POLITECNICO DI MILANO

School of Industrial and Information Engineering

Master of Science degree in Engineering Physics

Analysis of the Meyer-Neldel rule in MOSFET devices

Supervisor: Prof. Alessandro Sottocornola Spinelli
Co-supervisor: Prof. Christian Monzio Compagnoni

M.Sc. Thesis of
Giulio Franchini
Student ID Number 850913

Academic Year 2017-2018

Contents

Abstract

NAND Flash technology is nowadays one of the most implemented solutions for highly-performing non-volatile memories. The continuous development of its scaling processes brought the characteristic cell feature size (F) to values around 15 nm. This dimension reduction allowed to obtain excellent storage densities, while keeping a low cost-per-bit ratio, and high read/write performances, overcoming the efficiency offered by other technologies, such as magnetic hard-disk drives. The dimension reduction of the arrays led to different stability issues, becoming harder to control the more F was reduced, making the technological growth difficult to be carried on. In order to overcome these issues, a new approach was taken and the array structure was changed. The third dimension was employed and 3D NAND Flash arrays were developed, allowing greater storage densities while using a big enough feature size to overcome the mentioned problems. The new processes involved in the manufacturing of 3D NAND Flash employ the deposition of the semiconductor material, resulting in a polycrystalline channel.

Reliability represents the most stringent constraint to the technology development in both planar and 3D arrays, and must be carefully checked. To this aim, accelerated tests are usually carried out, in which the phenomenon under analysis is speeded up by application of voltages or temperatures higher than usual. Results are then extrapolated to normal operating conditions using appropriate relations. For the case of temperature acceleration, an Arrhenius-like relation is often found, controlled by an activation energy. The analysis of the temperature dependence of the device characteristics in both planar and 3D devices becomes of primary importance to understand the extrapolation process. This is the purpose of this thesis work along with the study of the Meyer-Neldel rule, an empirical law, usually found in amorphous and polycristalline materials, which links the activation energy and the prefactor of thermally-activated conduction processes.

Chapter 1 gives an overview of NAND Flash technology, explaining the schematic layout of the memory array and the performance of the main operations, followed by the presentation of the scaling trends and the above mentioned issues that took the

technology to 3D structures. The basic structure of different 3D NAND Flash strings is given along with the new issues brought by the implementation of this technology.

Chapter 2 describes the known temperature dependences of monocrystalline and polycrystalline devices with models taken from the literature. Emphasis is given on the variations brougut by the presence of trap states. At the end of the chapter different proposed explanations of the Meyer-Neldel rule are presented along with a Meyer-Neldel analysis performed on a TFT. Such analysis will set the guidelines for the analyses of the following chapters.

Chapter 3 has the main scope of verify the biunivocal relation that is often given between the Meyer-Neldel rule and trap states. This investigation is made by the means of simulations of monocrystalline devices, a nanowire MOSFET and a long-channel planar MOSFET without any kind of trap states. The results of the simulations are then analyzed with the aid of MATLAB. The analysis starts from the understanding of the electrostatic of the devices and employs Arrhenius plots of the current and the charge from which activation energies and prefactors are extracted. The obtained results are compared with experimental results and analytical models from the literature.

Chapter 4 presents the results of simulations with trap states of the same nanowire device presented in Chapter 3 along with the analysis of the Meyer-Neldel rule. The analysis is again based on the understanding of the underlying physical processes such as the occupation of the trap states and it employs again Arrhenius plots for the extraction of activation energies and prefactors. Throughout the chapter the obtained results are compared with the ones of the previous chapter, highlighting the role of the trap states.

Sommario

La tecnologia NAND Flash è ad oggi una delle soluzioni più adottate per memorie non volatili ad alte prestazioni. Il continuo sviluppo dei processi di *scaling* ha portato la *feature size* (F) della singola cella a valori intorno ai 15 nm. Questa riduzione delle dimensioni ha permesso di ottenere un'ottima densità di archiviazione, mantenendo un basso costo per bit, e alte prestazioni di lettura/scrittura, superando l'efficienza di altre tecnologie, come gli hard disk magnetici. La riduzione delle dimensioni degli array ha portato a diversi problemi di stabilità, che diventano più difficili da controllare al decrescere di F, rendendo sempre più complesso lo sviluppo della tecnologia. Per superare questi problemi, è stato adottato un nuovo approccio e la struttura dell'array è stata modificata. Sfruttando la terza dimensione sono stati sviluppati array 3D NAND Flash, consentendo di ottenere una maggiore densità di memoria e una feature size abbastanza grande da evitare i sopracitati problemi. I nuovi processi coinvolti nella produzione di stringhe 3D utilizzano la deposizione del materiale semiconduttore e di conseguenza il canale risultante è fatto di silicio policristallino.

L'affidabilità rappresenta il vincolo più severo per lo sviluppo tecnologico sia negli array planari che 3D, e deve essere attentamente controllata. A tal fine, vengono solitamente effettuati dei test accelerati nei quali il fenomeno in analisi viene accelerato tramite l'applicazione di tensioni o temperature superiori al normale. I risultati vengono poi estrapolati alle normali condizioni di funzionamento utilizzando opportune relazioni. Nel caso dell'accelerazione in temperatura, si trova spesso una relazione di tipo Arrhenius, controllata da un'energia di attivazione. L'analisi della dipendenza dalla temperatura delle caratteristiche del dispositivo sia nei dispositivi planari che in quelli 3D diventa quindi di primaria importanza per comprendere il processo di estrapolazione. Questo è lo scopo di questo lavoro di tesi oltre allo studio della regola di Meyer-Neldel, una legge empirica, solitamente presente nei materiali amorfi e policristallini, che collega l'energia di attivazione e il prefattore di processi di conduzione attivati termicamente.

Il capitolo 1 fornisce una panoramica della tecnologia NAND Flash, spiegando

schematicamente il layout dell'array di memoria e l'esecuzione delle principali operazioni, seguito dalla presentazione dei trend di scaling e dei problemi sopra menzionati che hanno portato la tecnologia alle strutture 3D. La struttura di base delle diverse stringhe 3D NAND Flash è poi presentata insieme alle nuove problematiche portate dall'implementazione di questa tecnologia.

Il capitolo 2 descrive le note dipendenze dalla temperatura delle caratteristiche dei dispositivi monocristallini e policristallini con modelli tratti dalla letteratura. Particolare enfasi è data alle variazioni dovute alla presenza di stati trappola. Alla fine del capitolo vengono presentate diverse spiegazioni proposte della regola Meyer-Neldel insieme ad un'analisi della regola eseguita su un TFT. Tale analisi definirà le linee guida per le analisi dei capitoli successivi.

Il capitolo 3 ha lo scopo principale di verificare la relazione biunivoca che è spesso data tra la regola di Meyer-Neldel e gli stati trappola. L'analisi viene effettuata mediante simulazioni di dispositivi monocristallini, un MOSFET nanowire ed un MOSFET planare a canale lungo senza alcun tipo di stato trappola. I risultati delle simulazioni saranno poi analizzati con l'aiuto di MATLAB. L'analisi parte dalla comprensione dell'elettrostatica dei dispositivi e utilizza dei grafici Arrhenius della corrente e della carica da cui vengono estratti le energie di attivazione e i prefattori. I risultati ottenuti saranno poi confrontati con i risultati sperimentali e i modelli analitici presenti in letteratura.

