration is occurring. The lead is pinched down to approximately 1 nm in size, corresponding to about 10 atomic channels of conduction. (e−h) Unzipping model illustrating the evolution of a nanogap, starting with a thermally excited edge atom easily transported away by the applied current along the high-mobility ⟨110⟩ directions, followed by adjacent atoms electromigrating and jumping to the edge through vacancies. Figure reproduced with permission from ref 53. Copyright 2008 American Physical Society.

In the images depicting Electromigration in the wire, when a certain temperature and current density are reached, atoms (especially those at the borders) become mobile and move in the same direction as the electron flow, reaching the edge of the gold wire. Due to the vacancies created in the lattice, this mobility is facilitated, allowing more atoms to move. If not controlled, this process can lead to an avalanche situation (breakdown). To control the flow layer by layer, certain conditions for current density and critical temperature must be met.

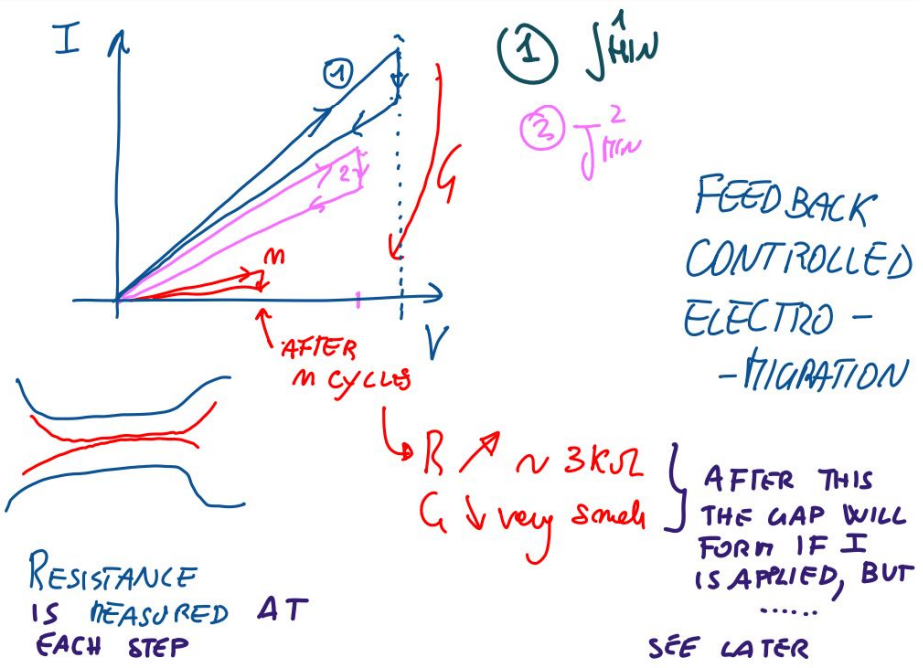


Figure 6.24: Current-Voltage (I-V) characteristics obtained during resistor measurement.

Another interesting aspect is that if the situation is well-controlled without Thermal runaway, the Electromigration process is reversible. This means we can accumulate ions using an electronic current. In the example on the left, we can see how a current flowing from left to right accumulates ions toward the wire, creating a valley (c). By reversing the current, moving ions from right to left (d), we can fill this valley. From an experimental perspective, we measure the current flowing across the bridge as a function of voltage, determining resistance and conductance values. We have critical points from theory, typically derived from the results of many experiments. The next step in this process involves the experimental analysis of the current flowing through the bridge concerning voltage. This examination allows us to measure resistance (R) and, specifically, determine values for both resistance (R) and conductance (G). These values are crucial, as they are based on theoretical models derived from extensive experimental data. The initial phase consists of applying a current (I) and conducting the first of a series of cycles, denoted as 'n.' During this phase, a specific current value, denoted as $J_{\mathrm{MIN},1}$, is required for Electromigration and is related to the parameter W_1. This requirement corresponds to a particular applied voltage (V). As a result, ions within the system begin to flow, leading to a reduction in wire thickness and, consequently, an increase in resistance (R) (and a decrease in conductance (G)). The cycle is terminated when resistance (R) reaches a predefined threshold. Subsequently, we continue with additional cycles, each associated with its unique critical resistance values linked to $J_{\mathrm{MIN},i}$. Throughout each cycle, conductance (G) diminishes, leading to a corresponding increase in resistance (R). The process is considered successful when resistance (R) reaches a significantly high level, typically around 3 kΩ6.24, resulting in extremely low conductance (G). At this point, the wire's dimensions are reduced to a few rows of atoms, marking the limit of the Electromigration process. It is from this point that we can determine the next course of action. One potential strategy involves creating the nanogap by applying a final run of current (I). While this is a viable option, we must carefully consider other factors before proceeding further.

6.5 What's Next?

A molecule must be trapped in the gap! But how?

Place the structure in a solution before the final nanogap (NG) step to avoid other impurities from filling the gap. Molecules are very important.
There are two possible choices: 2.1. Insert Gap Structure into Solution Apply the last current cycle to create the gap. (Set resistance, R, to infinity)

- 2.1. Insert Gap Structure into Solution Apply the last current cycle to create the gap. (Set resistance, R, to infinity)
- Wait for the molecule to connect to the gap's sides.

The second stage involves creating the gap, and we have two possible approaches:

1. The first approach is to place the system in a solution with a specific concentration and then apply a final cycle of current. This generates the gap, and we wait for molecules to connect (as they are present in the surroundings, and some may get trapped in the gap). However, this approach has two issues:
   - Firstly, even if we successfully create the nanogap with the molecule inside, applying current during the final process can lead to the accumulation of metal atoms around the gap, especially at the edges. This can hinder the connection between the molecule and the electrodes, affecting charge injection. We aim for robust connections to ensure efficient conduction6.25.

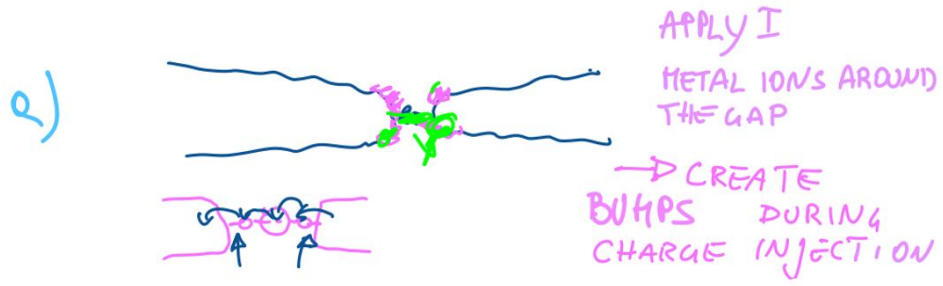


Figure 6.25: Deposition of the molecule

   - Secondly, if the molecule is present during the final cycle, high temperatures may be involved, and depending on the molecule, this could alter its conductance.
2. The other option leverages natural Electromigration. At the end of the previous process, we have a very thin chain of atoms. If we place this in a solution at a controlled temperature, Electromigration occurs naturally over time, leading to gap creation without the need for an external current (e.g., 1 hour @ 100 K)6.26. This happens because the electrons in this thin and unstable structure have sufficient energy to allow metal atoms to move freely. In this case, the gap's quality is usually excellent, with no metal atom-related irregularities near the ends and no damage to the molecule.

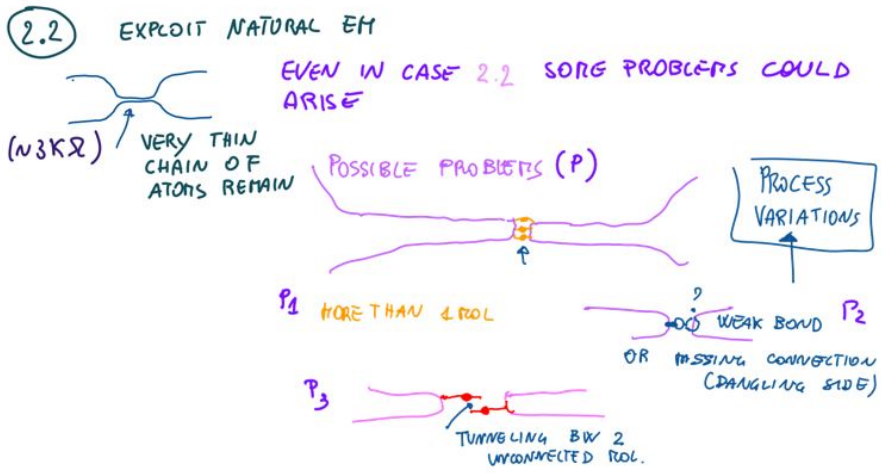


Figure 6.26: Controlled system of the Temperature

Typically, the second choice is the preferred method.
However, potential issues arise with this method. We might have created a pristine gap, but there's a possibility of having more than one molecule present. In such cases, we could experience weak bonds because the gap may not be perfect, leading to variations in bonding quality at the edges. There could also be scenarios where we have a gap with two molecules, one of which lacks a bond with one electrode, and the other lacks a bond with the opposite electrode. However, between them, we may observe tunneling effects.
These issues are referred to as "process variations." We need to address them, and while we can try to control them to some extent, we must anticipate what might occur. For instance, measuring resistance or conductance can help us understand these situations (more molecules in the gap result in higher conductance, a weak bond leads to lower conductance, and tunneling behavior depends on various parameters).
In general, predicting and managing these variations is challenging. Even in standard FinFET technology, similar problems exist. Transistor models incorporate probability density functions that characterize the behavior of various parameters. These probability density functions are filled with data obtained from numerous experiments. By integrating this data into computer-aided design (CAD) tools, we can evaluate the impact of these process variations at different technological levels, from the device level to the gate or system level.

6.6 Understanding Electromigration and its Modeling

Let's now explore Electromigration and its typical modeling approaches. Although not fully elucidated, from a theoretical perspective, it is often described using the "thermodynamics of irreversible processes theory." According to this theory, Electromigration can be comprehended through three fluxes denoted as J_e (the flux of electron particles), J_m (metal ions flux), and J_u (energy flux). These fluxes are induced by forces, or potential gradients, represented by X_j, where j can be associated with electron particles (e), metal ions (m), or energy (u).

- J_e - Electron Flux
- J_m - Metal Flux
- J_u - Energy Flux

$$X_k = -\nabla \mu_{ec} \mathbf{k} \tag{11}$$

The fluxes of electrons and metal particles are influenced by the potential gradient of electrochemical potential (μ_{EC}), which can be expressed as shown in the slide. Specifically, ϕ denotes the electrostatic potential, μ represents the chemical potential, and Z signifies the particle charge. Notably, the charge differs for electrons (-1) and metal ions ($+1$). Furthermore, the energy flux is governed by the local temperature gradient.

$$\mu_{\mathrm{EC}} = \mu + Z \cdot e \cdot \phi \tag{12}$$

Where:

ϕ - Electrostatic potential

μ - Chemical potential

Z - Particle charge (For electrons: -1, For metal ions: +1)

When we formulate equations for Electromigration in nanogaps, they are refined and simplified. The general description of the system is as follows:
For J_m (metal ions flux), we introduce constants such as $L_{m,m}$, $L_{m,e}$, and $L_{m,u}$. These constants are multiplied by various gradients, including the gradient of the electrochemical potential for metal ions over temperature (T) and the gradients of the electrochemical potential for electrons over T and temperature squared (T^2). As evident, J_m depends on these gradients and constants. A similar approach is applied to the other two fluxes, each with its specific set of constants.

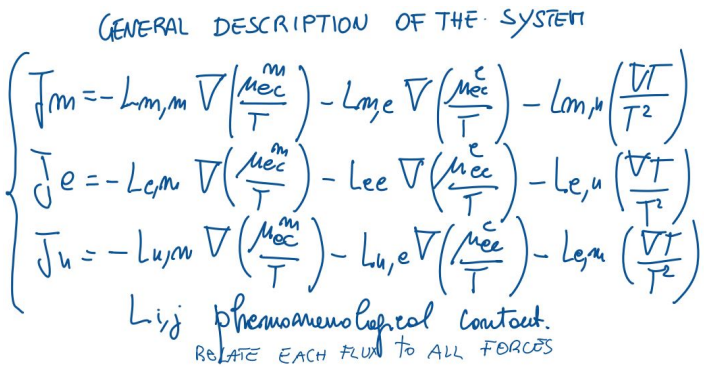


Figure 6.27: The general description of the system

$$J_m = -L_{m,m}\nabla\left(\frac{(\mu_{ec})^m}{T}\right) - L_{m,e}\nabla\left(\frac{(\mu_{ec})^e}{T}\right) - L_{m,u}\frac{\nabla T}{T^2} \tag{13}$$

$$J_e = -L_{e,m}\nabla\left(\frac{(\mu_{ec})^m}{T}\right) - L_{e,e}\nabla\left(\frac{(\mu_{ec})^e}{T}\right) - L_{e,u}\frac{\nabla T}{T^2} \tag{14}$$

$$J_u = -L_{u,m}\nabla\left(\frac{(\mu_{ec})^m}{T}\right) - L_{u,e}\nabla\left(\frac{(\mu_{ec})^e}{T}\right) - L_{e,m}\frac{\nabla T}{T^2} \tag{15}$$

Each flux intricately depends on the gradients of these three forces, and the constants involved are phenomenological constants. These constants are functions of the specific environmental conditions, geometry, and materials involved.
This is used not only to explain Electromigration but also to explain thermoelectric effects, thermo-diffusions, and so on. In each of these cases, some of the terms are more important than others. Solving these equations in 3D is very complicated, so we'll need some assumptions considering our specific case of Electromigration in nanogaps.
First of all, in materials with high conductivity, the gradient of the electrochemical potential for the electrons can be ignored. It has also been demonstrated that thermal diffusion for gold electrodes can be ignored. As a final condition, we say that in metal wires, the flux of charges is due only to the electrons, and we can introduce the resistivity. J_m can be rewritten as:

$$J_m = L_{m,m}\nabla\mu_m - Z^* e\rho J \tag{16}$$

Where ρ is the resistivity, J is the current density, and Z^* is an effective particle charge. The effective charge is given by the expression:

$$J_m = -L_{m,m}\left(\frac{\nabla\mu_m - Z^* e\rho j}{T}\right) \tag{17}$$

Electrostatic potential: ϕ
Metal flux (J_m): $-L_{m,m}\nabla\mu_m - Z^* e\rho J$
Voltage (V_1): $pJa^2 - L_{\epsilon s}T$
Current density (J)
The correction term depends on the fact that there is a transfer of momentum from electrons to ions of the metal. Actually, the atoms of the metal behave as if they have the charge Z^* due to this momentum transfer from the electrons. While quantifying this correction is not straightforward, it helps explain the actual charge that the ions will have. In general, this correction is much larger than Z, resulting in a negative Z^*, which means that the final net force that acts on the gold atoms will be in the same direction as the electron flow (as observed experimentally).

$$Z^* = Z - \frac{L_{m,e}}{L_{m,m}} \quad \text{if } \frac{L_{m,e}}{L_{m,m}} \gg Z \tag{18}$$

Generally, in Electromigration (EM), if Z^* is negative, the net force acting on the gold atom is in the direction of the electron force.
In practice, metal atoms behave as if they had a charge Z^* due to momentum transfer of ions. The gradient of μ for metal ions is given by $\Omega\frac{d\sigma}{dx}$, and the metal flux J_m is expressed as:

$$J_m = -L_{m,m}\left(\frac{\Omega\frac{d\sigma}{L} - Z^* e\phi j}{T}\right) \tag{19}$$

Another assumption is introduced, where the force due to stress is considered. The movement of ions creates a mechanical stress in the structure that can oppose the ion movement. The gradient for ions depends on an experimental quantity multiplied by the derivative of stress, denoted as Γ. If we consider a wire with a length L and substitute it into the previous equation for J_m, we obtain the expression mentioned earlier.

$$J_{\min} = \frac{\Omega\Delta\sigma_{\max}}{Z^*\rho eL} \tag{20}$$

In a typical electronic circuit, there's a specific level of current density that must be reached for ion migration to occur due to electromigration. If the current density is below this threshold, electromigration won't happen. We can determine this minimum current density using a parameter known as Blech's length
As long as the gradient of stress remains smaller than a certain maximum value (which is observed experimentally and depends on the metal and geometry), the Electromigration force is balanced by the stress gradient, resulting in $J_m = 0$. This implies that a certain force must be applied to induce Electromigration, and this explains why Electromigration is observed in wires of a certain length in integrated circuits (ICs), even with significant current. The wire length plays a role in creating the nanogap, as the mechanical stress can oppose ion movement in shorter wires.

$$D = L_{m,m}\left(\frac{d\mu}{dc}\right) \tag{21}$$

At this point, we introduce the *Diffusion constant*, denoted as D, through the Einstein relation, where C represents the concentration of metal ions, and μ is the chemical potential as defined previously. The diffusion constant D can also be expressed using the second formula, incorporating E_a, the activation energy for ion diffusion on the metal surface (which varies depending on the metal type, e.g., for Au, it is 0.12 eV). By utilizing this information, we can derive $L_{m,m}$, which can then be substituted into the expression for the flux of metal ions due to Electromigration, denoted as J_m.