Il capitolo 4 presenta i risultati delle simulazioni con stati trappola dello stesso dispositivo nanowire presentato nel capitolo 3 oltre all'analisi della regola di Meyer-Neldel. L'analisi si basa ancora una volta sulla comprensione dei processi fisici presenti come l'occupazione degli stati trappola con le variazioni di temperatura e utilizza nuovamente i diagrammi di Arrhenius per l'estrazione delle energie di attivazione e dei prefattori. Lungo tutto il capitolo i risultati ottenuti verranno confrontati con quelli del capitolo precedente, evidenziando il ruolo degli stati trappola.

Chapter 1

NAND Flash Fundamentals

Nowadays NAND Flash is the elected technology for non-volatile memories. This is due to their characteristic of having high reading and writing throughput while keeping a low cost per bit storage capacity. This last characteristic got better and better in the last years because the feature size (F) of one memory cell keep on reducing in the last decades due to the fast scaling process reaching 15 nm. This technology managed to reach magnetic Hard Disk Drives in term of costs and overcome them in term writing and reading speeds (almost x10), furthermore, because of the absence of mechanical moving parts, they are way quieter and have a big shock resistance. Here in the first chapter an overview on NAND Flash technology will be given in order to understand their principal working operations and the underlying physical processes with the objective of understanding how the scaling process and the reduction of the feature size can affect these processes and damage their stability. The solution to these issues was found in new 3D structures in order to keep up with Moore's law trend, this solution brought many advantages but has different drawbacks. The main disadvantage (i.e the use of polysilicon in the fabrication process) has great consequences on the physic of the devices.

1.1 2D NAND Flash

1.1.1 Array structure

The fundamental element of each cell is the *Floating Gate* MOSFET (FGMOSFET), a MOSFET with a conductive layer buried in the gate oxide. This layer allows the storage of charges which can be translated into one or multiple bits, this conductive layer take the name of *Floating Gate* (FG). Many cells are connected in series to create strings which are connected on one side to a single *source-line* (SL) while on the other side one different *bit-line* is present for each string. Control gates of FGMOSFET on different

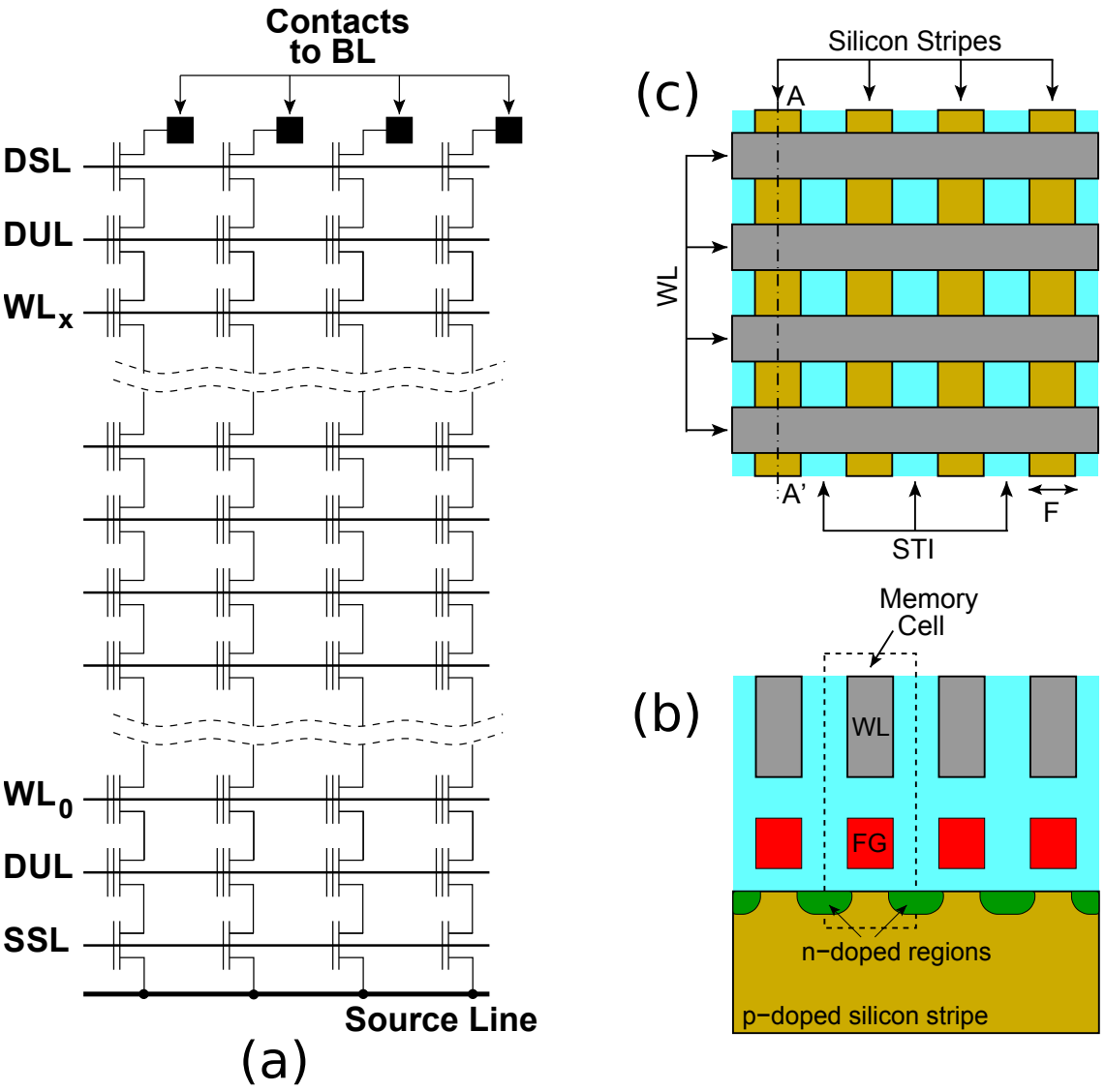


Figure 1.1: (a) Schematic circuit of a NAND Flash array. WL = *word-line*, BL = *bit-line*, SSL = *source-select line*, DSL = *drain-select line*, SL = *source-line*, DUL = *dummy line*. (b) perpendicular section along string. (c) schematic layout of the array (from [1]).

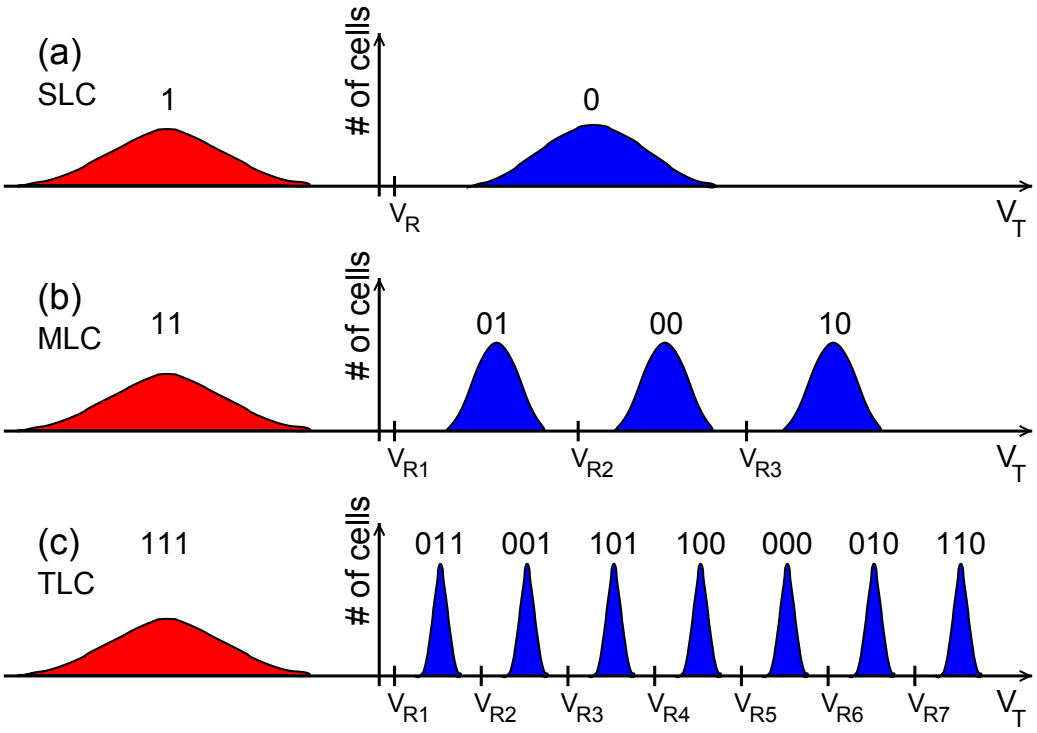


Figure 1.2: Schematic representation of V_T distributions for (a) SLC, (b) MLC e (c) TLC technologies (from [1]).

strings are formed by shared *word-lines* (WLs) which are positioned orthogonal to the strings mentioned above, therefore we obtain a matrix arrangement which allows us to select the desired cell thanks to the joint use of word-line, *Drain Select Line* (DSL) and *Source Select Line* (SSL). DSL and SSL connect and disconnect each string respectively to the *bit-lines* and the *source-line*. The FGMOSFETs just before SSL and DSL are called *Dummy cells* and are controlled by WLs named *Dummy Lines* (DUL) because they are not used to actually store any data because their behaviour can be different from all the other cells in the middle of the string due to border effects. All these components are shown in Fig. 1.1(a).

In Fig. 1.1(c) is depicted a planar view of the memory array, *Shallow Trench Insulation* (STI) is clearly visible. The silicon stripes are formed by etching removal on a wafer followed by the deposition of a dielectric material (light blue in the figure) which guarantee the isolation between adjacent strings. In the figure is also indicated the feature size (F) which is the width of the silicon stripe. The pitch of the strings and WLs is usually identical and equal to 2F, therefore the total area of a single cell is equal to $4F^2$. A single cell is highlighted in Fig. 1.1(b) which is a view orthogonal to the AA' line in Fig. 1.1(c) and orthogonal to the memory array. In this figure word lines are represented in gray, floating gate in red, oxide insulator in light blue; in the bottom we can see the

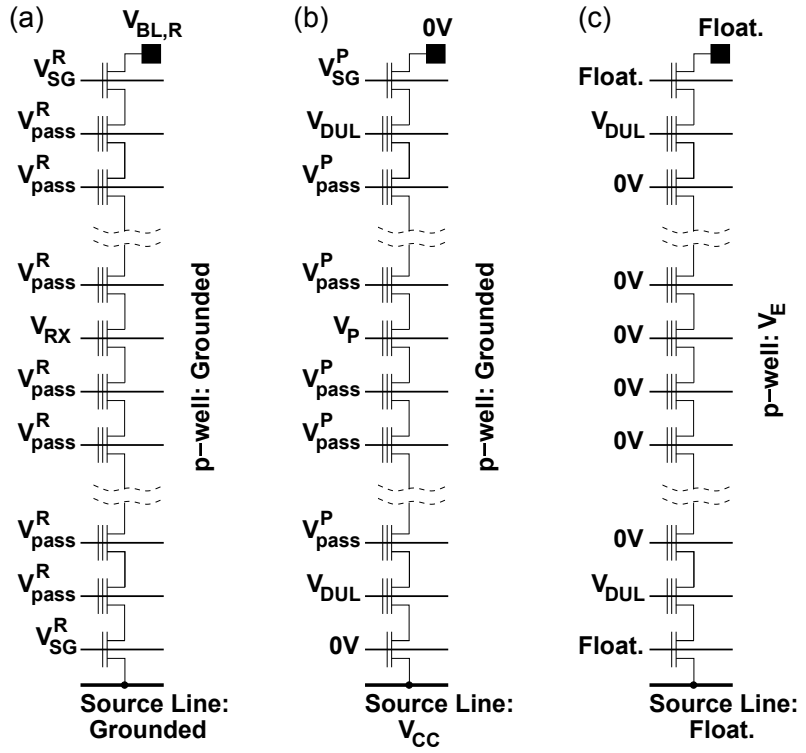


Figure 1.3: Schematic description of the applied voltages to a NAND string contacts fot the (a) read, (b) program and (c) erase. During the read and program operations the affected cell is, respectively, the one to which V_{RX} and V_P is applied. The erase operation, instead, is done on the whole string (from [1]).

p-doped silicon substrate and the source and drain n+ implantations. Note that every source and drain is in common between 2 cells. We can finally define the Net Bit Storage Density (NBSD) as $\frac{n}{4F^2}$, where n is the quantity of storable bits per cell. We will see that this parameter describes well the improvement with respect to older technologies such as magnetic hard disk drives.

1.1.2 Array operations

The threshold voltage (V_T) of a cell is defined as the gate bias such that the channel current is above a certain value (usually 10 nA). When charges are injected into the floating gate the V_T of a cell changes, more specifically it rises when the charge on the floating gate becomes more negative. This mechanism is the one exploited to store data, they are related to the charge stored in the floating gate of the cells.

In Fig. 1.2 is shown how is possible to store multiple bits in a single cell by discretizing the threshold voltage in 2^{BPC} (Bits Per Cell) levels. One, two and three BPC correspond respectively to *Single Level Cell* (SLC), *Multiple Level Cell* (MLC) and *Triple Level Cell*

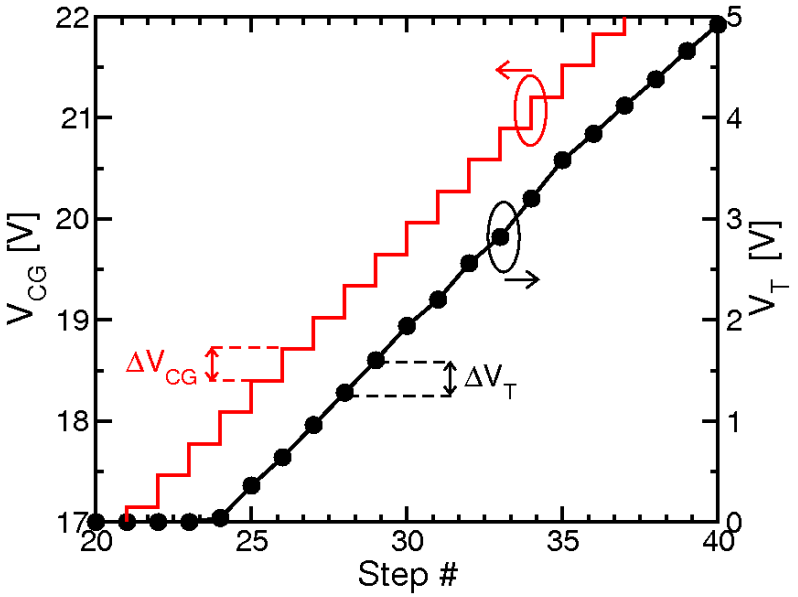


Figure 1.4: Waveform example of the voltage applied (in red) to the control gateduring the programming of a NAND flash in 60 nm technology and the corresponding programmed V_T (in black). In stationary regime $\langle \Delta V_T \rangle = \Delta V_{CG}$, allowing an accurate control of V_T (from [2]).