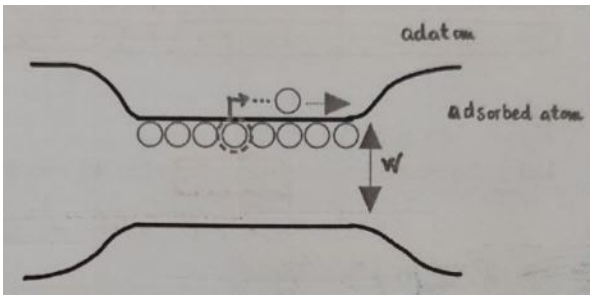


Figure 6.28: Adatom with the differences

$$\mu = kT\ln(C)$$
$$D = L_{m,m}\frac{kT}{C}$$
$$D = D_0 e^{-\frac{E_a}{kT}}$$

Upon substituting $L_{m,m}$ into the equation, we observe that J_m depends on the difference between the current density J and the minimum current density required to initiate Electromigration, known as J_{MIN}. The term α encompasses all remaining factors from the previous equation.

$$J_m = -L_{m,m}\left(\frac{\Omega\Delta\sigma}{L} - Z^* e\rho j\right)\cdot\frac{1}{T} \tag{22}$$

$$J_m = f(J_{\min}, T) = -L_{m,m}\left(\frac{\Omega\Delta\sigma}{L} - Z^* e\rho j\right)\cdot\frac{1}{T} \tag{23}$$

$$J_m = \frac{\alpha}{T}\cdot(J - J_{\min})\cdot e^{-\frac{E_a}{kT}} \tag{24}$$

$$\alpha = \frac{CD_0\cdot Z^* e\rho}{k} \tag{25}$$

This equation describes the flux of metal ions but only when $J > J_{\text{MIN}}$. Otherwise, J_m will be approximately zero. Notably, J_m exhibits an exponential dependence on temperature and especially on E_a. This relationship bears similarity to Black's equation, which we discussed earlier in the context of Electromigration in nanogaps. It highlights the significance of temperature and a certain critical current in initiating the Electromigration process.
So far, we've developed a formulation specific to our case from a general theory. However, we still need to consider Joule heating and integrate it with the Feedback Controlled Electromigration (FCE) process we discussed earlier. It's important to note that this discussion primarily addresses a single occurrence of the Electromigration phenomenon, whereas in practice, it occurs repeatedly with different voltage values, each associated with varying minimum current conditions due to changing circumstances.

Returning to J_m, it describes the evolution of ion flux, with its temperature dependence implying that as current continues to flow, the temperature increases, affecting ion flow. The theory and practice dictate that Electromigration only commences when the ion flux reaches a sufficient level. This not only involves the minimum current J_{MIN} related to resistivity and mechanical strength but also a limit associated with temperature and Joule heating.
In essence, when J reaches a certain critical value J_{CRIT}, such that the temperature in the specific region where Electromigration occurs (commonly referred to as the nanogap) surpasses a critical temperature T_{CRIT}, Joule heating is activated. This implies that the local temperature must exceed T_{CRIT} when the power dissipated in that section of the wire, denoted as P, exceeds a critical value P_{CRIT}. Notably, P is proportional to resistivity and J^2, leading to P_{CRIT} being dependent on resistivity and J^2_{CRIT}. Therefore, when J reaches J_{CRIT}, sufficient power is dissipated, raising the temperature—a phenomenon known as Joule heating.
STRESS: If $J_m = 0$, it means that the stress required for ion migration has not been reached. Stress is a crucial factor in initiating the migration process.
TEMPERATURE: The expression for J_m ($J_m = \alpha \cdot (J - J_{\text{min}}) \cdot e^{-\frac{E_a}{kT}}$) shows that it depends on temperature (T).
FLUX OF IONS: The flux of ions, J_m, is related to the formation of the nanogap. It starts when the conditions are right, specifically when the flux of ions is large enough.
GAP FORMATION: The process of gap formation depends on several factors, including stress and temperature. It initiates as soon as the local temperature in the gap zone is influenced by Joule heating, causing an increase in temperature.
Consequently, it's crucial for the local temperature to reach a certain threshold for the process to initiate. When T_{CRIT} is achieved through this mechanism, ion mobility is enhanced to a point where vacancies and *adatoms* (adsorbed atoms) are formed. *Adatoms* typically rest on the metal surface, while vacancies are created as atoms break bonds and begin to move. Only when this occurs does the flow of ions become possible.
However, as migration happens, the wire reduces in size. Consequently, the resistance (R) increases, and the resistivity (rho) also increases. Additionally, the temperature (T) increases. In this scenario, there is a risk of thermal runaway, which could be dangerous.
Let's take a moment to examine two different experimental scenarios involving two distinct types of wires, each structured differently. This will help us understand the differences in behavior.

1. **Terrace Type Connection**: In this setup, when voltage is applied across the wire, the current gradually increases until a point where Electromigration begins. After some time, the wire breaks. This process unfolds relatively slowly. Following the break, we're left with a 1 nm gap, and a measurable current of around 1 nA, attributed to tunneling 6.29. The measurement of this current allows us to confirm the presence of the gap. If, instead, we measure 0A current, it indicates a significant short circuit, indicating the gap is present but not in a desirable condition.

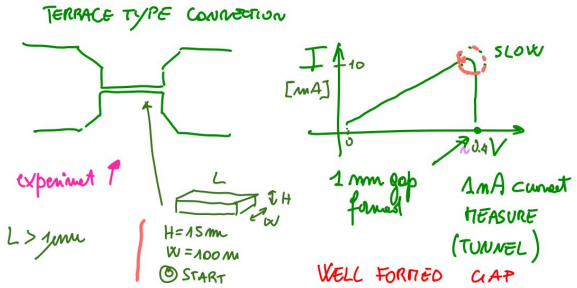


Figure 6.29: Terrace Type Connection

2. **Bow Tie Connection**: In this scenario, the current also increases with applied voltage but abruptly drops to zero at a certain point, which happens more quickly than in the Terrace Type Connection 6.30. After the break, we observe clusters of metal atoms surrounding

the gap. However, in this case, the gap measures around 100 nm, which isn't suitable for our purposes. The measured current is 0A, signifying a more unfavorable outcome.

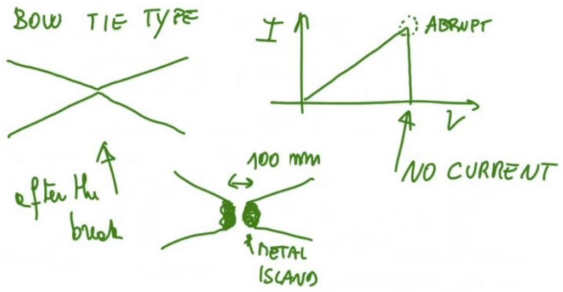


Figure 6.30: Bow Tie Connection

These two cases illustrate different behaviors and outcomes in Electromigration experiments, each with its own implications and challenges.
The difference in behavior and results between these two conditions is attributed to the occurrence of thermal runaway in the Bow Tie case, where the local temperature eventually reaches a critically high value due to excessive power dissipation. This leads to the gold wire reaching its melting point, resulting in the formation of islands of atoms. However, this phenomenon does not occur in the Terrace Type case.
The occurrence of thermal runaway and the subsequent melting of the metal is undesirable. But why does this happen?

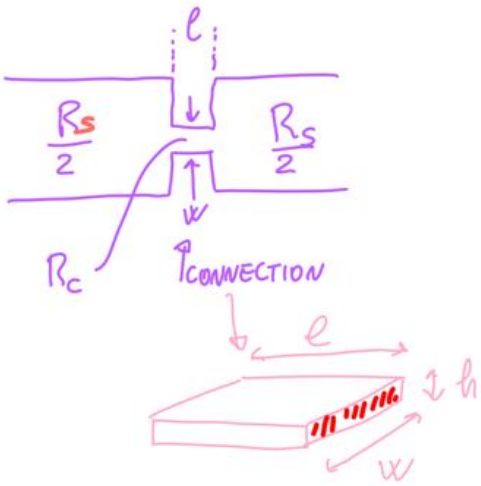


Figure 6.31: Connection Model

$$R_{\mathrm{TOT}}(t) = R_S + R_C(t) \tag{26}$$

$$R_C(t) = \frac{\rho \cdot \mathrm{L}(t)}{h(t) \cdot w(t)} \tag{27}$$

Consider a small connection, as shown above, with certain dimensions l and w at the connection point where Electromigration is expected. Additionally, there are two larger branches where we

have a certain resistance, which we model as a series resistance R_S equally divided between the two branches.
Now, let's introduce a time-dependent resistance $R_C(t)$ at the connection point due to the ongoing Electromigration process. This means that the total resistance can be expressed as $R_{\mathrm{TOT}}(t) = R_S + R_C(t)$. Specifically, $R_C(t)$ follows the expression provided in the slide, where we have the resistivity ρ and the length 'l', which are assumed to be constant. However, the thickness 'h' and the width 'w' of the wire (the section if we consider their product) change over time due to the Electromigration process.
If we increase the applied voltage ΔV, we observe an increase in current. At a certain point, J_{MIN} is reached, allowing J_m to initiate. This entails some movement of ions, and a certain level of mobility is required, which can be achieved through temperature (Joule heating), as previously mentioned.

$$p(t) = p_{\mathrm{critic}} \left(\frac{1 + \frac{R_s}{R_c(0)}}{1 + \frac{R_s}{R_c(t)}} \right)^2 \tag{28}$$

Using the parameters we have, Electromigration initiates, leading to the formation of narrow cuts within the wire. This signifies an increase in the resistance of the connection point, denoted as R_C, and consequently, an escalation in the locally dissipated power, $P(t)$. This power can be described by the equation mentioned earlier, capturing the rise in local dissipation at the connection point.
In this context, if the ratio $R_S/R_C(0)$ is relatively small—meaning that the series resistance R_S is significantly smaller than the initial $R_C(0)$—the dissipated power tends to hover around a critical value, P_{CRIT}. This critical value depends on the square of the critical voltage V_{CRIT}, which is necessary to achieve the critical power corresponding to a critical temperature. This critical temperature is where Joule heating and temperature effects activate the desired mobility. Additionally, P_{CRIT} is influenced by resistivity (ρ), the length (l) of the wire, and the initial ratio $R_S/R_C(0)$.

$$p_{\mathrm{critic}} = \frac{V_{\mathrm{critic}}^2}{\rho \cdot L^2} \left(1 + \frac{R_s}{R_c(0)} \right)^2 \tag{29}$$

However, when $R_S/R_C(0)$ is greater than one (indicating a high series resistance compared to R_C), $P(t)$ increases rapidly. This implies that the local temperature at the connection point can swiftly approach the melting point of the electrode material.
This analytical explanation aligns with experimental findings. In the two cases we discussed earlier, the Bow Tie connection exhibited a high series resistance compared to the resistance of the connection point itself, while the Terrace Type connection did not.
One important consideration is that during Electromigration, the wire experiences cuts, causing the resistance $R_C(t)$ to increase over time. If R_S is too large, the ratio $R_S/R_C(t)$ becomes excessively high, resulting in a significant increase in the dissipated power. Conversely, if R_S is small, the power dissipated over time remains relatively constant.
This provides practical insights into what to expect. In the Terrace Type connection, the power dissipated remains controlled and limited, whereas in the Bow Tie connection, it cannot be effectively managed.
Now that we have a better understanding of the process involving Joule heating and its implications for design, let's examine some values in an example experiment to gain further clarity.
In this context, we are examining a Terrace type connection. Our critical temperature (T_{CRIT}) is approximately 345 K, and the minimum current (J_{MIN}) is around 4×10^8 A/cm^2 for gold. An intriguing aspect to consider is that we are primarily concerned with the local temperature of the connection, which is not significantly influenced by the surrounding environment. However, there exists an indirect relationship with the environmental temperature. To achieve the critical power (P_{CRIT}) necessary to elevate the local temperature to the required level, it becomes essential to

increase the environmental temperature. For instance, if the environmental temperature shifts from 295 K to 4.2 K, as was the case in this experiment, the critical current (J_{CRIT}) must increase by 50% to attain the critical power (P_{CRIT}) and consequently, the critical temperature (T_{CRIT}).

6.7 Parallel Fabrication of Nanogaps via FCE

In our system, we aim to achieve parallel fabrication of nanogaps using a method known as Field-Controlled Electromigration (FCE). This involves an array of electrodes (wires) where we incorporate inputs, power supply, and ground connections. Positioned strategically among these nodes are molecular transistors, arranged in a specific manner to enable their interconnection and the generation of logic functions.

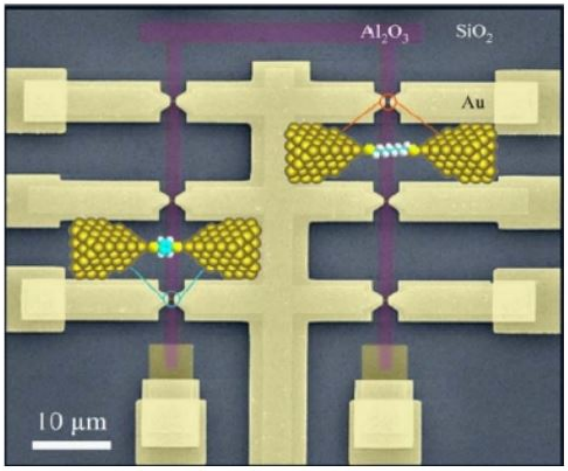


Figure 6.32: Nanogaps array based on the Nanogaps

This concept bears a resemblance to the fundamental idea behind Field Programmable Gate Arrays (FPGAs), which are characterized by an abundance of transistors that can be "programmed" to configure their connections in various ways. This flexibility allows for the rapid creation of logic functions, enabling the reconfiguration of our system to suit different applications.
Before the advent of FPGAs, we had Programmable Logic Arrays (PLAs) in the past. These were simpler in structure and comprised different identical elements, including transistors, which connected arrays of wires to perform logic functions.

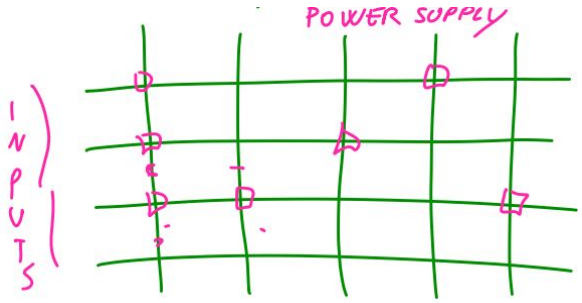


Figure 6.33: The expected system based on Digital aspects

Structures like the molecular elements can currently only be organized in a specific manner, utilizing an array of electrodes interconnected at specific points. This arrangement allows us to

synthesize logic functions using molecular transistors generated within this setup. Initially, we create the gaps and subsequently place the molecules.
If we could fabricate multiple gaps in parallel, we would simply immerse the entire array in a solution of molecules, enabling them to connect simultaneously. However, transitioning from a single element to multiple elements often necessitates a reevaluation of how we design the individual device to seamlessly integrate it into the larger system—a challenging task.

6.7.1 Single Device

In the first scenario, we focus on a single device. This device encounters the resistance of the lead (the metallic wire surrounding it) and the resistance at the connection point. When considering the conductance (G) in relation to the applied voltage (V), it starts at a stable point and gradually decreases as a specific voltage is applied, initiating Electromigration. There's a critical point where Electromigration begins, and we must exercise caution to avoid reaching the melting point of the electrodes. We then reduce the applied voltage and repeat this process until the desired gap is created. The behavior of this process is illustrated in the diagram below.

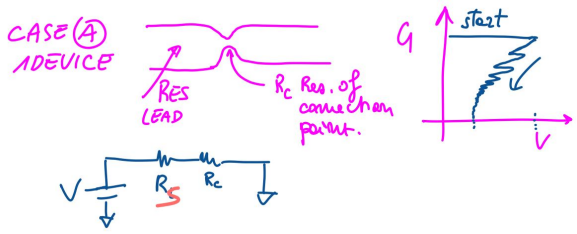


Figure 6.34: Single Device

We can model this situation using the sketch we can notice. In the Field-Controlled Electromigration (FCE) process, we adjust V so that the power P at the junction is maintained above the critical value P_{CRIT}. This ensures that all necessary processes are activated while still being below the value required to reach the melting point of the electrodes. We employ a feedback system to optimize this delicate balance. After several iterations, the resistance of the connection point (R_C) stabilizes at a final value, which is approximately $4\,\text{k}\Omega$. This represents the limit just before the break occurs.
At this stage, as we've discussed, we have the choice to either proceed with a final step in the process or to wait for natural Electromigration to take its course. At this stage, as we've discussed, we have the choice to either proceed with a final step in the process or to wait for natural Electromigration to take its course.

6.7.2 Two Devices

In this scenario, we have two devices that are separated but controlled in parallel, and in theory, they should be identical. The equivalent model for this situation is depicted on the right in the previous slide, where we have the lead resistance for each device in series with the resistances of the two connection points.

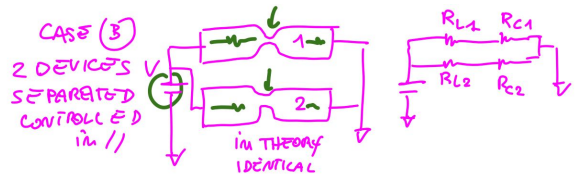


Figure 6.35: Two Devices

Since theoretically they are identical, we expect to have $R_{L1} = R_{L2}$ and $R_{J1} = R_{J2}$. We also anticipate that when we apply a voltage V, they should break at the same moment and in the same manner. However, what actually occurs is that in one device, the conductance follows the expected behaviour initially but then abruptly increases again. In the end, one gap behaves as expected, while the other exhibits irregularities, indicating that the quality of the connection for the molecule will not be good.

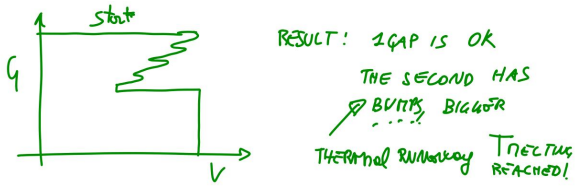


Figure 6.36: Gap vs Voltage

For the second device, we experience thermal runaway, signifying that the melting point has been reached. This discrepancy arises due to process variations, meaning that the two devices are not truly identical, even if we designed them the same way. In practice, R_{J1} differs from R_{J2}, and R_{L1} differs from R_{L2}, albeit slightly.
For instance, after some time into the Electromigration process, R_{J1} might be around $4.2\,\Omega$, whereas R_{J2} is approximately $4\,\Omega$. This implies that one of these two devices will break before the other, most likely the one with the higher resistance, as it has eroded slightly more (with fewer atoms remaining). Once this break occurs, even though it's in the expected manner, all the current flows into the other device. Unfortunately, our measuring system cannot react immediately, leading to a melting event.

6.7.3 Two devices with series resistance

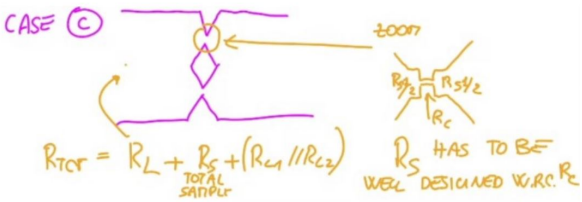


Figure 6.37: Two devices with series resistance

Here, we are incorporating a series resistance (R_S) connected with the junction resistance. If we zoom in on the connection point, we have R_S in series with R_C. Therefore, the total resistance (R_{TOT}) can be expressed as:

$$R_{\mathrm{TOT}} = R_{\mathrm{LEAD}} + R_S + \left(\frac{1}{R_{C1}} + \frac{1}{R_{C2}}\right) \tag{30}$$

Here, R_{LEAD} represents the resistance of the larger wires, R_S is the series resistance for the entire sample, taking into account R_{S1} for the first junction and R_{S2} for the second one. It's crucial that R_{S_i} is well designed with respect to R_{C_i}.

Figure 6.38: Gap vs Voltage

In practice, if the total R_S is smaller than both R_{C1} and R_{C2} (which also implies that the individual series resistances are smaller than each of the connection resistances), then the power at the junctions remains constant.

$$\frac{\partial P_1}{\partial R_{c1}} < \frac{\partial P_2}{\partial R_{c2}} \tag{31}$$

This prevents us from entering the runaway process. So, even if, for example, $R_{C1} > R_{C2}$ due to process variations, initially, the current will flow to the second connection point, increasing the power there. Then, at some point when $R_{C1} < R_{C2}$, the current will shift to the first connection point, increasing the local power. This means that the two Electromigration processes proceed in parallel with a form of compensation, resulting in the correct generation of nanogaps at the same time. If we examine the behavior of conductance (G), we'll observe significant steps initially, followed by more refined steps.
This approach enables the creation of several connections in parallel, but it requires precise design and reliability. Tens of elements in parallel have shown promising results. However, it's worth noting that this isn't the only procedure used; there are other effective techniques, such as mechanical methods. In fact, one approach is to use Electromigration after an initial mechanical application.

6.8 Generation of the gaps based on mechanical stress

The generation of the gaps based on mechanical stress is the process of creating nanogaps by applying mechanical forces to a material. This can be done using a variety of methods, such as mechanically controlled break junction (MCBJ), crack-defined break junction (CDBJ).

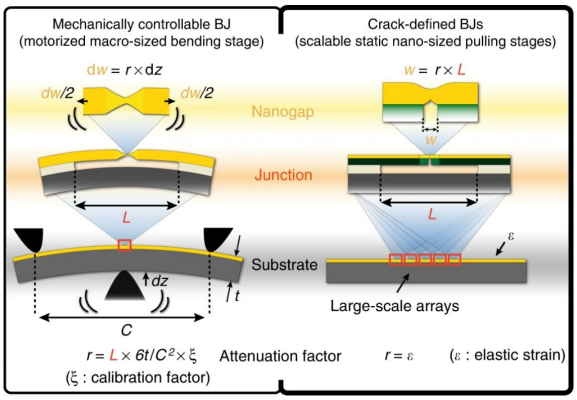


Figure 6.39: Cross-sectional schematics of a MCBJ integrated in a 3-point bending stage and of CDBJs. The dynamically controllable mechanically controllable break junction (MCBJ) allows for continuous monitoring of the breaking process and tuning of the resulting inter-electrode separation.

6.8.1 Mechanically Controlled Break Junction (MCBJ)

MCBJ is a powerful technique that leverages mechanical forces to meticulously manipulate and regulate nanoscale gaps. This method offers precise control over gap formation and measurement, making it invaluable in various scientific and technological applications.

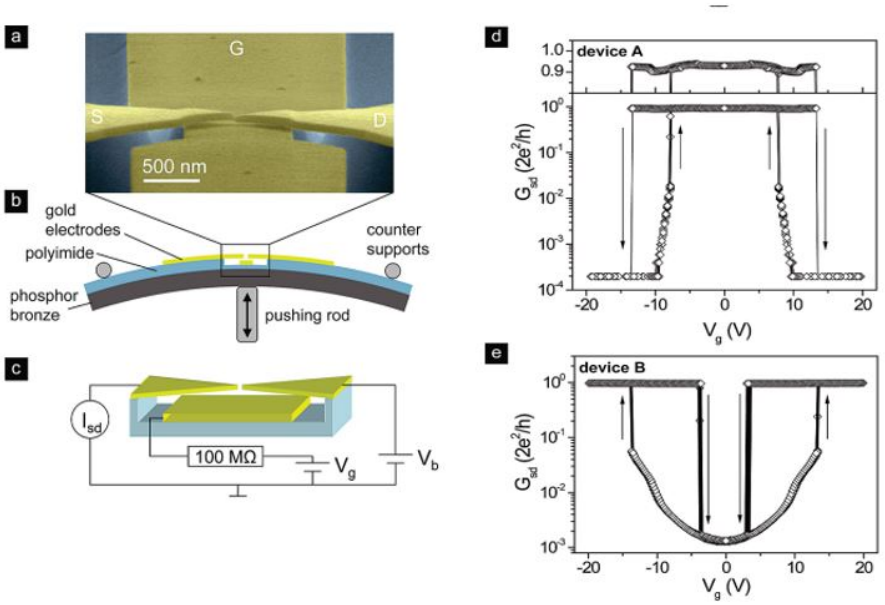


Figure 6.40: Gate-induced reversible switching of single-atom contacts in mechanical break junctions. (a) Electron micrograph of a gated mechanical break junction (Device A, colorized for clarity). The source-drain electrodes (S, D) are suspended 45 nm above the gate (G) and have been stretched and broken by bending the flexible MCBJ substrate. Their tips are slightly misaligned, possibly due to the release of processing-induced stress in the metal film. (b) Schematic of the three-point bending setup. The pushing rod controls the distance of the electrodes via the bending of the substrate. (c) Schematic of the electrical measurement circuit, including a protective series resistor. (d) Switching characteristics of a gated mechanical break junction during 20 consecutive gate voltage cycles at a bias voltage of 50 mV, presented on a linear and semilogarithmic scale. Each gate voltage sweep took about 200 s. (e) Switching characteristics of another device in which a single atom contact is established at high gate voltages. Measurement parameters are identical to those for Device A, but the switching direction is inverted.

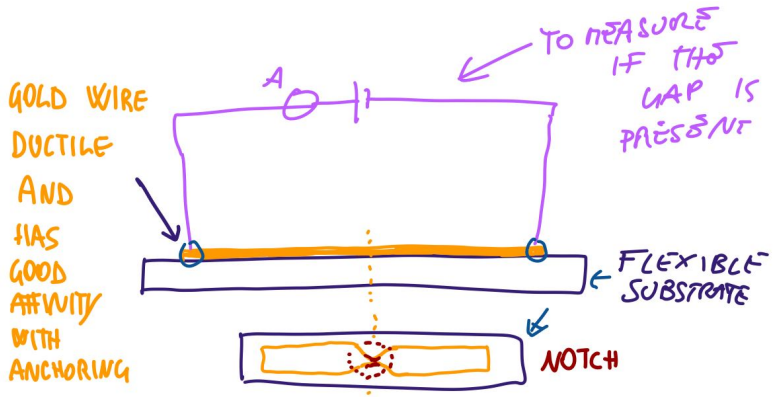


Figure 6.41: mechanically controlled break junction

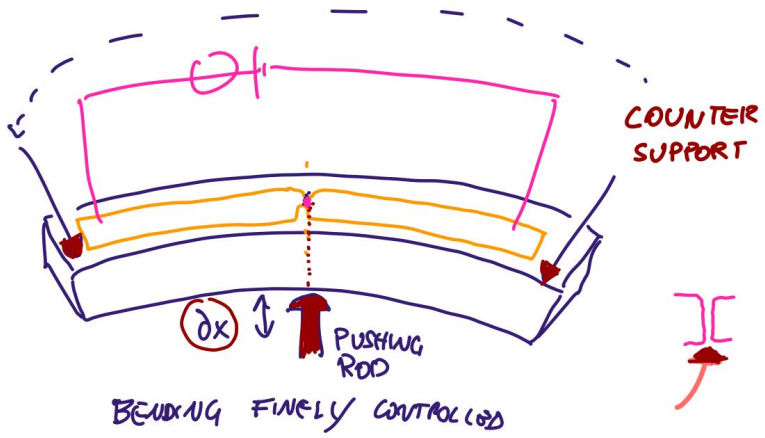


Figure 6.42: Bounding finally get controlled

Imagine having a flexible substrate onto which a layer of Gold is deposited. Gold is chosen for its chemical affinity with anchoring groups used for molecules and its ductile nature. A current measurement system is integrated to detect the presence or absence of a gap; if the current flows, there is no gap, and if it ceases, a gap has formed.

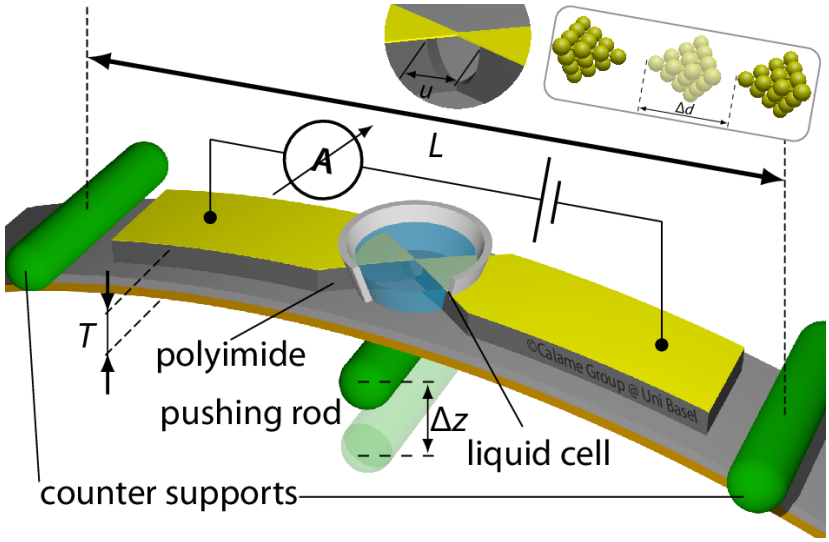


Figure 6.43: Artistic View of Atomically Sharp Electrodes and Relative Displacement

The flexible substrate is secured to supports, with counter supports at the ends and a central rod for substrate manipulation. To promote the desired break, we typically focus on a point coinciding with the rod placement, employing a finely controlled 3-point bending mechanism. The narrow section of Gold on the substrate is stretched until it fractures, creating nanogaps. This method results in high-quality gap formation, essential for anchoring molecules.
However, calibrating the system to precisely control this point is a challenge, especially when generating multiple nanogaps in parallel. Parallel generation of gaps remains impractical, even with precise control of individual gaps. This limitation mainly pertains to laboratory-level experiments.
The attenuation factor (a), which signifies the ratio of relative electrode displacement (Δd) to substrate bending (Δz), is a crucial parameter in the MCBJ system. It is determined by the sample's geometry and can be approximately estimated using the formula: $a = 6uT/L^2$. Here, 'u' is the length of the free-standing metallic bridge, 'T' is the substrate's thickness, and 'L' is the distance between side counter supports.

Historical Developments

Historically, the first MCBJ devices were constructed using notched metallic wires affixed to the substrate with epoxy. This approach simplified device fabrication, making it accessible without specialized equipment. It enabled the observation of quantized conductance in metallic wires. However, notched-wire MCBJ devices suffered from a small attenuation factor, typically around 10^{-2}.
Advancements in the field emerged with the adoption of lithographic techniques for MCBJ fabrication. Electron beam lithography was used to prepattern metallic contacts on a flexible substrate, often composed of steel covered with polyimide. This substrate could then be positioned within a three-point bending mechanism. In this setup, the attenuation factor was improved to approximately $(1.5 - 4) \times 10^{-5}$, providing precise control over wire elongation and gap formation with sub-angstrom precision.

Molecular Junction Formation

In the final step, molecules can be introduced into the system by attaching a liquid cell containing molecules in solution. When the wire breaks, one or more molecules can spontaneously assemble between the contacts, forming a molecular junction. This molecular junction can then be studied using electronic and optoelectronic techniques.

6.8.2 Crack Defined Break Junction(CDBJ)

Introduction

This technique offers scalability, distinguishing it from its predecessor. While we will delve into the intricacies for a single device, it's important to note that this approach can be effectively applied at a chip-level scale. Additionally, it demonstrates an intriguing fusion of this technique with Electromigration methodology.

Fabrication Process

Examining a specific section of the initial setup, we start with a Silicon substrate as the foundation. Atop this substrate, we cultivate a substantial layer of SiO2 through the process of wet oxidation. Overlaying this SiO2 layer is another stratum, approximately 100 nm thick, consisting of amorphous Silicon meticulously deposited via Chemical Vapor Deposition (CVD). This amorphous Silicon layer serves a sacrificial role in the process.

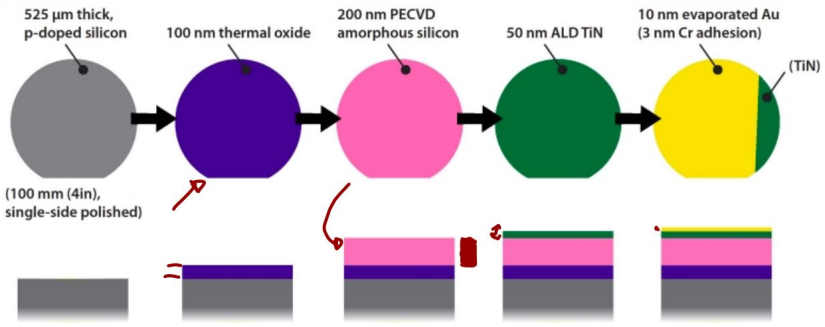


Figure 6.44: Gold Patterning top view

Above the sacrificial layer, we introduce another layer, roughly 50 nm in thickness, comprising TiN (Titanium Nitride). TiN is selected for its unique characteristics—it combines hardness with reliability in breaking. Moreover, it's deposited using Atomic Layer Deposition (ALD), which introduces a tensile stress into the lattice structure.
Proceeding beyond the TiN layer is an exceptionally thin adhesion layer formed by evaporating Chromium (Cr). Finally, we reach the Gold layer, approximately 10 nm thick, where the gap

will be created. It's worth emphasizing that everything described thus far primarily serves as a structural support system for the process.