(TLC). Depending on the chosen technology the threshold voltage distribution changes, the state with the lowest V_T usually corresponds to the erased state while higher values refer to programmed states. The read operations are performed by $2^{BPC} - 1$ specific gate voltages named V_{RX} applied to the control gates and measuring the current flowing through the cell. There are 3 main operations: read, program and erase.

1. *Read Operation*: as written above the read operation is performed by measuring the current flowing through the cell when a certain value V_{RX} is applied. All the other cells that belong to the same string must not limit the string current in order not to make errors in the reading operation. Hence their gate voltages are set to a value V_R^{pass} that is well above the maximum threshold voltage to make them work as pass transistors. The string selection is performed by having DSL and SSL working as pass transistors by applying a value V_R^{SG} to their control gates. The SL is grounded while the corresponding BL has an applied voltage $V_{BL,R}$>0 V so that a current can flow through and the value of this current must be affected only by the programming state of the cell. To detect if the threshold voltage of the cell is below or above the applied V_{RX} a sense amplifier is used. One of the most important characteristic of the read operation that must be taken into account is the time needed to perform such operation: it starts from a few tens of microseconds for SLC technology, rising for MLC and TLC because they need multiple reading steps. This time is much longer if compared to other technologies such as NOR Flash.

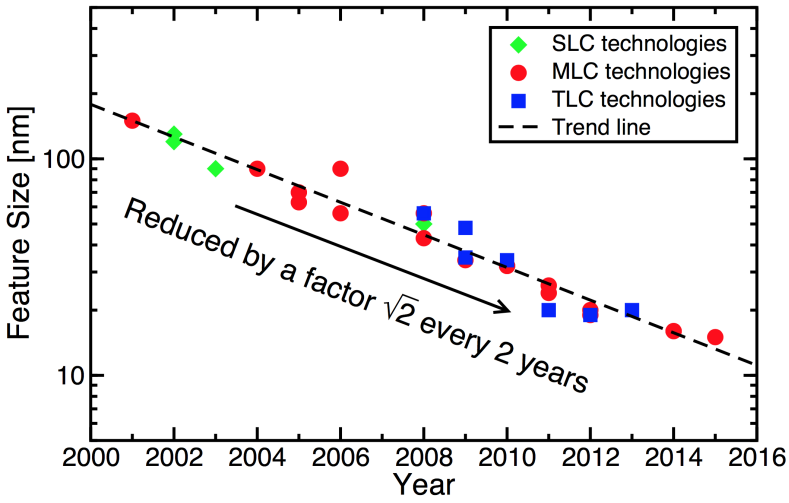


Figure 1.5: F values for the NAND Flash manufactured by the leading semiconductor companies from 2001 to 2015. The employment of a 3D Flash technology allows to increment the technological node while keeping a constant feature size (from *International Solid-State Circuits Conference* (ISSCC)).

The origin of this time overhead are all the parasitic resistances and capacitances along the word-lines. Nowadays different strategies are employed to reduce the read times, like sequential reads and exploiting the parallelism of multiple strings all driven by the same set of WLs. The bias configuration is shown in Fig. 1.3(a)

2. *Program Operation*: as previously said it consists in injecting electrons into the Floating Gate till the threshold voltage of that cell will reach a specified value. The physical process exploited for this operation is called *Fowler-Nordheim tunneling*. It allows the flow of electrons through the oxide layer that separates the silicon channel and the floating gate, also called tunnel oxide. The programming bias V_P is relatively high ($\simeq$ 20 V), it is applied on the control gate, being the body grounded, a strong electric field is generated enabling the tunneling. The other cells have a bias (V_{pass}^{P}) such that they will work as pass transistors. The string is disconnected from the source line by applying 0 V to the SS, while the DS is activated with a voltage higher than 0. The dummy lines are isolated from the string selectors with a bias V_{DUL} lower than V_P^{pass}. This is shown in Fig. 1.3(b). The string that have to be programmed has its BL biased at 0 V while the others have the same bias of their selectors (DS) hence they are disconnected. All this bias scheme is set to maintain the bias in the selected cell to 0 V, while the capacitive coupling between the channels of the other cells that share the same WL and the WL make the bias of these channels to rise. This is called *self-boosting* and prevent the unwanted

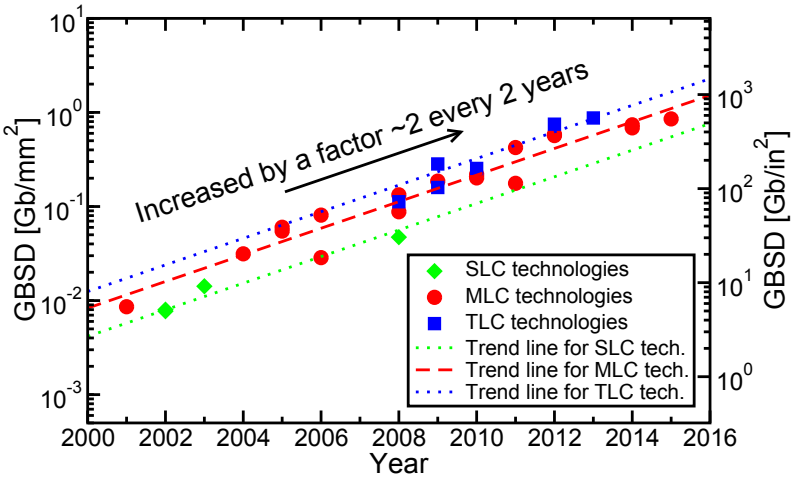


Figure 1.6: GBSD technological growth for cells using SLC, MLC e TLC technologies (from [1]).

programming of the other cells. The law that is followed by the threshold voltage of the programmed cell is:

$$V_T = \frac{q\Delta n}{C_{PP}} \tag{1.1}$$

where q is the elementary charge, Δn the number of electrons injected in the floating gate and C_{PP} the capacitance between the control and floating gates. The method to allow a precise control of the amount of charge injected is called *Incremental Step Pulse Programming* (ISPP). It is a sequence of incremental programming pulses and verify operations. The verify operations check the reached V_T and control that is equal to the desired voltage V_{PV}. The ISPP method is shown in Fig. 1.4, the time needed are in the order of hundreds of microseconds and few milliseconds for SLC MLC and TLC respectively [3] [4]. The average time can be reduced by exploiting the parallelism of the structure.