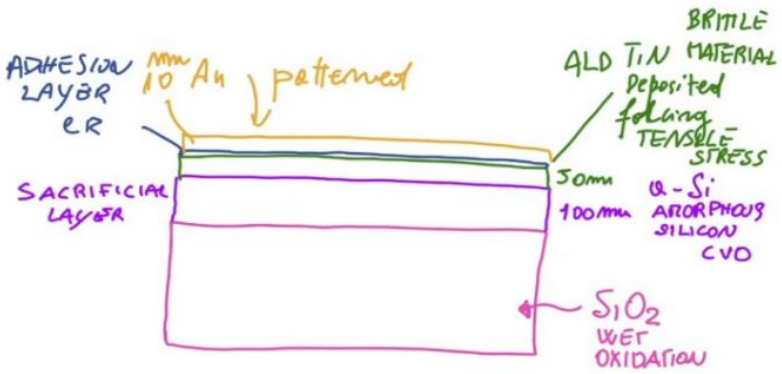


Figure 6.45: Normal vs. Undercut Etching Layers

The Gold layer, as we've seen previously, is meticulously patterned. When observed from the top view, the Gold is arranged in a specific configuration. It starts as a broader section, gradually tapering down to a thinner wire. Following this, there is a distinct notch, strategically placed where the gap is intended to be generated. This notch is then succeeded by another section of the thinner wire, which, in turn, transitions into yet another broader section. The overall length of this bridge, denoted as 'L,' holds significant importance, and this entire pattern is crafted using Photolithography.

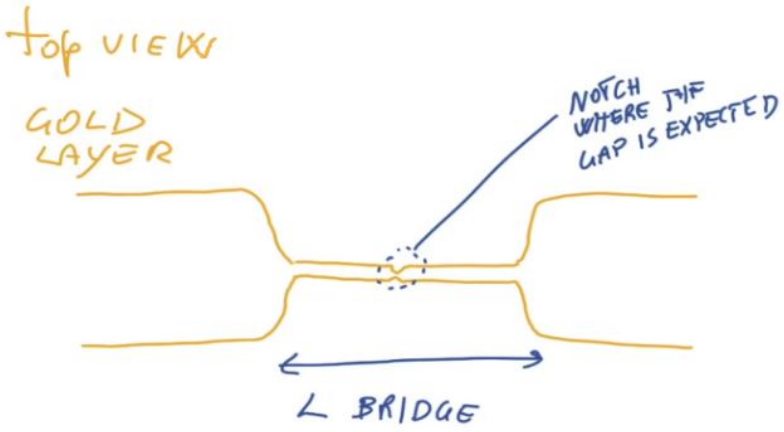


Figure 6.46: Bridget zone

Following the preparation of this element, a critical procedure called 'undercut etching' is employed. This procedure is focused on the sacrificial layer of amorphous Silicon. Through undercut etching, an additional portion of the amorphous Silicon beneath the Gold substrate is selectively removed. This differs from a standard etching process, which would only remove the portion of amorphous Silicon not associated with the areas below the broader regions and the Gold-Cr-TiN bridge.
Effectively, this means that amorphous Silicon is no longer present beneath the Gold-Cr-TiN bridge. This meticulous engineering ensures precise control over the region where the gap will ultimately be formed.

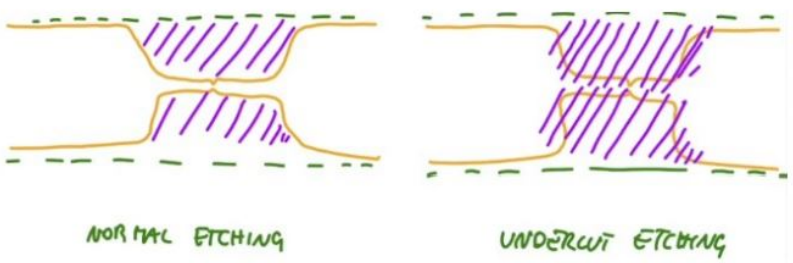


Figure 6.47: TiN Layer Behavior

After performing the undercut etching, we arrive at a new cross-sectional view. What's intriguing at this point is the behavior of the stressed TiN layer. Since it no longer has anything to anchor to, it becomes inherently fragile and eventually breaks apart. During the etching process, the void created in the amorphous Silicon layer induces increasing stress in the suspended TiN layer below the bridge until it fractures.

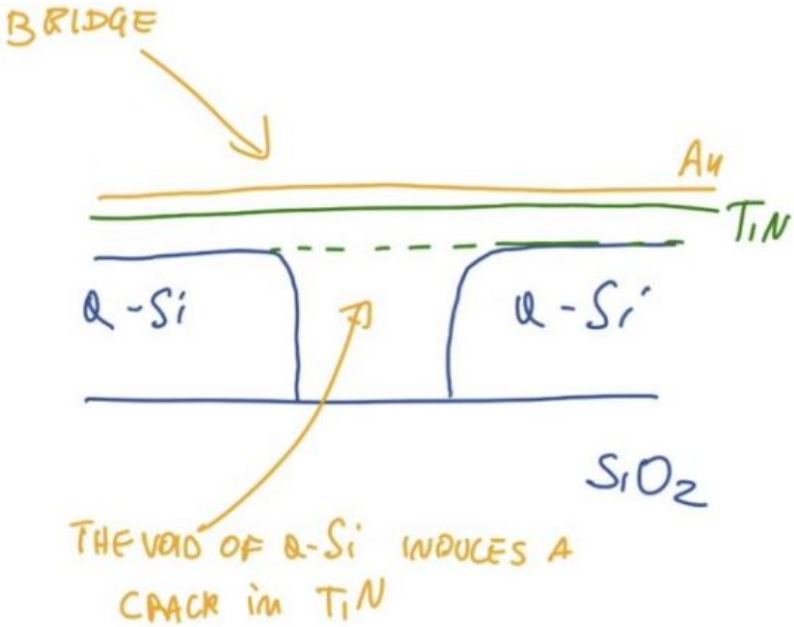


Figure 6.48: The void of a-Si induces the cracks

Upon the breaking of the TiN layer, the part of the Gold bridge with the previously mentioned notch comes into focus. Here's what unfolds: when the TiN breaks, each part of it exerts a pulling force on the Gold in opposite directions (effectively, TiN acts as a cantilever). Consequently, the part of Gold containing the notch experiences considerable strain, eventually leading to a crack formation, and thus, the creation of a nanogap. Importantly, the distance 'd' of the nanogap depends on the opening 'W' of the TiN. While there isn't a well-defined formulation, 'W' is influenced by the length of the bridge 'L' and the elastic strain ε of the TiN, which measures around $2.7\,\mathrm{nm}/\mu\mathrm{m}$.

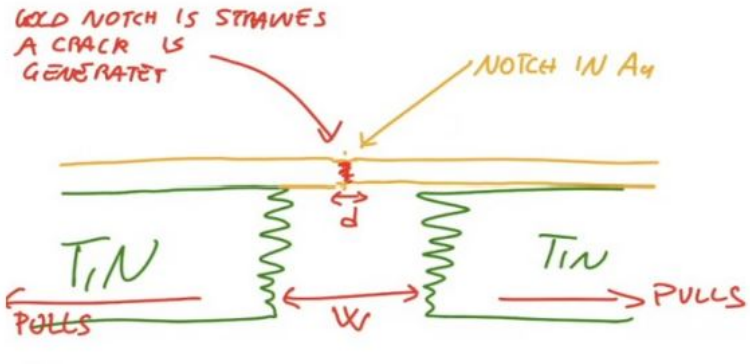


Figure 6.49: The gold notch is strained and a crack is generation

As depicted in the figure above, knowing 'L' and ε allows us to determine 'W,' and with a corresponding relation to 'd,' we could potentially predict and design the nanogap's dimensions. This is critical since it's within this distance 'd' that we will host the molecule—a significant factor given that the molecule's length significantly influences the current in the molecular transistor and overall conductance.
This technique exhibits scalability and can be applied at the chip level. In fact, the experiment has managed to generate a substantial number of bridges in parallel, approximately 780,000 on a single chip. Remarkably, over 99% of these bridges successfully produce nanogaps, with the desired gap size averaging around 3 nm, aligning with our target. However, not all generated gaps have precisely this 3 nm dimension. Depending on the relationship between 'L' and 'W' that has been established, three situations can arise:

- In cases where the bridge's length is too short ($L < 1\,\mu$m), no gap forms at all.
- Conversely, when the bridge's length is excessive ($L > 6\,\mu$m), the gap becomes too large due to the high mechanical stress induced by the TiN.
- In intermediate cases, we may have both gap formation and the presence of 'scratches' or 'ligands' within the notch. The yield of 3 nm gaps stands at around 80%. In instances where ligands are present, an Electromigration process without feedback can effectively work to break the final wire generated due to stress.

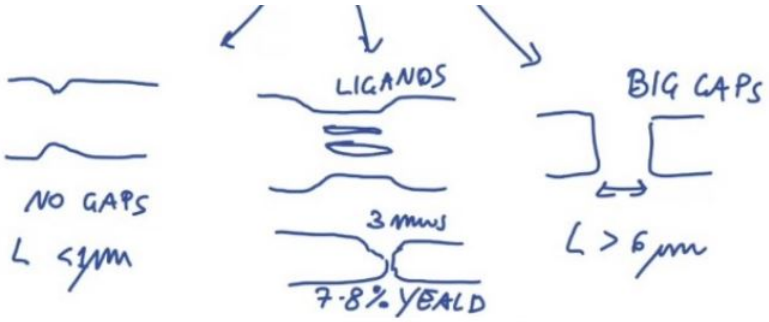


Figure 6.50: Gaps formed of the Various type

6.9 Molecular Transistor used as Sensor

There are a lot of possible sensors based on molecules, but normally in those cases, we are referring to molecular structures that utilize organic elements and molecules for sensing. In our case, we place a molecule inside a nanogap between two electrodes using anchoring groups. Optionally, we can include a gate in the setup.

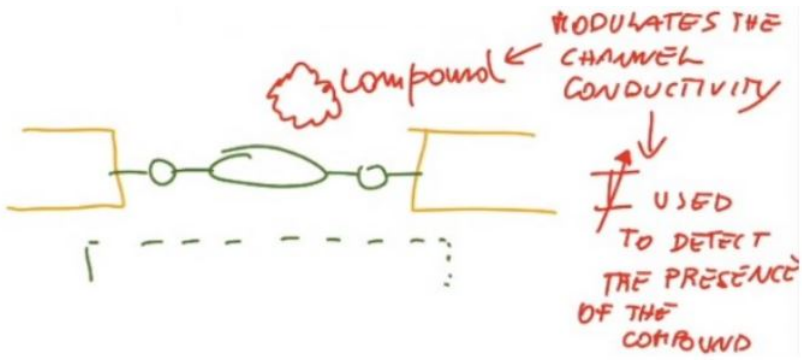


Figure 6.51: Setup of the Sensor

In this configuration, we introduce a compound that, when it is in close proximity to the molecule, modulates the channel's conductivity. As a result, we observe a change in current compared to the case without the compound. This change in current serves as an indirect method to detect the presence of the compound itself. Of course, there can be multiple elements of the same type of compound, with varying concentrations. The specific configuration that works depends on the type of molecule and the type of compound, and several variables must be considered.
Additionally, we have external forces that can influence the conductivity of our device. One of the distinguishing features of this molecular transistor is its ability to detect the presence of a specific compound.

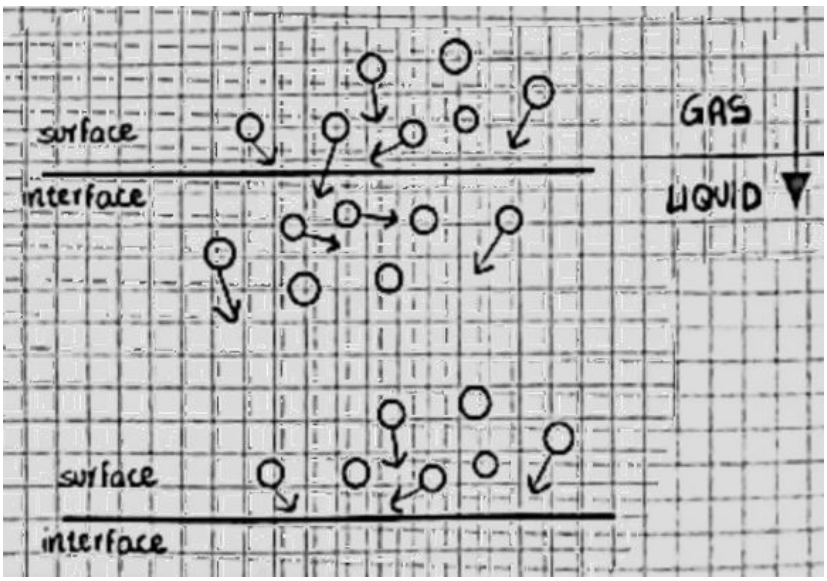


Figure 6.52: Gaps formed of the Various type

This detection is achieved by modulating the conductivity of the channel, and we observe variations in current to determine the presence of the target compound. The gate structure is sketched, and its usage is contingent on the application. For instance, it can be employed to switch the sensing function on or off with respect to the target.
It's important to distinguish between two processes:

1. **ABSORPTION:** This involves the dissolution of chemical species by a liquid or solid and occurs at interfaces between materials in different phases.
2. **ADSORPTION:** In this process, molecules, ions, or atoms adhere to a surface (adsorbent). Here, molecules make contact with the surface and are deposited in the interfacial region, leading to surface reactions, but not internal ones.

The reverse process is called **desorption**.
In the context of molecular sensors, our primary focus is on adsorption, which can be categorized into two main types:

1. **Physisorption:** This type of adsorption involves weak binding energy, typically around 300 mEV.
2. **Chemisorption:** In contrast, chemisorption involves the formation of a chemical bond. The transition between these two types of adsorption can be sharp or gradual.

Absorption processes can initiate with physisorption and then progress to chemisorption, often due to the involvement of additional forces over time. In this scenario, the binding energy is higher. These variations are illustrated in diagrams for better understanding.

6.9.1 Advantages of Molecular Sensors

Now, let's evaluate the advantages of molecular sensors:
Molecular sensors offer a distinctive advantage in that they do not require intermediary labels. In traditional sensors, especially bio-chemical sensors, a molecule is typically used to detect the target. However, there is usually a label between the target and the molecule. This label is responsible for changing its behavior in response to the presence of the target. The interaction between the label and the molecule is what helps us determine if the target is present or not.
In contrast, molecular sensors eliminate the need for an intermediary component (the label). This absence of a label provides significant advantages, primarily because it allows us to adapt the sensor to different targets, increasing its versatility.

Specific Considerations for Molecular Sensors

In the case of molecular sensors, it is essential to consider how the external compound interacts with the sensor. These interactions can be categorized into three main situations:

1. Only physisorption occurs.
2. Only chemisorption takes place.
3. Initial physisorption is followed by chemisorption, usually due to external forces.

In an ATK simulation conducted using ab initio methods, we observe gold electrodes with a fullerene molecule positioned in between. Notably, there are no direct connections between the gold electrodes and the fullerene. The focus of this simulation is on the orbital surfaces of both the gold electrodes and the fullerene, and these orbitals depend on the distance between them.

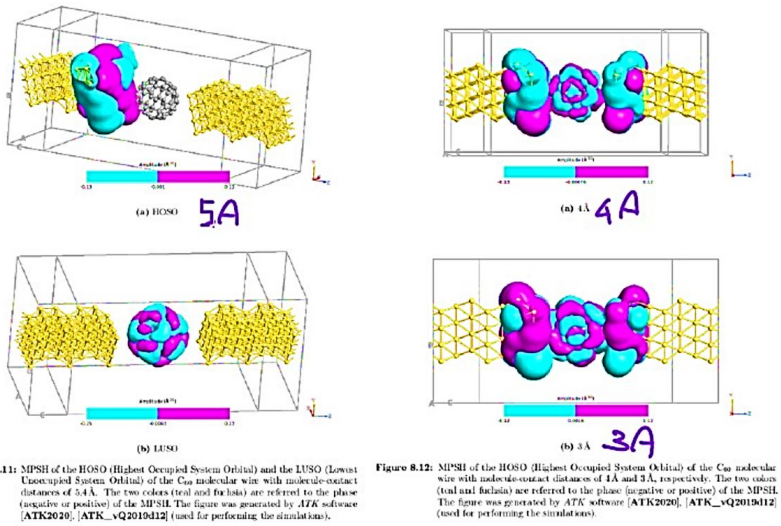


Figure 8.11: MPSH of the HOSO (Highest Occupied System Orbital) and the LUSO (Lowest Unoccupied System Orbital) of the C_{60} molecular wire with molecule-contact distances of 5.4 Å. The two colors (teal and fuchsia) are referred to the phase (negative or positive) of the MPSH. The figure was generated by *ATK* software [**ATK2020**]. [**ATK_vQ2019d12**] (used for perfoming the simulations).

Figure 8.12: MPSH of the HOSO (Highest Occupied System Orbital) of the C_{60} molecular wire with molecule-contact distances of 4 Å and 3 Å, respectively. The two colors (teal and fuchsia) are referred to the phase (negative or positive) of the MPSH. The figure was generated by *ATK* software [**ATK2020**]. [**ATK_vQ2019d12**] (used for performing the simulations).

Figure 6.53: Gaps formed of the Various type

1. At a distance of 5 Å, the orbitals are not in communication.
2. At 4 Å, there is a partial superposition of the orbitals.
3. At 3 Å, the orbitals are significantly superposed, leading to mutual interference.

This simulation can also be applied to the interaction between the external compound and what we aim to measure. In this context, the orbitals of the external compound interfere with the orbitals of our system when they come into close proximity.
To gain a better understanding of these interactions, we will explore three examples: one involving chemisorption, one involving physisorption, and one that falls in between these two categories. These examples will shed light on how these different modes of interaction work.
The absence of a label means that the sensor operates directly, without requiring an additional action to be active, unlike systems that need human intervention. For example, in a COVID test, we take a sample, and a specific solution is required to verify the presence of the COVID virus, which necessitates human involvement. In general, human intervention can be a drawback, as it can be time-consuming and challenging.
This kind of process can also be time-consuming, taking hours to yield results. This is a common problem with many sensor systems. Molecular sensors have the advantage of potentially offering high selectivity and sensitivity. However, since this technology is still in its early stages, we are in the preliminary phase and face some critical challenges.
For instance, we know that the molecular channel can be specifically chosen for the target. Furthermore, we can adjust the bias window (transmission in the nanowire) exactly where we want to detect the peak of a specific species. We can also manipulate the gate to fine-tune the transmission spectrum. This level of selectivity is absent in conventional bulk sensors. High selectivity implies that we can have significant variations in the current, on the order of 10-20%, which can be used for the final detection. The current can also be small, as measuring it is not particularly challenging nowadays, thanks to amplifiers and converters.
Molecular sensor technology offers good scalability and versatility. It can be integrated into other systems in integrated circuits (ICs) and provides very fast detection times, typically in the order of nanoseconds, microseconds, or milliseconds, which is much faster than the typical sensor systems that take hours to produce results.
Moreover, once the technology is accessed, it becomes cost-effective, and it has the potential for reversibility. In several other systems, especially those involving bulk semiconductors, achieving reversibility can be quite challenging, but it's not a default property of many sensors.

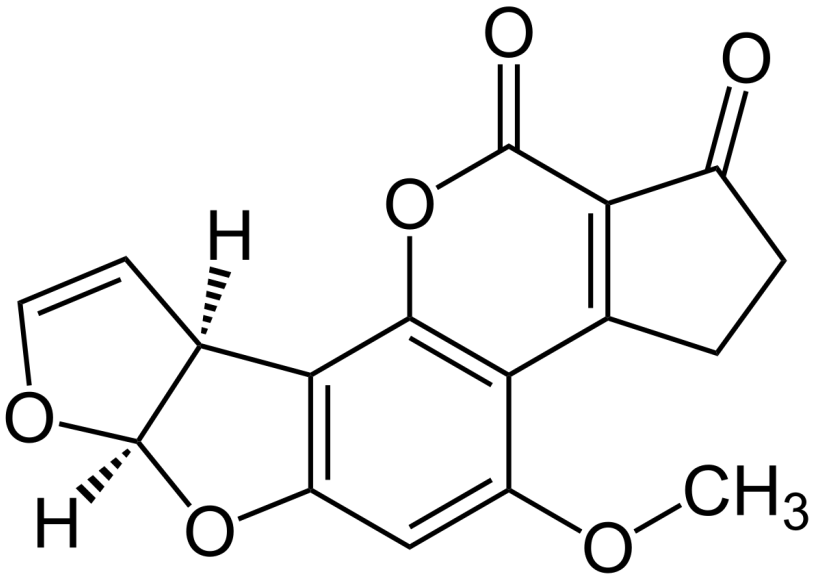


Figure 6.54: Gaps formed of the Various type

In conclusion, while molecular sensors offer promising advantages, we must also acknowledge that we are in the early stages of this technology, and there are still critical challenges to overcome.

6.9.2 Advantages of Molecular Sensors

- **No Need for Labels:** Molecular sensors operate without the need for intermediary labels. This direct approach enhances versatility by allowing changes based on the target.
- **Simplicity:** Molecular sensors eliminate the need for human intervention. This is advantageous as it avoids the time-consuming and challenging steps required in other systems.
- **Selectivity and Sensitivity:** Molecular sensors offer the potential for high selectivity and sensitivity. Customization of the molecular channel and fine-tuning of the transmission spectrum provide a level of selectivity absent in conventional bulk sensors.
- **Fast Detection:** Molecular sensor technology provides very fast detection times, typically in the order of nanoseconds, microseconds, or milliseconds. This is significantly faster than traditional sensor systems, which often take hours.
- **Cost-Effectiveness:** Once accessed, molecular sensor technology becomes cost-effective. The potential for reversibility is an added advantage, as it is not a default property of many other sensors.

6.9.3 Potential Problems with Molecular Sensors

- **Early-Stage Technology:** Molecular sensor technology is still in its infancy, and variations in detection capabilities may occur. Yield in the manufacturing process is relatively poor, leading to uncertainties about the functionality of sensor gaps.
- **Small Interaction Areas:** While small interaction areas can be beneficial for limiting interference, they may hinder broader exposure when necessary.
- **Noise:** Electromagnetic noise can affect the system, potentially interfering with accurate readings and detection.
- **Linearity:** The linearity of the system depends on its design and characteristics. While linearity simplifies the system, it may not always be desirable, as non-linear responses can provide valuable information.
- **Reliability:** Due to its early-stage development, the reliability of molecular sensor technology can be a significant challenge that needs to be addressed.

Examples of sensors commonly found in the literature include:

- **Light Sensors:** These sensors detect various aspects of light, such as intensity, wavelength, or color. They are widely used in applications like photography, ambient light control, and optical communication.
- **Bio-Chemical Sensors:** These sensors are designed to detect and measure specific biological or chemical compounds. They have critical applications in healthcare, environmental monitoring, and the detection of biomarkers.
- **Gas Sensors:** Gas sensors are used to detect and quantify the presence of specific gases in the environment. They find use in air quality monitoring, industrial safety, and gas leak detection.
- **Metal Sensors:** These sensors are designed to identify and quantify the presence of metal substances or ions. They are essential in various industries, including metallurgy, environmental monitoring, and the detection of heavy metals in water.

These sensor types have diverse applications across scientific research, industry, and daily life.

6.10 Sensor Cases Analysis

We will analyze three sensor cases:

- 1) Cyclophane-Based Molecule Sensors (Chemisorption)
  Designed to detect Cyclophane Rings (CR) through chemisorption. Offers high selectivity and sensitivity.
- 2) Fullerene-Based Gas Sensor (Physisorption)
  Utilizes physisorption to detect gases. Simplicity, but lower selectivity compared to chemisorption.
- 3) Opebase Sensor for Thymine Detection (Chemisorption)
  Designed for precise thymine detection in DNA sequencing using chemisorption.

6.10.1 Case1: cyclophane based molecular (single molecule) sensor for Cr – chemisorption

We have already observed this phenomenon in the case of logic application for the transistor. In that instance, we had a 2,2PCP (two rings connected by two bridges composed of three carbon atoms, or 3 Methylene molecules). However, we now have a 3,3PCP, as mentioned earlier (the two rings in the middle are connected by two bridges made of three carbon atoms each). The 3,3PCP is widely used, and we can adjust the number of carbon atoms in the bridges to enable various potential applications.

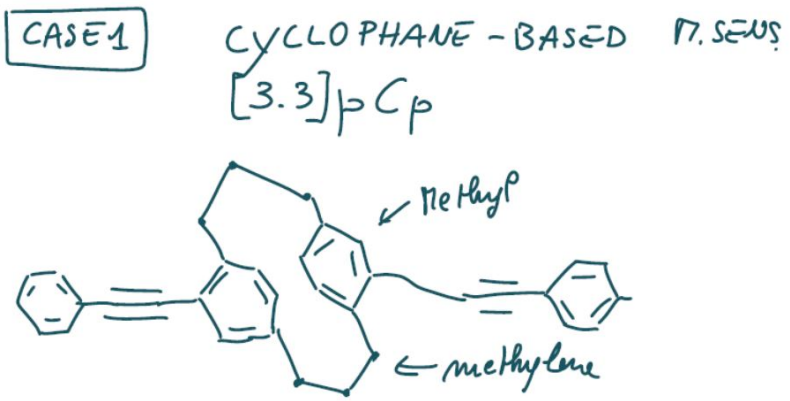


Figure 6.55: Cr – chemisorption

If we examine a cross-section of the central part, we will see one ring positioned above the other, connected by the bridges. The interior part is suitable for incorporating atoms that can be sensed, thereby creating a bond with the rings. In this case, we are considering a chromium atom (a metal atom) that will form a bond by connecting to the PZ orbitals of the two rings. The effect of chromium is to enhance the overall stability of this molecule, reducing the Highest Occupied Molecular Orbital energy level (HLG) from 5.12 eV to 3.84 eV.

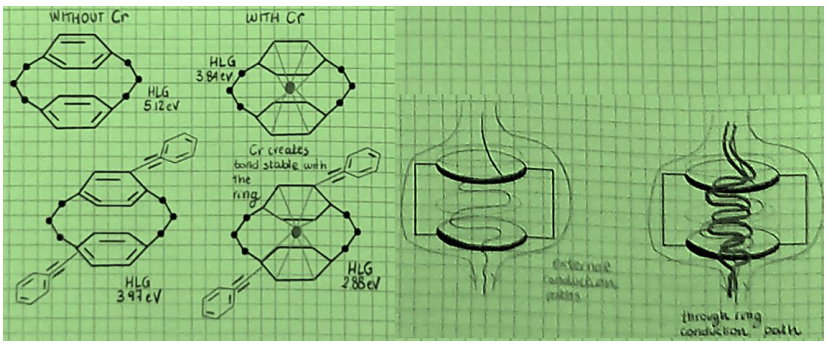


Figure 6.56: The view of the HLG for each steps

If we also account for the connections with the anchoring groups and the metal electrodes, then the HLG decreases further, from 3.97 eV to 2.88 eV when chromium is introduced. This disparity provides a means to distinguish between the two scenarios.

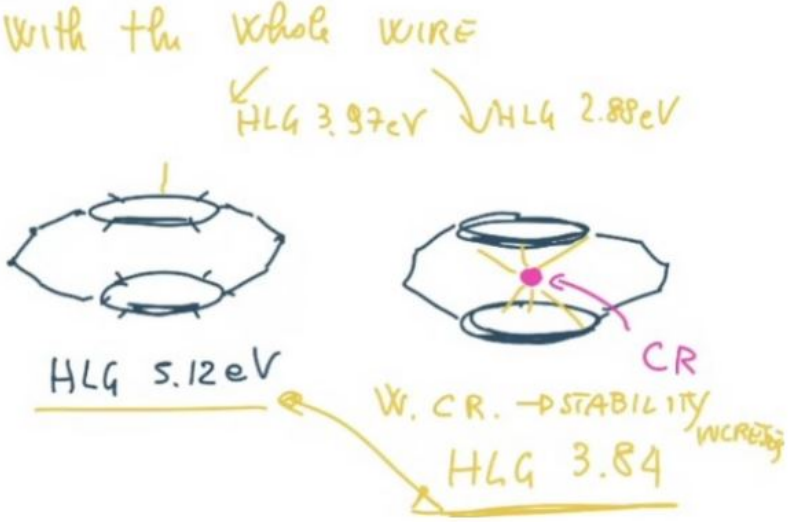


Figure 6.57: The whole HLG wire

On the right, we have a clearer view of this structure. The upper case represents the configuration without chromium, while the lower one includes chromium in the center of the central rings. Additionally, there is a structural deformation after the adsorption of chromium. The two central rings are initially separated by a specific distance due to the length of the bridges, but they draw closer after the adsorption of chromium in the middle. This alteration modifies the organization of the orbitals in the 3,3PCP molecule, as depicted in the figure above, and consequently impacts the molecule's properties.

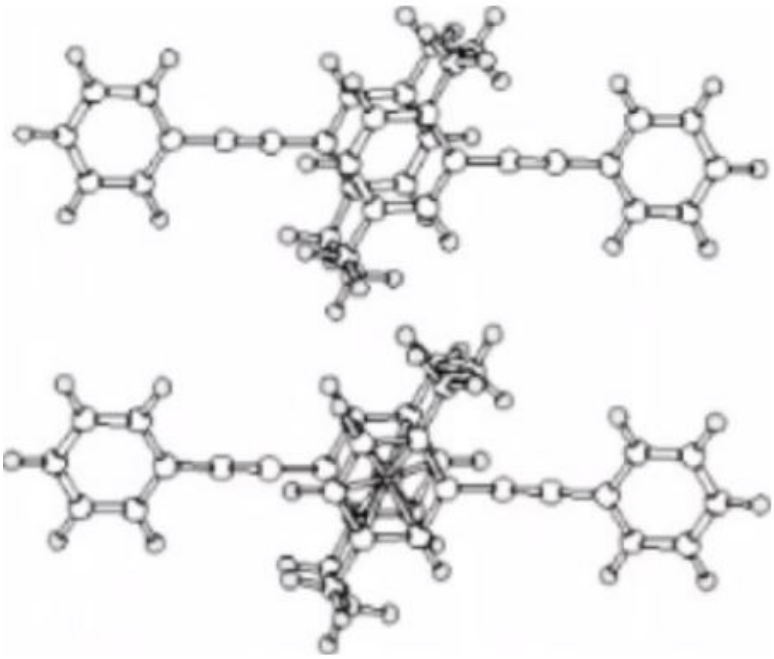


Figure 6.58: The bonding rings and connection

We use [3.3)PCP, which includes methylene bridges. We have three rings, with the middle ring substituted with two smaller rings and bridges in between. In this configuration, we have three points to consider: it increases the width of the boat-like structure, making it slightly larger and capable of accommodating certain elements in the middle.
We can distinguish:

1. A central circle
2. The entire structure

Without Cr (Chromium):
Here, we have a front view of the molecular arrangement. Cr forms bonds with the carbon atoms, sharing 6 electrons. This alters the previous double bonds. Overall, the combination of Cr and C creates an 18-electron stable system.

$$\begin{aligned} \text{HLG without Cr:} &\quad 5.12\,\text{eV} \\ \text{HLG with Cr (HOMO):} &\quad -5.73\,\text{eV} \\ \text{HLG with Cr (LUMO):} &\quad -0.61\,\text{eV} \end{aligned}$$

In the figure on the left, you can see a sketch of the structure with an empty central circle and the case where an atom of Chromium is present in the middle.
With and without Cr:
Here are the values of HLG for different structures, considering only the central elements and the entire structure:

$$\begin{aligned} \text{HLG without Cr:} &\quad 5.12\,\text{eV} \\ \text{HLG with Cr (HOMO):} &\quad -5.73\,\text{eV} \\ \text{HLG with Cr (LUMO):} &\quad -0.61\,\text{eV} \end{aligned}$$

The presence of Cr clearly reduces the HLG, indicating a significant change. Cr forms stable bonds with the structure, resulting in more charges and higher energy involved in the overall system.

HLG without Cr: 3.91 eV
HLG with Cr: 2.88 eV

The introduction of Cr leads to a substantial reduction in HLG, demonstrating a distinct change in the system's properties.

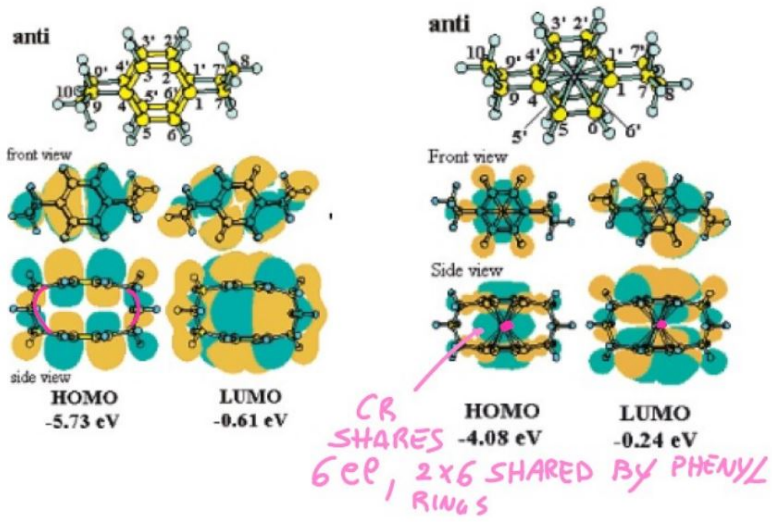


Figure 6.59: The HOMO and LUMO views front

The sensing principle is still not completely understood. However, there are still two currents contributing to conduction. The first one accounts for the external path of the ring, whereas the other accounts for the internal path. When chromium (Cr) is introduced, there are many more paths inside the rings, increasing the contribution of current. As a result, the transmission coefficient T(E) changes, and consequently, the current will change. In this case, we do not have data regarding quantum interference with or without Cr through the rings. Cr increases the paths through the rings, affecting T(E) changes.