3. *Erase Operation*: if writing data is about injecting charge is straightforward that erasing means removing charges from the floating gate, the same tunnel mechanism is exploited. It is not possible to erase one single cell because the substrate bias is in common between many cells, the area with the same substrate (p-well) bias is called block (typical size of 4 MB). The electric field must be opposite of the one of the writing operation, therefore all the WLs will be biased at 0 V while the p-well is set to a high positive bias V_E allowing the tunnel of electrons from the floating gate to the p-well. In order to reduce the electrical stress on the device, all the string selectors and source and bit line are left floating while the dummy lines are biased to a bias $V_{DUL} > 0$. The operation takes few milliseconds.

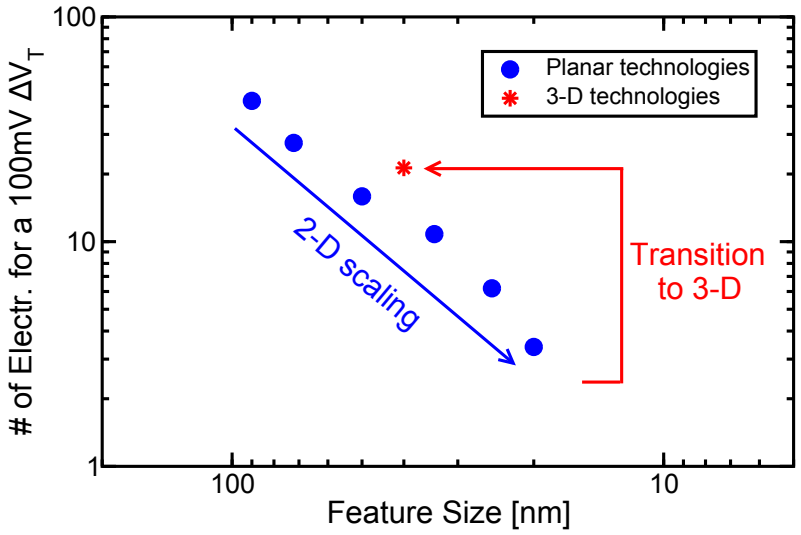


Figure 1.7: Number of electron transferred to/from the FG to cause a $\Delta V_T = 100$ mV depending of the F value (from [5]).

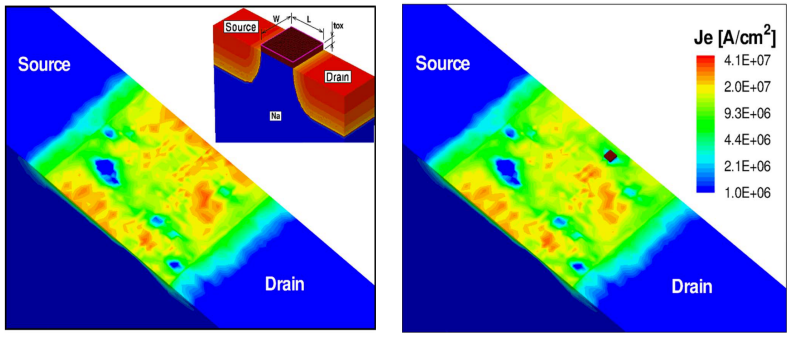
1.2 Technological growth

The demand for big data storage devices never stop rising and brought all the memories companies to increase the capacity of their products. The improvement of the processes capabilities allowed by the technological development permitted the reduction of the feature size F of a factor $\sqrt{2}$ every 2 years. This is completely consistent with the prediction of the Moore's law as we can see in Fig. 1.5. Another figure of merit is the *Gross Bit Storage Density* (GBSD) that is the memory capacity per unit area of the device, if C_{chip} and A_{chip} are respectively the total capacity and total area we have $GBSD = \frac{C_{chip}}{A_{chip}}$. In Fig. 1.6 it is reported its agreement with the Moore's law, it can also be seen the vertical shift between the different technologies SLC MLC and TLC. The miniaturization process reached a feature size of 15 nm and this brought many drawbacks, analyzed in the next subsection, that made the companies moving to a 3D structure. Exploiting the third dimension (i.e. perpendicular to the silicon wafer) it will be shown how it is possible to increase the GBSD while using a higher feature size, hence avoiding the just mentioned drawbacks.

1.2.1 Physical scaling issues

The scaling of the feature size brought many advantages in term of GBSD and costs, but have severe drawbacks in term of stability caused by the reduction of the cell size:

1. *Program Noise*: The reduction of the cell size implies also a reduction of C_{pp} which in turn increases the sensitivity of V_T over Δn. The number of electrons needed to cause



(a) (b)

Figure 1.8: (a) Example of a 3D simulated current density in a planar MOSFET in which an atomistic doping has been considered. (b) A fixed electron was placed on the silicon/oxide interface to simulate a filled trap and the current density was extracted again (from [6]).

a shift of 100 mV with respect to the feature size F is shown in Fig. 1.7. It decreased dramatically reaching values in the order of units. Each voltage step during the ISPP inject a certain number of electrons, if the number needed is very small the width of the step has to decrease, not to affect the accuracy of the programming operation. Besides, the variability among different cells is not a problem since the verify operation performed get rid of this source of instability. The increase of the feature size brought by switching to a 3D architecture can overcome this problem because the number of electrons needed is increased proportionally with F.

2. *Time dependent variabilities on* V_T: in the tunnel oxide are present some defects in which electrons can be trapped and the reduction of the dimensions of the cell enhances the effect of these discrete defects. The capturing or releasing of charges inside these defects are random events that affect the threshold voltage of the cell. Furthermore, also the dopants are discrete and this cause the current to be percolative, the relative distance between the charge trapped in the defect and the percolative path of the current strongly affects the V_T shift. Specifically, a charge closer to the percolation path has a stronger effect on the threshold voltage.

The dimension reduction led to increasing issues because the average number of defects gets smaller with the dimension, making the statistical fluctuation of this number more and more important for the array operation. The combined effect of this increased statistical fluctuation and the enhanced effect of one single capture/release event has strongly increased the statistical dispersion of the threshold voltage instabilities, creating big challenges for a correct data retention. In Fig. 1.8 it is shown the effect of a discrete defect on the percolative current while in Fig. 1.9 we can see an example of instability coming for capturing/releasing in

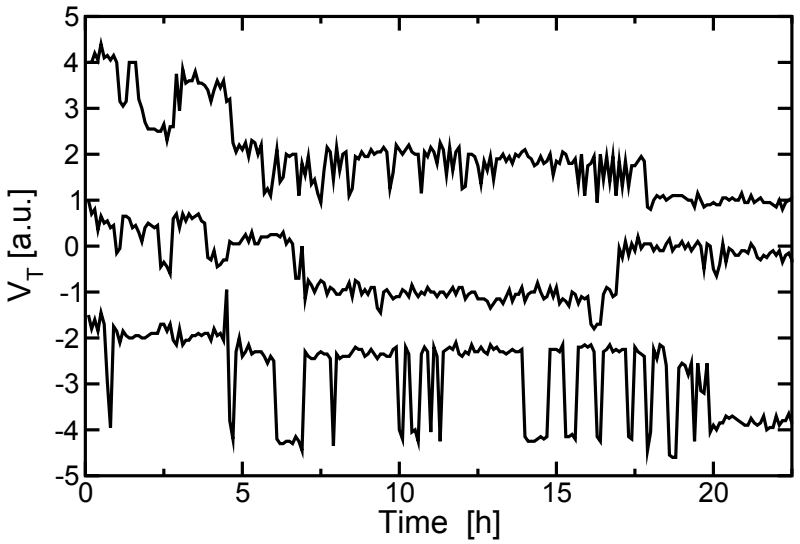


Figure 1.9: V_T profile for three different cells with respect to the retention time. Multiple program/erase cycles were performed before the measurement (from [1]).

the tunnel oxide defects.