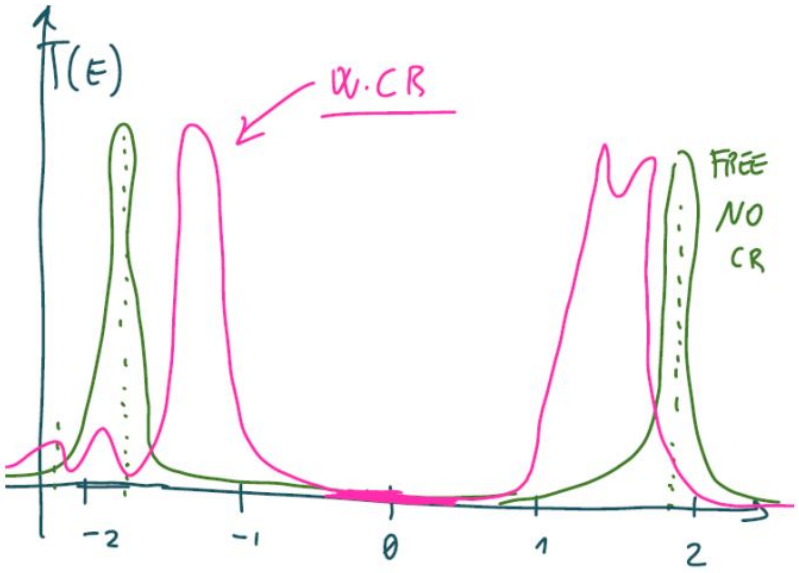


Figure 6.60: T (E) of the whole wire

Here is a sketch of the transmission spectrum. The Highest Occupied Molecular Orbital energy level (HLG) is reduced, leading to a smaller expected current. The two peaks are closer and larger, resulting in a change in current from 0.05 A to 0.3 A. This represents an increase of almost one order of magnitude. Experiments have been performed, and the table below reports the measured conductance values.

Condition	**Conductance (µS)**
Without Cr	0.015
With Cr	0.1

What is not clear in this kind of sensor is what happens when we create several sensors. This raises questions about repeatability and the interactions that can occur between different sensors. A single sensor appears to work properly; however, when multiple sensors are placed together, there is a possibility that they may not function properly in unison. Additionally, it's important to note that not all targets can be reliably detected.

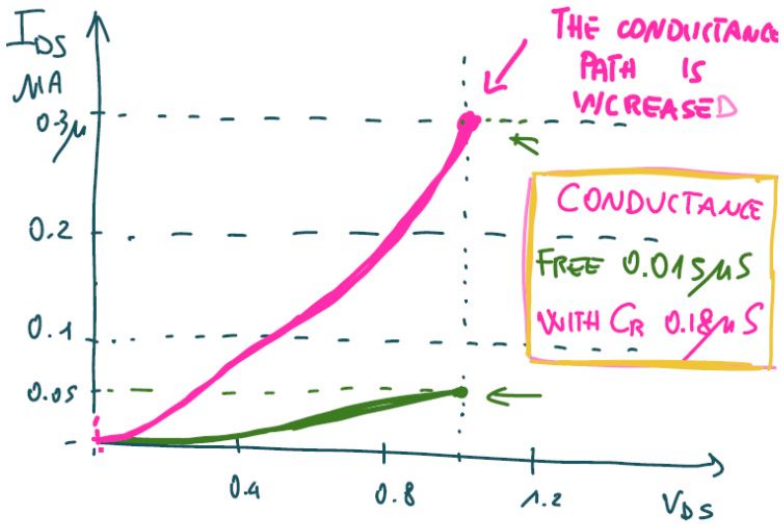


Figure 6.61: I_{DS} and V_{DS} characteristics

Here we have the current in two cases. In the free case, we observe a very small current with a maximum of around 0.05 µA, while in the case of Cr (Chromium), the current reaches a maximum of around 0.3 µA. This demonstrates a significant increase in the conduction between the central rings.
Additionally, we have reported the conductance in both cases, and their ratio is approximately 12. This is highly promising because it suggests that we can use this sensor in a circuit where we leverage the difference in the two currents to determine the presence or absence of the Cr atom. For instance, we can envision creating a current with this transistor after adsorption, transforming it into a voltage, and then using it to control another transistor. By organizing the system properly, the second transistor can be switched on or off depending on whether the Cr atom is present, essentially serving as a simple analog-to-digital converter (ADC).
However, the most straightforward way to utilize this sensor is to measure a voltage difference across a resistance where the current flows, and then use this resulting voltage drop directly outside the sensor. These are examples of the practical applications enabled by this sensor.

6.10.2 Case 2: Gas Molecular Sensor Based on Fullerene – Physisorption

There is a full family of structures like this, where there are $2 \cdot (10 + N)$ carbon atoms, and N represents the number of hexagons. In our particular case, the Fullerene is called C60 because $N = 20$, making a total of 60 carbon atoms. The structure is extremely regular and possesses various symmetries.

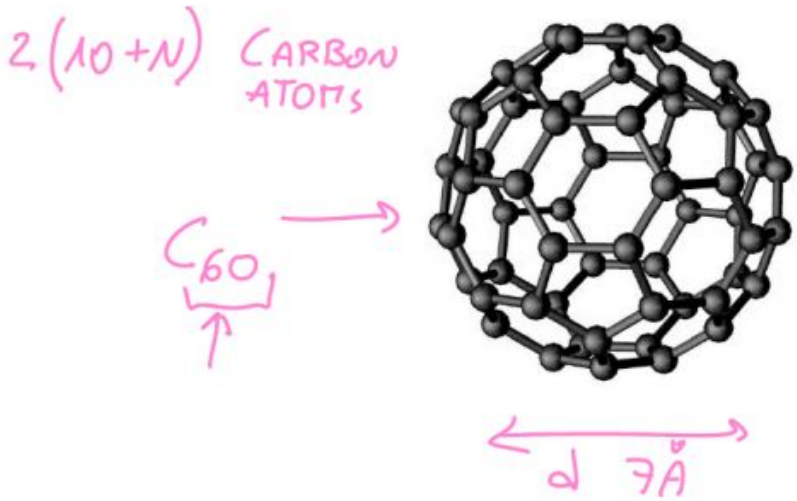


Figure 6.62: the figure of the family structure

Its diameter is approximately 7 Å. This is a very stable system, which is why the Fullerene, unlike what we've seen so far, does not require anchoring groups. It remains between the electrodes due to Van Der Waals forces. However, as we've discussed, if the distance between the two electrodes is small enough, coupling remains strong.
This means that depending on the distance d, we could have weak or strong coupling for this system, as discussed in Chapter 2. In the case of a sensor, we need something in between these extremes, not too weak but not as strong as when using anchoring groups.

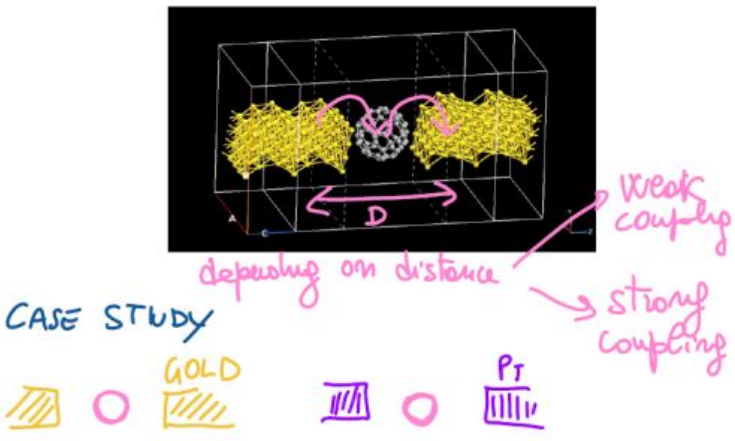


Figure 6.63: The Decoupling and the Distance

The examples we'll see in the following relate to the possibility of using gold or platinum electrodes for our system. The Fullerene is sandwiched between them, and this will help us understand the impact of the different types of electrodes on the coupling. It's not just a matter of distance.

Here we have the transmission spectra of the system for the cases in which we use Au or Pt for both contacts. As we can see, the system using gold is very close to weak coupling since the broadening is minimal, with deltas more significant than Lorentzian distributions. In the case where we used platinum instead, the broadening is much more pronounced, indicating strong coupling.

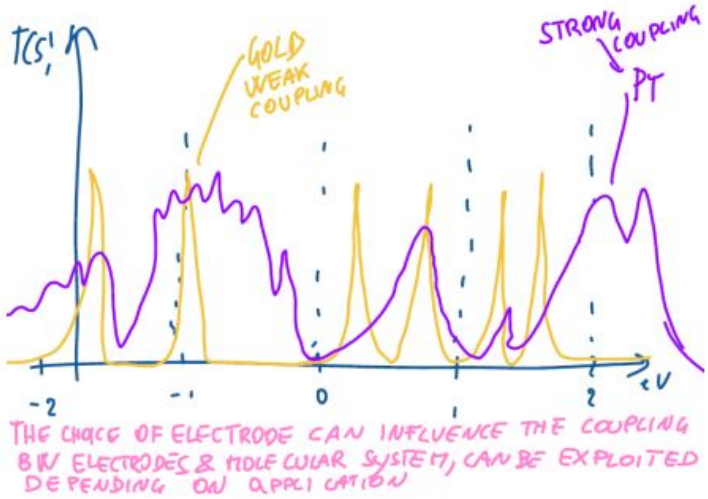


Figure 6.64: The Impact of Electrode Selection on Coupling Between Electrodes and Molecular Systems

One isn't necessarily better than the other; it depends on what we want to achieve. Having a very clear High-Level Gain (HLG) as in the transmission spectrum of the weakly coupled system could be very useful. In contrast, in the strongly coupled case, we have a much higher current. In terms of adsorption, we could say that, concerning the Fullerene within the nanogap (not the compound, which is currently absent), in the case of gold, we have something similar to physisorption, whereas with platinum, it resembles chemisorption. Therefore, the choice of the type of metal can impact the conduction properties, and the choice depends on the application.

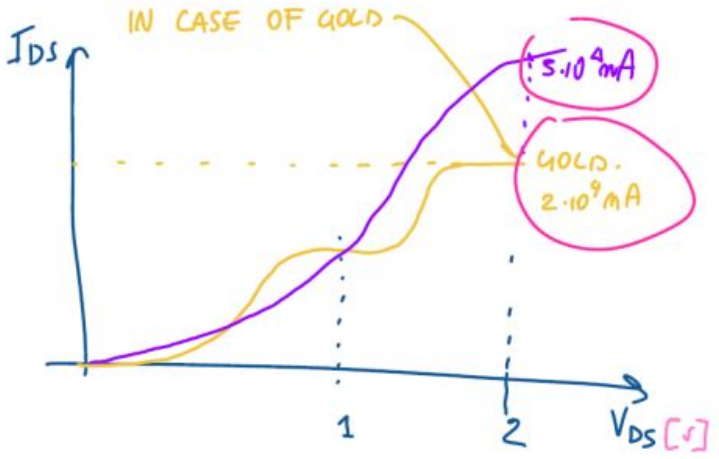


Figure 6.65: Comparing IDS Characteristics: Gold vs. Platinum Electrodes in Voltage Range [0, 2] V characteristics

The IDS in these two cases is sketched here. Of course, the current using Au is less than that using Pt and displays a step-like behavior, similar to when we have used deltas as the Density of States (DOS). The maximum current is around 2×10^4 nA within the voltage range of $[0, 2]$ V

for VDS. Using Pt, we'll have a smoother and larger current, with 5×10^4 nA as the maximum value in the same voltage range.

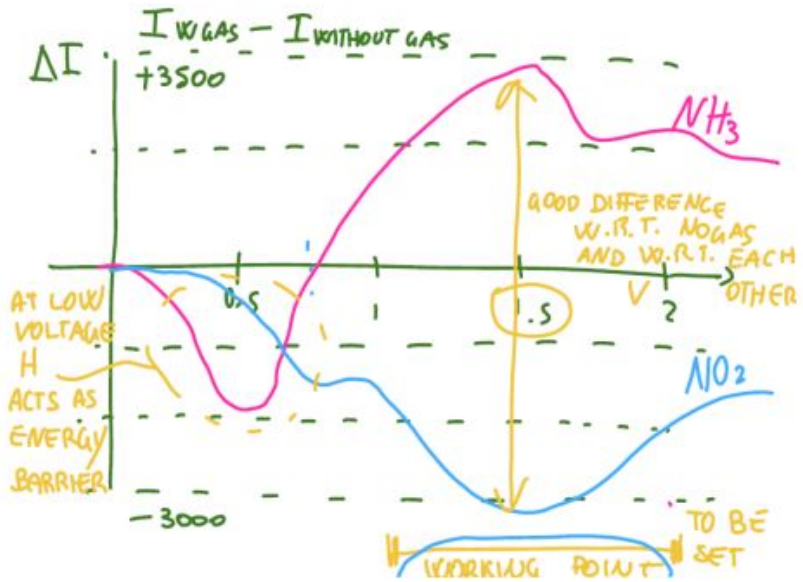


Figure 6.66: using Au as metallic contacts

If we want to increase the current, Pt seems better, but if we want to enhance selectivity (the capability of our sensor to detect the presence of an external compound), then this may not be the best solution. It depends.

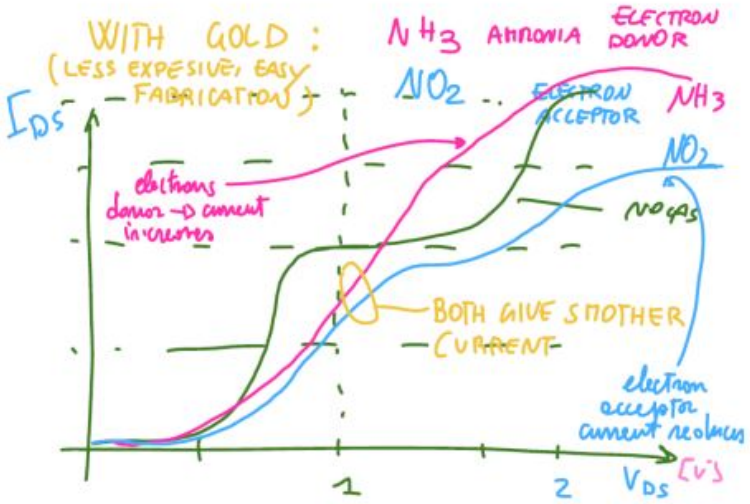


Figure 6.67: I_{DS} and V_{DS} characteristics

In the rest of the example, we'll choose the gold case, as it is easier to fabricate and less expensive than platinum. Using Au as metallic contacts, let's see what happens when different species are introduced in the presence of C_{60}.
In this example, we have two gas species: NH3 (Ammonia) and NO2 (Nitrogen dioxide). Both are common gases in real life and can be potentially hazardous. The key difference between them is that NO2 is an electron acceptor due to the presence of oxygen, while NH3 is an electron donor due to the presence of hydrogen.
First, we've depicted the current in the case without any compound, as we've seen before. When NH3 is introduced, the current increases because this compound donates electrons to our system. In the case of NO2, the current decreases as it accepts electrons from our system.

The crucial point is that this variation in current doesn't occur over the entire voltage range. To maximize sensitivity, we need to engineer the sensor to operate within this range. Here, we observe the relative differences between the NH3 and the NO_GAS cases and between the NO2 and the NO_GAS cases, making it much clearer to detect variations.
First of all, it's worth noting that the position around 1.5 V is where we observe the maximum difference in this range, and it's where we should set the working point of our sensor. Not only do we have the maximum variations compared to the case without gas, but we can also differentiate between the two gases, indicating good selectivity.
Additionally, there's an interesting aspect around 0.5 V. The explanation is that hydrogen, at low voltages, creates a kind of 'barrier' and later amplifies the electron donation. In this case, the correct working point for detection is the same for both gases, which is a fortunate situation.

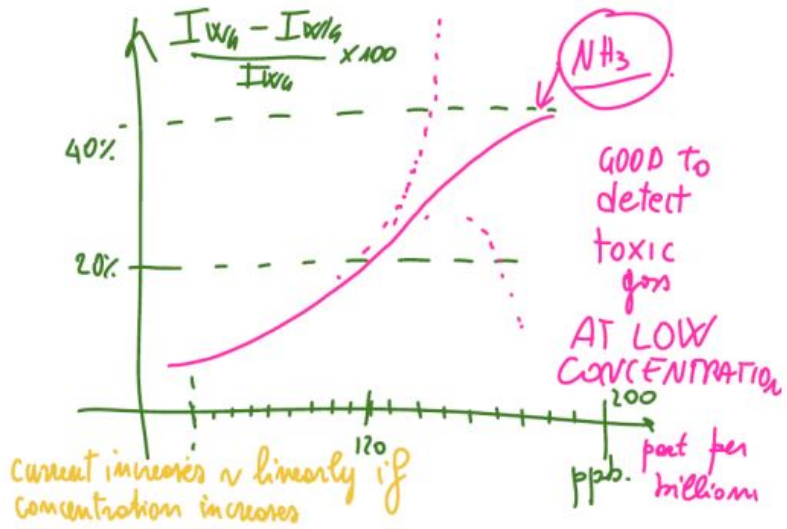


Figure 6.68: Species characteristics: $100 \times \left(IWG - \frac{IW}{G}\right) / IWG$

There's also an intriguing aspect in terms of the number of external compounds present and measured in parts per billion. Here, what is measured is the relative variation of the current with a specific species, calculated as $100 \times \frac{IWG - IW/G}{IWG}$.
What happens is that there is a quasi-linear behavior with an increasing concentration in the case of NH3. This means the current increases with the rising concentration of NH3 compound. This behavior could be useful, for example, to trigger an alarm if the concentration exceeds a predefined threshold, providing information based on current measurements. This demonstrates that C60 could be a good element for detecting toxic gases at low concentrations, even lower than what is typically encountered in gas sensors.

Case 2: Gas Sensor $PbSO_4$

In an effort to comprehensively understand the intricate changes in transmission spectra when a compound approaches a molecule, we present an in-depth simulation featuring a C60-based sensor designed for PbSO4. This study delves into the auto-states of the quantum transmission operator, visualized as isosurfaces that delineate regions in space where electron transmission can take place, as previously discussed in relation to $T(E)$.

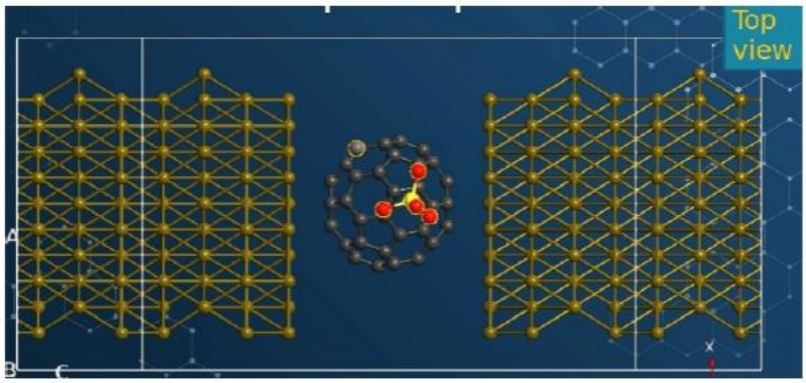


Figure 6.69: $PbSO_4$ is an ionic compound.

A standout observation within this simulation is the emergence of a distinctive electron pathway denoted by a prominent blue isosurface. This pathway owes its existence to a complex interplay of various elements: the gold electrode, the C60 molecule, and the compound PbSO4. In the absence of the compound, the molecule and electrode would remain in close proximity but fail to establish the necessary electrical connection. While electron transmission is theoretically possible without the compound, its efficiency is notably diminished, resembling a tunneling process with limited effectiveness.

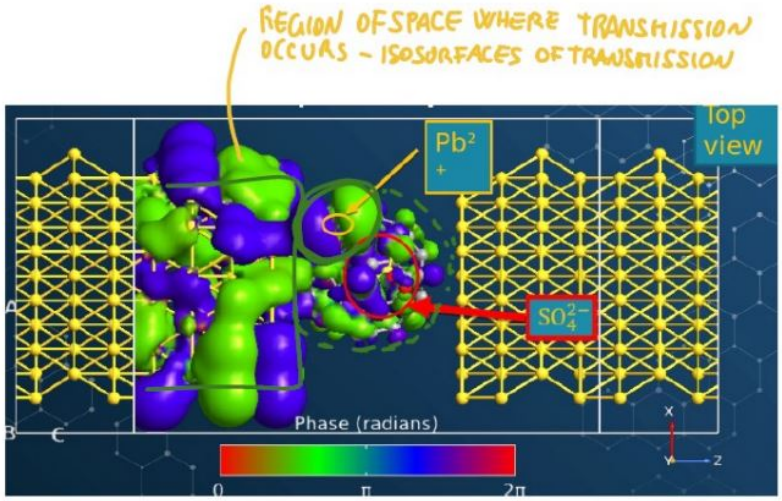


Figure 6.70: Isosurfaces of Transmission: Representing Regions of Space Where Transmission Occurs

This simulation provides a comprehensive and detailed perspective on how compound interactions dynamically shape transmission spectra. It underscores the pivotal role of synergy between compounds, molecules, and electrodes in enhancing electron transport efficiency. The intricate interplay of these components, as elucidated in this study, offers a nuanced understanding of the changes in transmission spectra, which is invaluable for the design and optimization of molecular-scale electronic devices.

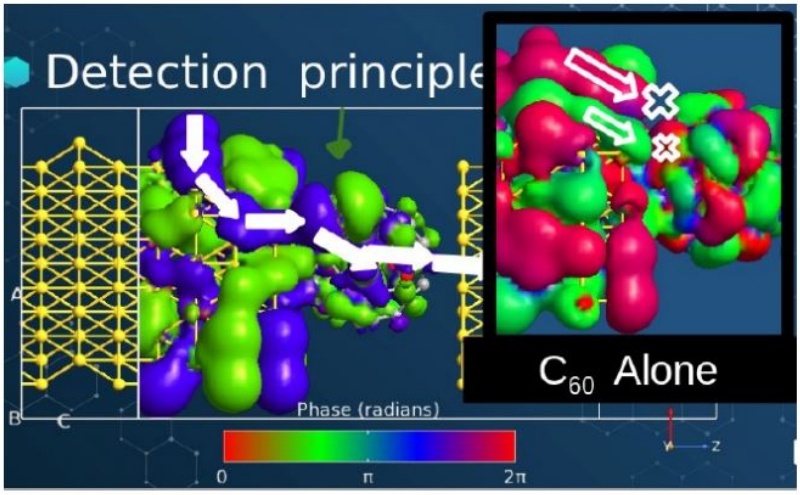


Figure 6.71: Transformation and Modulation of Transmission Path due to Interaction

Impact on $T(E)$: Linear Variation at 1 V In the case highlighted in orange, we observe larger peaks in the transmission function $T(E)$, resulting in significant differences, even with a low V_{DS}. However, in other regions, such as between -1 and 0.5, there are fewer discernible distinctions. This emphasizes the importance of carefully considering the bandwidth (BW).
To selectively alter the region of interest, we must adjust the bias windows. This flexibility is particularly valuable when studying gate activity, as it allows us to explore various parts of the transmission spectrum.

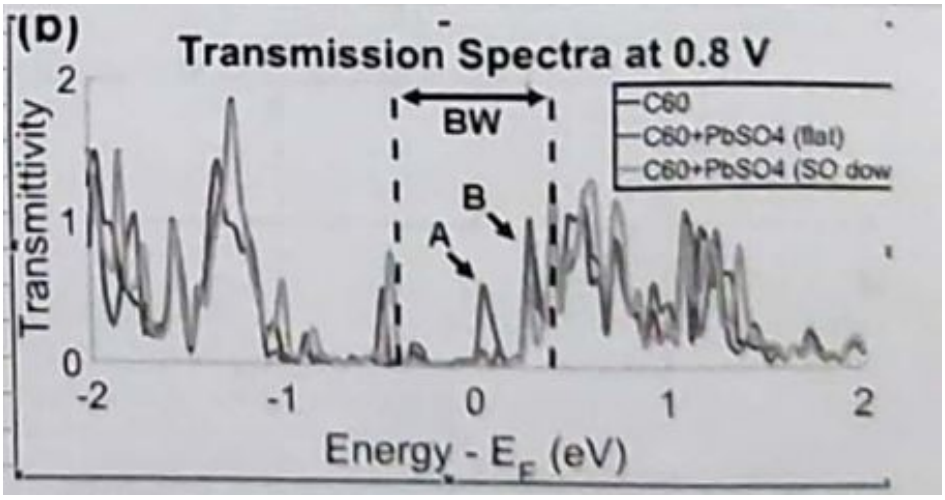


Figure 6.72: Transmission Spectra at 0.8 V

In the figure, you can see the transmission spectra at 0.8 V. The bandwidth (BW) and bias voltage (V_{DS}) play crucial roles in the observed behaviors. Different current behaviors are evident for various pollutants present in the air. Relevant variations are observed in some cases, while they are less pronounced in others. The selectivity with respect to other gases is reasonably good, and clear current variations occur at around 1 V, although this behavior is not consistent across all scenarios.

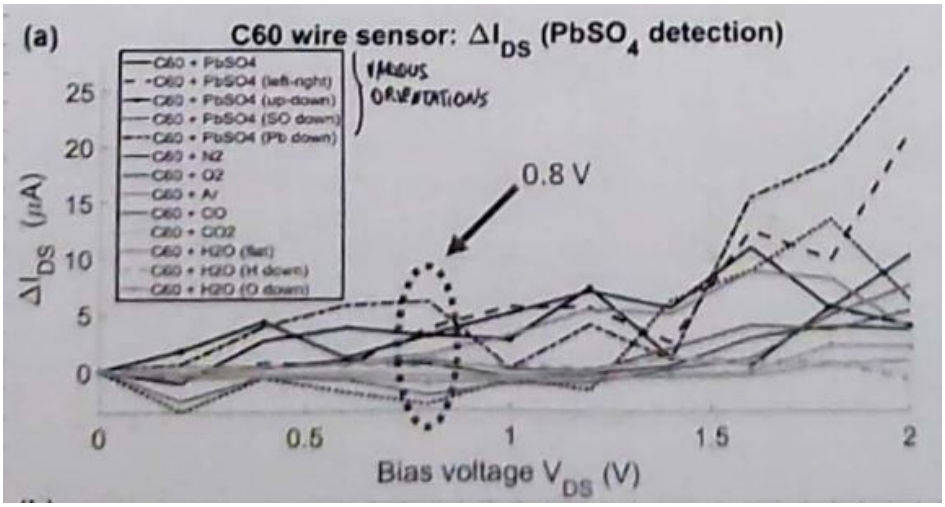


Figure 6.73: Selectivity with Respect to Other Gases in Air at 0.7 A

The selectivity with respect to other gases in the air is assessed, with a clear selectivity value presented at 0.7 A, illustrating the varying sensitivity to different gases.
Case of NO with Fullerene (a) Without NO Target (b) With NO Target

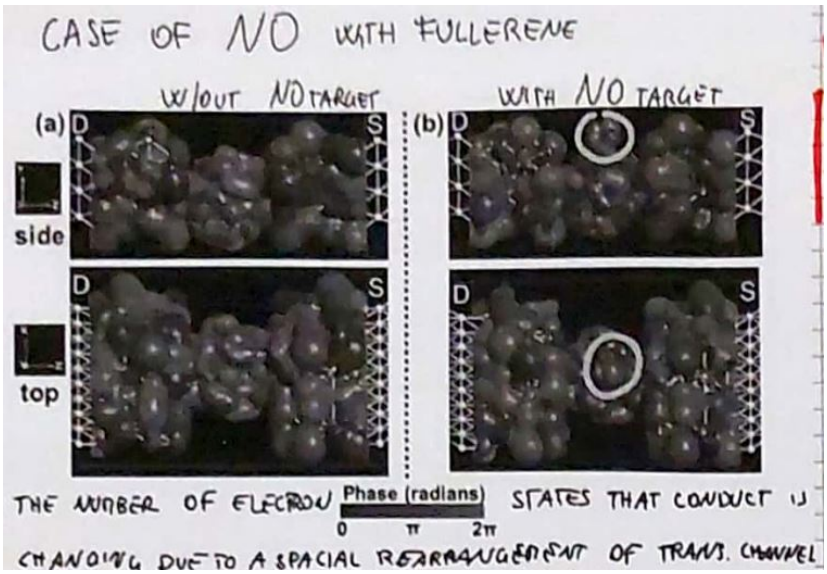


Figure 6.74: The Number of Electron States that Conduct

In the figure, the number of electron states conducting is shown, indicating a change of 2π radians due to a spatial rearrangement of the transmission channel.
Impact on $T(E)$: Clear Variation at 1 V Transmission spectra at 1 V are presented, showing the impact on $T(E)$ with a noticeable variation.

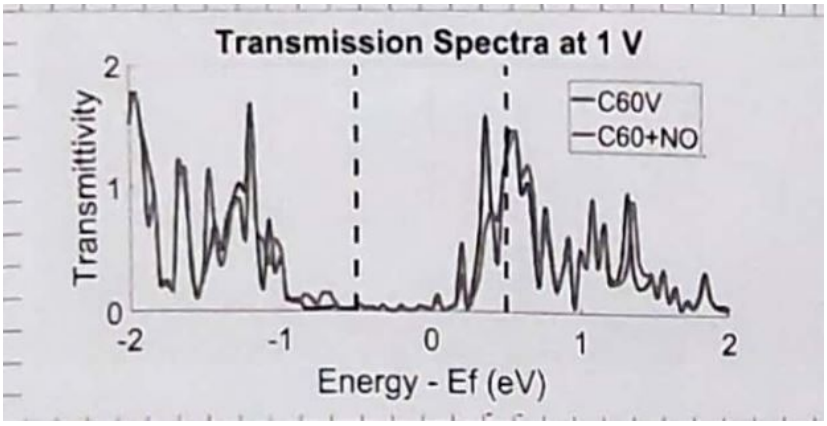


Figure 6.75: Transmission Spectra at 1 V

In other V_{DS} points, the transmission function $T(E)$ exhibits smaller differences, highlighting a particularly significant point.

Same Structure, Different Gas: NO (Nitrogen Oxide) with Fullerene

Here, we explore the case of NO (Nitrogen Oxide) with fullerene. The top and side sections of the case are depicted, demonstrating that there are more paths for electrons to move due to a special rearrangement of the transmission channel. It's important to note that this behavior is not consistent across all V_{DS} points, and specific selectivity is observed only in certain bias windows, which is not the case in bulk sensors.

Impact on Current In the figure, we observe the impact on current, specifically for a C60 wire sensor using the EHT method for NO sensing. Clear sensitivity is evident at 1 V.

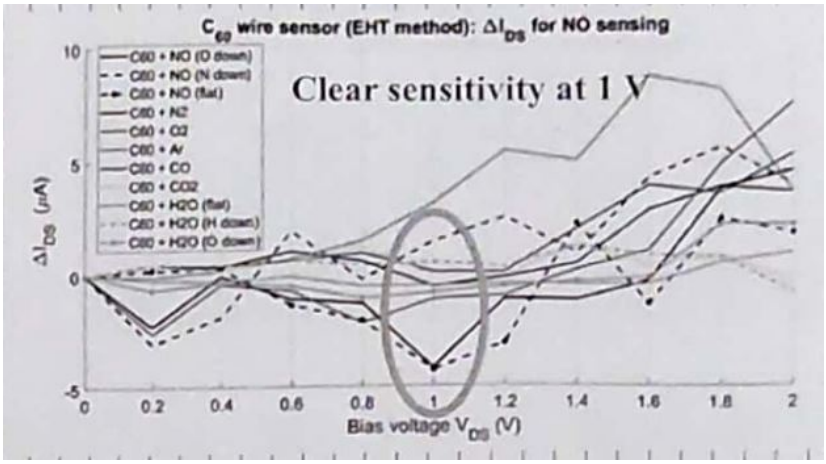


Figure 6.76: Impact on Current

6.10.3 Case 3: OPE-Based Molecular Sensor for Thymine Detection in DNA Sequencing – Chemisorption

In this example, we delve into the capture of a bio-molecule by a sensor, a process that involves various experiments. The specific application here is DNA sequencing.