There are two main effects: The first one is *random telegraph noise* (RTN) occurring when tunnel-oxide defects periodically capture and release single charge carriers from/to the channel of the memory cell, leading to a two-state fluctuation of V_T. RTN affects the stability as soon as the program operation is finished because the distribution of time constants for capturing, releasing events is very wide. The second is called *postcycling charge detrapping*. This phenomenon arises from the capture of negative charge in the cell tunnel oxide during the program and erase operations and from the subsequent neutralization of this charge during the idle periods of data retention. When the captured charges are released the threshold voltage shifts towards lower values as time elapses as shown in the picture. This process enlarge the page V_T distribution towards lower values. Both the mentioned phenomena worsen with the number of program/erase cycles performed on the array, limiting the endurance of the new memory chips to a few 10^3 program/erase cycles in the state-of-the-art technologies.

3. *Cell-to-Cell Electrostatic Interference*: In order to achieve an high GBSD while reducing F, not only W and L have been reduced but also the relative distance between cells in the memory array. This made the electrostatic interference between adjacent cells more and more relevant for the array operation: the charges stored in a cell significantly affect the threshold voltage of adjacent cells. Notwithstanding that electrostatic interference between neighboring cells may show up any time dur-

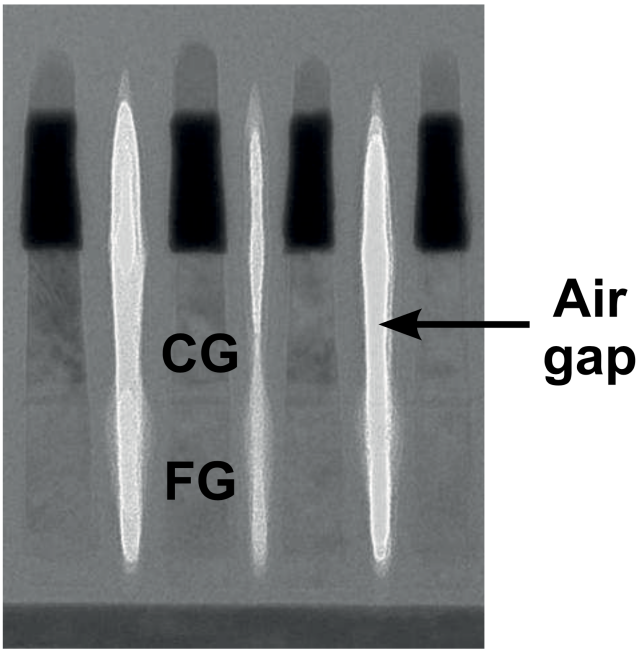


Figure 1.10: Vertical cross section of a NAND Flash array showing the presence of air gaps between adjacent cells (from [7]).

ing array operation, the most critical phase is during programming. The program operation causes huge changes in the charge in the floating gate on the selected cell and hence affects a lot the V_T of the adjacent cells, thus enlarging the distribution of threshold voltages among cells and worsen the performances of the array, requiring very precise ISPP step to inject the desired charge with a low error. Many solution to this issues have been studied and implemented, the most successful has been the introduction of air gaps in-between cells belonging to adjacent WLs as shown in Fig. 1.10. In this way the dielectric constant is reduced and hence the parasitic capacitance, limiting this phenomenon.

1.3 3D NAND Flash

As written above the solution of all the issues presented in the previous section was moving to a 3D structure. The aim was to keeping on increasing the GBSD avoiding the problems presented in section 1.2. The first experiment was done by stacking multiple 2D arrays one on the top of the other. This solution increased the GBSD but did not have advantages in term of process costs and complexity (i.e the expensive lithographic steps). The two main solutions are called *vertical channel memories* and *vertical gate memories*. The former is now presented, highlighting the benefits and the drawbacks of the structure, because is the most important array solution to pursue an equivalent

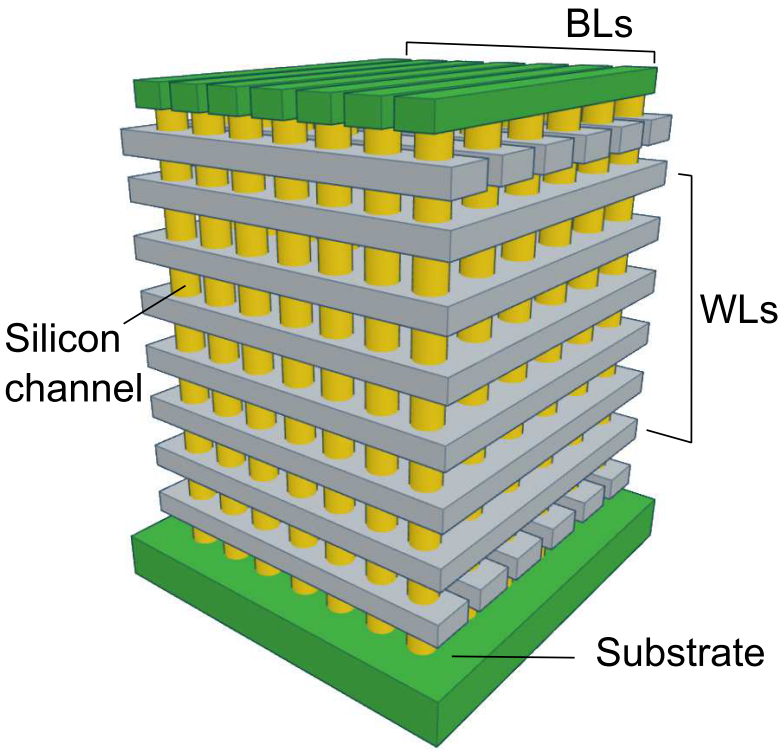


Figure 1.11: 3D vertical-channel NAND Flash array schematic representation (from [1]).

scaling of the technology.

1.3.1 3D vertical channel array structure

One of the main ideas was that a reduction of the critical lithographic steps could be achieved if the silicon channel could be implemented orthogonal to the wafer surface. In Fig. 1.11 is presented the schematic structure of a vertical-channel 3D NAND Flash array. The silicon channels runs vertically from the substrate to the BLs while the WLs are planes intersecting the channels. Like in the planar case the first and last WLs are select transistors, their arrangement (one orthogonal to the other) in a matrix-like manner makes possible to select a single channel and thus a single cell; the cells of this selectors WLs are dummy cells. The structure implemented is the so called *gate-all-around* (GAA) that allows the best electrostatic control of the channel by the gate. With this structure the quadruple lithographic steps required for the 15 nm planar chip could be avoided, it has the same GBSD with only 24 layers and a feature size F of 40 nm. The WLs are made from metal or high-doped polysilicon, the different solution depends on the manufacturing process. The substrate, connected to all strings, takes the role of the source line.

In Fig. 1.12 the so-called *macaroni* structure is shown. This solution has been

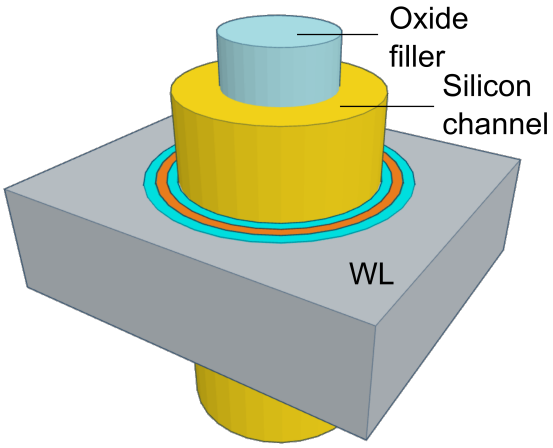


Figure 1.12: Schematic view of a single Macaroni cell (from [1]).

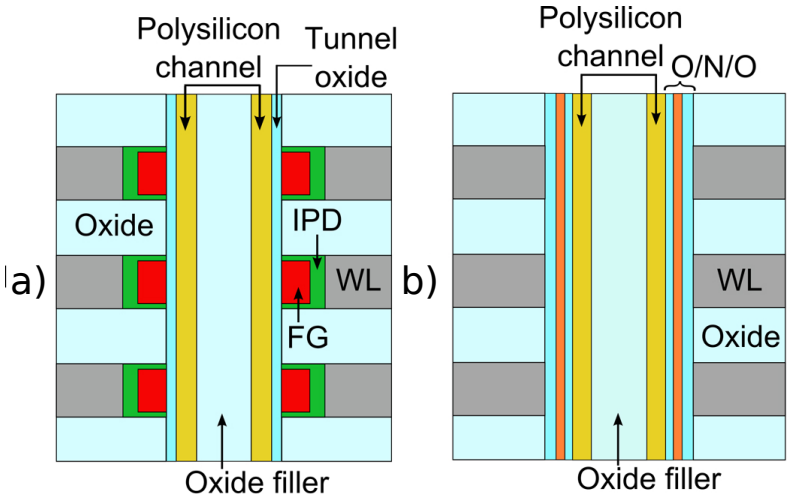


Figure 1.13: Representation of the vertical section of a string based on (a) polysilicon floating gate and (b) charge trapping in an ONO stack (from [1]).

proposed because the process needed for the production of these structures involves the deposition of silicon in the etched holes. Thus dangling bonds and lattice distortion will be present in the deposited silicon, hence the semiconductor will have a *polycrystalline* structure, with the presence of defects on the grain boundaries (GBs). To limit the absolute number of these defects, one solution is to limit the volume of silicon by filling the middle of the channel with a cylinder of oxide, called *oxide filler*. Different solution for the data storage have been proposed, in Fig. 1.13 the two principal ones are presented. In the case a) the solution is similar to the planar case. An isolated floating gate is present (red) that have been isolated from the WL by the use of the *Inter-poly Dielectric* (IPD). The second solution exploits an high defective material: *Oxide-Nitride-Oxide*

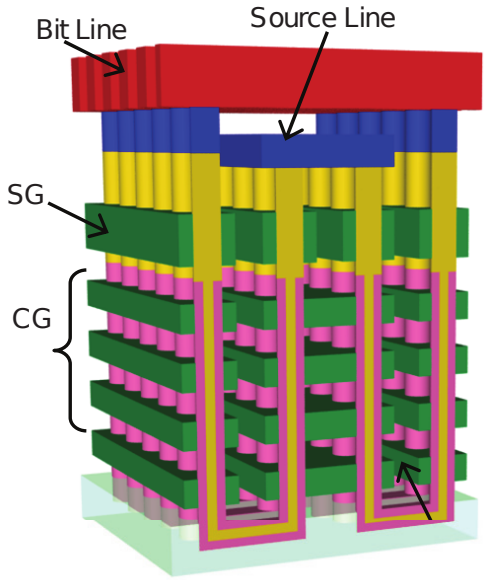


Figure 1.14: Rendering of an array based on P-BiCS strings (from [8]). *SG* stands for *Selector Gate* and *CG* for *Control Gate*.

(ONO) made of $SiO_2 - Si_3N_4 - SiO_2$. The charge is stored in the defect present in the nitride layer isolated by the two oxide films.

In Figs. 1.14 and 1.15 are respectively shown the *Pipe-shaped Bit Cost Scalable* (P-BiCS) presented by Toshiba in 2009 and the *Terabit Cell Array Transistor* (TCAT) by Samsung. The main difference between this two solutions is that in the latter a p-substrate is present and thus the erase operation can be performed similar to the planar case. In the other structure is exploited a phenomenon called *Gate Induced Drain Leakage* (GIDL) in which the holes needed for the erasing operation are obtained by a band to band breakdown.

1.3.2 Reliability tests and temperature

We have already mentioned that a tunneling process is exploited for the operations of the memory array. This process causes the aging of the device because it damages the thin insulator layer [9]. Recovery of damages induced by program/erase (P/E) cycles is a major source of threshold voltage (V_T) instability during data retention for deca-nanometer NAND Flash memories. This instability represents the worst reliability issue coming from spurious charge trapping in the tunnel oxide and interface state creation during repeated (P/E) cycles [10], arising from the possibility for the same oxide charge to detrap and for interface states to anneal out when cells have to preserve their data.

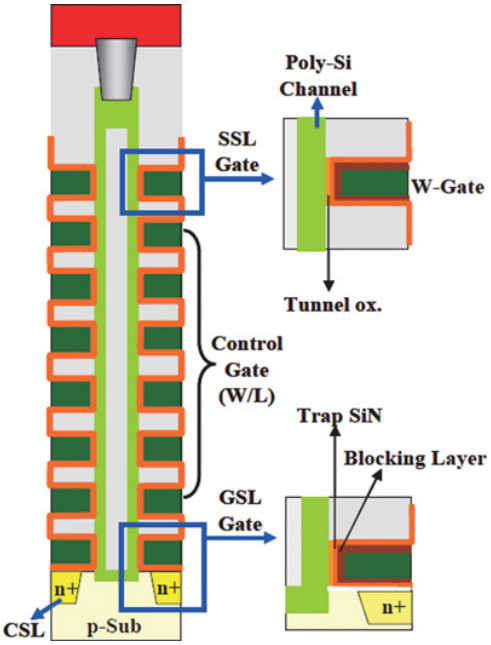


Figure 1.15: Schematic representation of a TCAT NAND Flash string with a zoomed view of two sections (from [11]).

The amount of cell damage contributing to V_T instabilities during data retention is the result of not only the number of P/E cycles previously performed on the array but also the time delay between cycles and the cycling temperature [12]. Characterization tests where P/E cycles are performed in quick succession to minimize the required experimental time, which is usually referred to as *fast-cycling* tests, provide only worst case results for the V_T instabilities during data retention. A more realistic test should, instead, reproduce the time distribution of P/E cycles that is reasonably expected in real device operation. To this aim, *distributed-cycling* experiments should be designed, trying to solve the tradeoff between a low characterization time and a correct reproduction of the amount of damage at the end of cycling. In so doing, the increase of the cycling temperature to obtain in affordable experimental times the same damage recovery that should be obtained at the device working temperature on a much longer cycling time scale appears as the most practical solution.