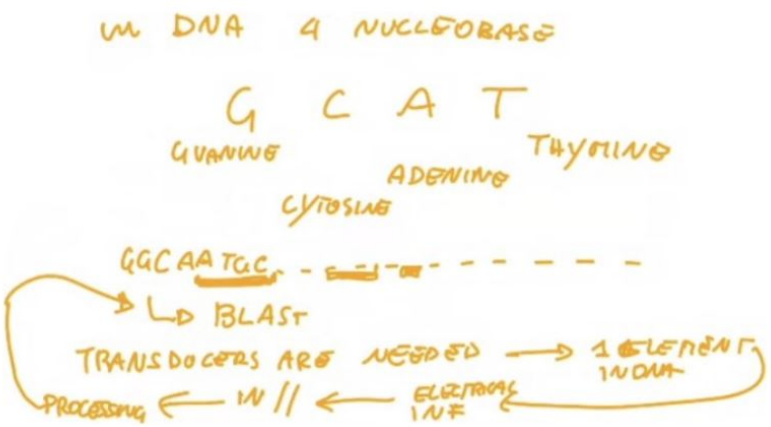


Figure 6.77: The last kind of sensor we see is for bio-application

The final type of sensor we will explore is designed for bio-applications, specifically to detect Thymine in order to optimize DNA sequencing. In DNA, there are four fundamental nucleotides: Guanine, Cytosine, Adenine, and Thymine. DNA consists of long sequences of these basic elements, and any deviations from a reference sample (often referred to as a "healthy" sample) indicate mutations, which can lead to various problems or diseases.
These mutations can occur at various positions along the DNA sequence. When researchers search for mutations in DNA, they typically compare the sample to a reference sample using computationally intensive algorithms such as BLAST. These algorithms rely on data extracted from what is known as "sequencing," which provides information about the DNA sequence related to a specific sample, facilitating the comparison with the reference sequence.
To perform sequencing on a sample, efficient transducers are required to translate each nucleotide into electrical information. Performing this process for all nucleotides in parallel with multiple transducers enables high-speed sequencing.
The concept is to use very small transducers based on molecular sensors capable of generating information about a part of the DNA, creating a compact and fast element that can be organized in parallel to achieve efficient sequencing. Now, let's delve into the design of a single sensor.

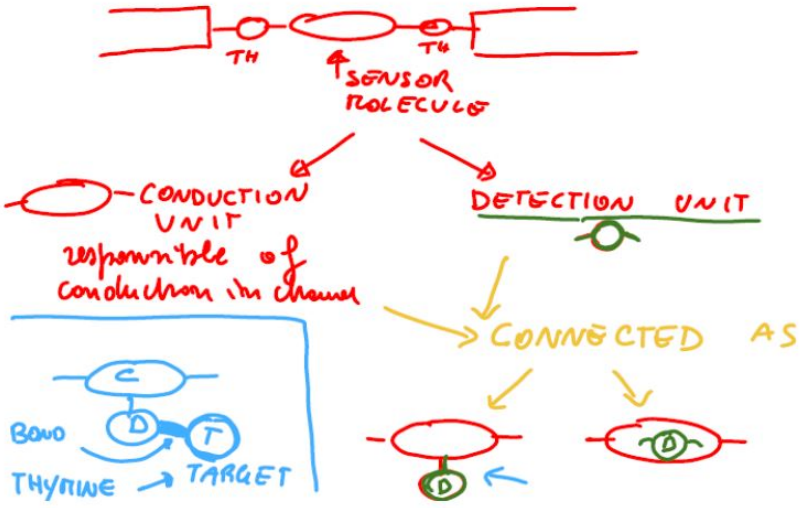


Figure 6.78: the initial situation on the detection unit

In the initial configuration described in the previous slide, the gate effect is not present. Instead, the system consists of metallic contacts, anchoring groups (such as Thiol in this example), and the molecule in between. The molecule is divided into two parts: the conduction unit responsible

for electron transport in the channel, and the "detection unit," a sub-part connected to the conduction unit. The detection unit can be positioned in two ways: it can be outside the conduction unit or embedded within it, depending on the molecule used as the conduction unit. Both the conduction unit and the detection unit are connected to the target, which in this case is Thymine. This subdivision into sub-units is necessary because conduction units are typically chains of molecules with good conductive properties, but they may not readily connect to a target, especially a biological one. To facilitate this connection, a strong chemical bond is required, as opposed to simple physisorption, as observed in gas sensors.
During the sensing process, a bonding occurs, and ideally, desorption should take place afterward. Desorption is necessary for multiple detections using the same detection unit. After measurement, cleaning the sensor involves breaking the chemical bond between the detection unit and the biological molecule to enable sensor reusability. The desorption process and the type of bonding must be considered and designed carefully.

Molecular Structure of the Sensor

In this specific case, the conduction unit is a three-ring OPE molecule, while the detection unit consists of a 2-AminoPyrimidine molecule. To understand the detection unit better, let's examine the structure of Pyrimidine.
Pyrimidine is a ring with two nitrogen atoms connected to hydrogen. This constitutes the fundamental structure of our detection unit. What varies is that the carbon atom located between the two nitrogen atoms is linked to another segment consisting of a single nitrogen atom connected to two hydrogen atoms. This unique structure forms our detection unit, which will be connected to the conduction unit.

Function and Structure of the Sensor

The conduction unit and the detection unit have distinct functions. Additionally, the target biomolecule plays a significant role in the sensor's operation. In many cases, an assembly structure is employed. By altering the detection unit while keeping it connected to the same conduction unit, the sensor's adaptability to different targets is facilitated. In the example provided, Thymine serves as the target. In this particular case, a chemisorption bonding occurs between the detection unit and Thymine. This is just one illustrative instance of sensor operation.
The conduction unit, as previously mentioned, is OPE, which features three rings in the theoretical study. In experimental settings, three rings are often insufficient, necessitating modifications.

Details of the Detection Unit

The detection unit is primarily based on Pyrimidine, specifically 4-AminoPyrimidine. It is an aromatic element that replaces two carbon atoms in the ring with two nitrogen atoms. The diagrams provided depict different representations of the same molecule. One of the Pyrimidine rings is substituted with an NH group to create the final detection unit, 2-AminoPyrimidine.

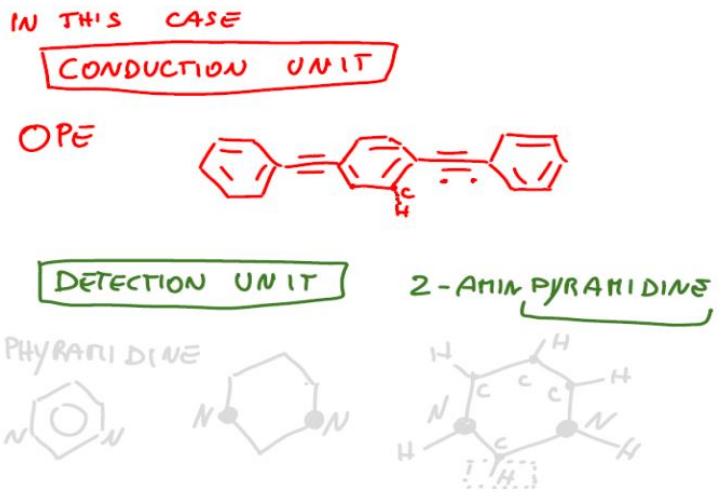


Figure 6.79: Structure of 2-AminoPyrimidine Detection Unit

The detection unit is connected at certain points to both the conduction unit and the target. Regarding the connection to the conduction unit, it links through the nitrogen (N) located at the intermediate link of the OPE molecule to a carbon atom (C).

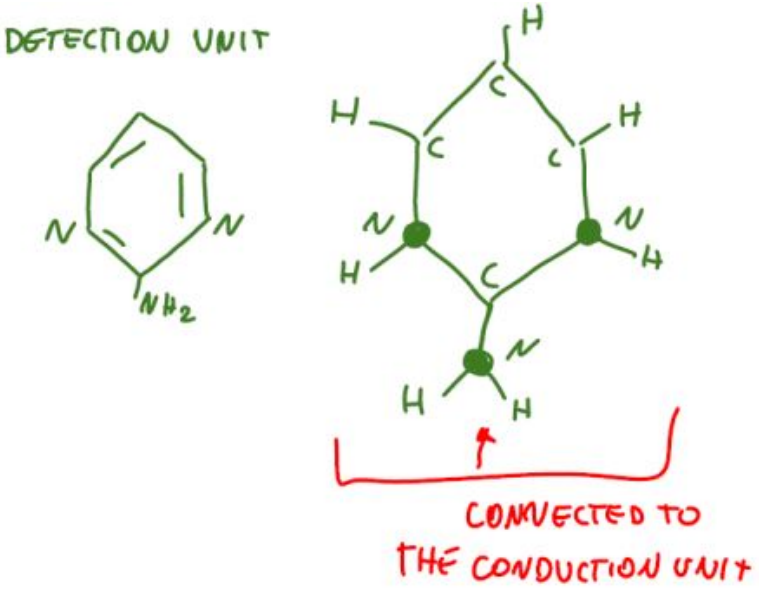


Figure 6.80: Connection of 2-AminoPyrimidine to Conduction Unit

The complete sensor molecule consists of both the conduction and detection parts. In this molecular structure, a carbon atom of the central ring in the conduction unit connects to the nitrogen atom that we introduced in the Pyrimidine.
It's worth noting that both the conduction and detection units feature π bonds, forming a conjugated system throughout the structure. Furthermore, both subparts are situated in the same plane, resulting in a planar structure. This planar arrangement implies the presence of delocalized states both above and below the molecule. As a consequence, electrons can flow through various paths from source to drain.

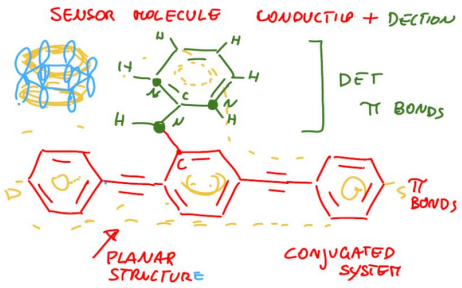


Figure 6.81: Structure of the Sensor Molecule

Target: Thymine Molecule
Our target is the Thymine molecule, which is derived from Pyrimidine. Thymine features two oxygen atoms instead of hydrogen, and one carbon atom in the ring is connected to another carbon with three hydrogen atoms. Our detection unit's role is to capture this target, and this is achieved through bonding between these two components, as depicted on the right side of the molecule above.

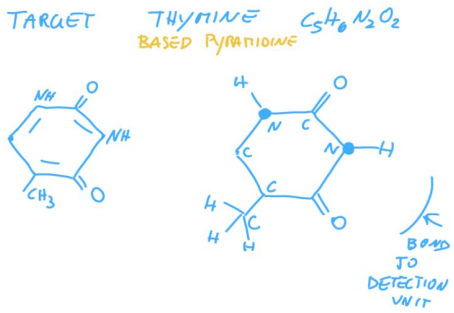


Figure 6.82: Structure of Thymine Molecule

This is the overall structure that our sensor will exhibit when it detects Thymine. In this structure, hydrogen bonding occurs between the target (Thymine) and the detection unit. The key concept here is that when we introduce the sensor molecule and the target into a solution, the similarity between the detection unit and the target promotes the breaking of the previously mentioned bonding and the formation of hydrogen bonding. While hydrogen bonding is not particularly strong, it proves effective in this context.

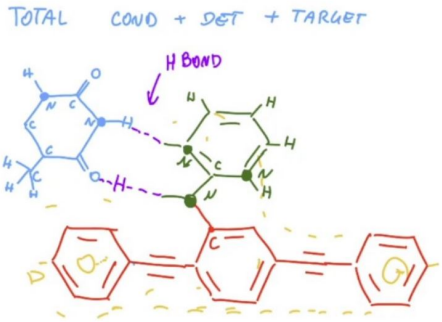


Figure 6.83: Interaction between Sensor and Thymine

Upon closer analysis, it is evident that to achieve the lowest energy state with this bonding, a certain twist in the detection unit and target is required. During bonding, this twist causes the detection unit and target to rotate by an angle of 35° relative to their original plane. This rotation means that the delocalized electrons will no longer have access to the same orbitals as before. The paths that were previously ensured by the detection unit are no longer present due to the misalignment with the conduction unit. Consequently, the delocalized bonds become less effective, resulting in an anticipated reduction in current due to the capture of Thymine.

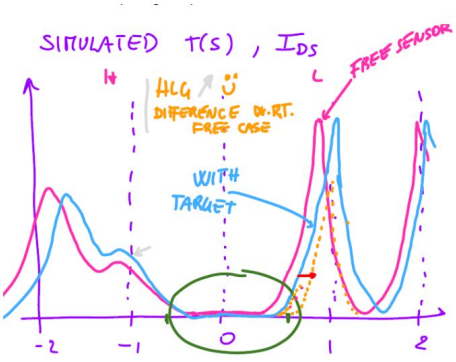


Figure 6.84: Simulated T(E) and I_{DS}

To illustrate this concept, we provide transmission spectra obtained through ab initio simula-

tions. These spectra compare the system's behavior with and without Thymine attached to the detection unit. The results align with our expectations, as the HLG (Highest Occupied Level Gap) increases when Thymine is detected. This implies that if we open the bias window, aligning with the shift towards higher energy in the LUMO peak due to the detection, the current will indeed reduce.
The current characteristics in both cases are presented, with a fixed voltage range of [0, 1] V for the drain-source voltage (VDS). However, focusing solely on the capture of the target is insufficient to fully explain the observed behavior. While we do see a difference, it's relatively small.
These results may not perfectly align with our initial expectations. The 35° rotation was anticipated to lead to a reduction in current overall. However, we do not have precise information about how the transmission spectrum changes due to this rotation. For example, we might have expected a decrease in the LUMO peak, but it remains more or less the same. This suggests that the 35° rotation does not have a substantial impact on the current. This is because, even after the rotation, we maintain a very strong channel between the source and drain, primarily due to the three rings of the OPEs. While the HLG does decrease due to the 35° rotation, the effect on the primary conductive part is not very significant.

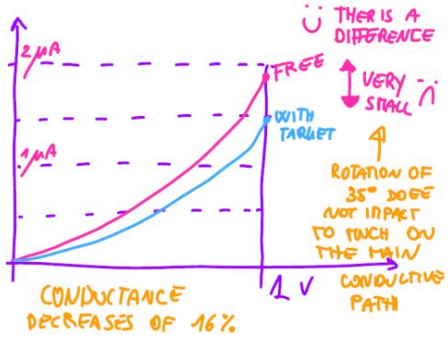


Figure 6.85: simulations of transmitted spectra and current

This experiment, involving simulations of transmitted spectra and currents through ab initio methods, has provided valuable insights. It indicates that we can indeed construct a sensor like this, but we need to find alternative ways to maximize its impact.
One challenge with the current implementation of this sensor was the nanogap, which was not sufficiently small. As a result, the length of the OPE chain had to be increased to five rings instead of just three. This change is expected to increase the HLG, and as a result, the current within our VDS range should decrease further. It's important to note that this situation differs from the impact of chain length we discussed earlier. In that case, the current increased, but the contacts were different. These systems are extremely sensitive to various factors, so careful consideration is crucial.

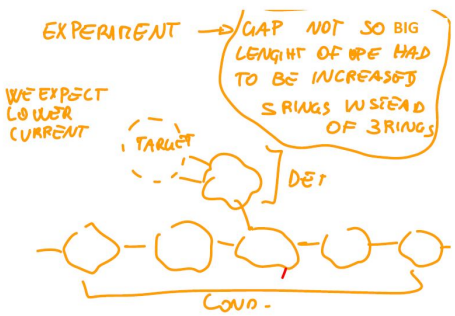


Figure 6.86: implementation of this sensor

In this particular case, the total conductance decreased by 14%. The currents are much smaller than what we observed in the ab initio simulations due to the increased length of the conduction unit. Ideally, we aim for currents of more than a few microamperes. The variation we currently have is not very significant.

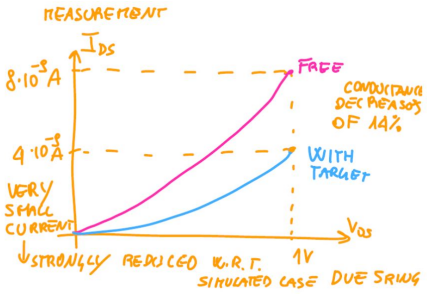


Figure 6.87: Simulated I_{DS} and V_{DS}

In this case, we have a very small current due to the length of the conduction unit. While we are not satisfied with this situation, we can overcome it by reducing the nanogap size. Although there is only a slight variation between the "free" and "with target" cases, we find this encouraging. It indicates that we need to explore other solutions to make progress. For instance, we could consider changing the type of molecule used for the conduction unit, altering the number of rings, or experimenting with the chemical aspects of detection and bonding.

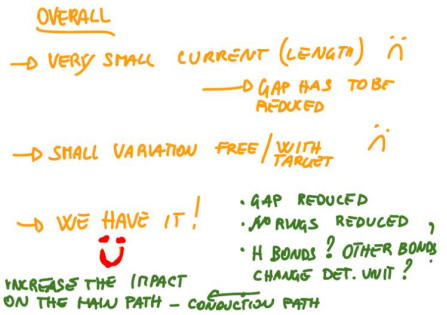


Figure 6.88: Overall of the process

The key objective here is to have a measurable impact on the current due to detection. If the rotation alone is insufficient, we could explore embedding the detection unit within the conduction unit. This way, when capturing occurs, it could trigger substantial changes within the core of the conduction unit, potentially increasing the impact on the primary conductive pathway.