In particular, the temperature increase is used with the aim of reproducing with short bake times the same V_T instabilities taking place at room temperature on a much longer time scale. To practically manage the temperature effect, an Arrhenius law is usually assumed to describe the temperature dependence of the time needed to reach a selected ΔV_T, introducing, in so doing, an activation energy E_a for the damage-recovery process.

In this framework it is evident that the study of the temperature behavior of the memory devices is needed along with the investigation on the activation energies involved in the various processes exploited in the operations of the devices.

Chapter 2

Temperature dependences

In this chapter the main temperature dependences that affect current conduction of different semiconductor devices are reported. The focus will be on conduction in the sub-threshold region, because it is the most interesting for the study of the Meyer-Neldel rule. The first device chosen for our analysis is a monocrystalline device such as a planar long-channel transistor, with a brief review on the interface traps that could affect the conduction, followed by polysilicon devices and their temperature dependences. Grain boundaries and trap states will be investigated with focus on how temperature changes the occupation of the trap states. At the end of the chapter a few Meyer-Neldel rule believed origins are outlined and an example of MNR analysis is reported in detail. Such example will set the guidelines for the analysis of the rule performed in the thesis work.

2.1 Monocrystalline device

2.1.1 Long channel planar MOSFET

In the depletion N-Type MOSFET a p-type substrate is brought to an inversion condition at the semiconductor-oxide interface by the bias applied to the gate, V_{gs} in Fig. 2.1. The positive bias will bend the conduction band towards the Fermi level, creating a depletion region from which the holes are pushed away, making the electrons concentration rise at the interface.

The inversion condition is considered to have been reached when the minority carriers (electrons) concentration at the interface equals the majority carriers (holes) concentration in the substrate creating a highly-conductive path between two heavily n-doped regions: source and drain. If a voltage (V_{ds}) is applied between these two terminals when the inversion condition is reached, a current will flow. This state is

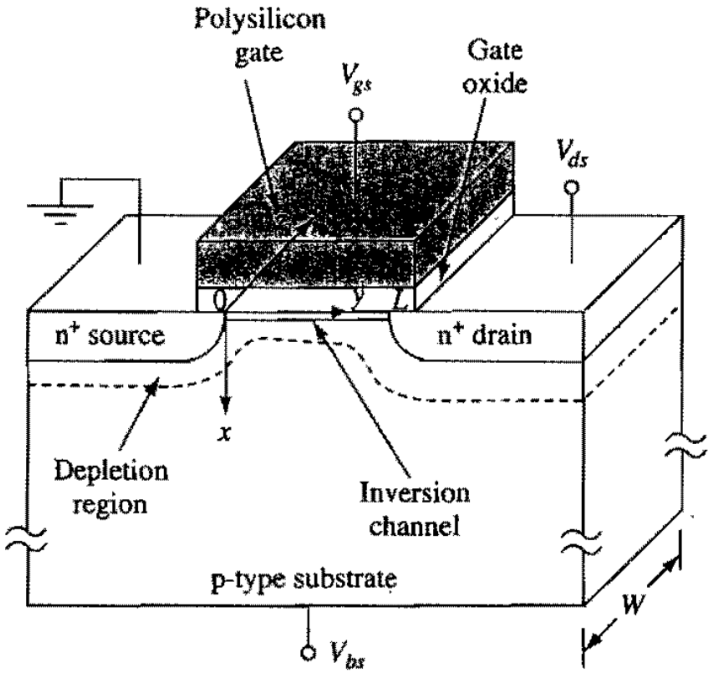


Figure 2.1: Schematic MOSFET cross section, showing all the biases (grounded source) and the dimension L: channel lenght; W device width).

known as *on-state* or *above-threshold* region, referring to the threshold voltage V_t that V_{gs} must reach in order to enable the inversion condition.

Naturally the electron concentration does not follow a step-like relation; the latter is an approximation of a steep exponential relation, and even for $V_{gs} < V_t$ enough carriers are present at the interface to make a current flow if a bias $V_{ds} > 0$ between source and drain is applied. This is called *sub-threshold* region. This region is the one the analysis will be focused on because, as it will be shown, it is relevant for the analysis of the Meyer-Neldel rule.

2.1.2 Subthreshold current

In the subthreshold region the inversion charge is much smaller than the depletion charge, which is created by the removal of the holes and has the magnitude of the doping concentration. The inversion charge sheet density is regulated by [13]:

$$-Q_i = \sqrt{\frac{\epsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right) \left(\frac{n_i}{N_a}\right)^2 e^{q(\psi_s - V_{fn})/kT}, \tag{2.1}$$

where ψ_s is the voltage drop in the silicon substrate, ϵ_{si} the silicon permittivity, N_a the substrate doping, q the elementary charge, and $V_{fn} = E_{fn}/q$ is the electron *quasi-Fermi potential* at the semiconductor-oxide interface where E_{fn} is the electron *quasi-Fermi level*. Since the inversion charge density is small ψ_s is a function of V_{gs} only, independent of V_{fn}. Hence the electric field along the channel direction is small and in turn the drift current is negligible.

The current in this condition is a diffusion one, therefore it will follow:

$$J_{n,diff} = kT\mu_n \frac{dn}{dx} \tag{2.2}$$

Performing the integration along the channel direction one obtains:

$$I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\epsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right)^2 \left(\frac{n_i}{N_a}\right)^2 e^{q\psi_s/kT} (1 - e^{-qV_{ds}/kT}). \tag{2.3}$$

If V_{ds} is about only 50 mV, the last exponential takes a value one order of magnitude lower than 1 and can be neglected, leading to:

$$I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\epsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right)^2 \left(\frac{n_i}{N_a}\right)^2 e^{q\psi_s/kT}. \tag{2.4}$$

It should be noted that, to a first approximation the current is proportional to charge, mobility, temperature and dimensions.

2.1.3 Subthreshold slope

To highlight the effect of temperature on the subthreshold region, Eq. (2.3) can be rewritten as function of V_{gs} assuming that ψ_s has little deviation from the inversion value $2\psi_B$ [14], where $\psi_B = \frac{kT}{q} \ln \frac{N_a}{n_i}$.

The device parameter

$$m = 1 + \frac{\sqrt{\epsilon_{si} q N_a / 4\psi_B}}{C_{ox}}, \tag{2.5}$$

where C_{ox} is the unit area oxide capacitance, can also be expressed as:

$$m = 1 + \frac{C_s}{C_{ox}} \tag{2.6}$$

since:

$$C_s = \sqrt{\frac{\epsilon_{si} q N_a}{2\psi_s}} = \sqrt{\frac{\epsilon_{si} q N_a}{4\psi_b}} \tag{2.7}$$

when the inversion condition is reached.

Therefore:

$$m = \frac{C_s + C_{ox}}{C_{ox}} = \frac{C_s}{C_{tot}} \tag{2.8}$$

where C_{tot} is the series of the substrate and oxide capacitances. m can also be interpreted as the sensitivity of the substrate voltage drop to the gate bias, i.e. $\Delta V_{gs}/\Delta\psi_s$.

Reversing the definition of ψ_B one obtains:

$$e^{-\frac{q\psi_B}{kT}} = \frac{n_i}{N_a}. \tag{2.9}$$

Since the device is operating near the inversion condition and thus the substrate bias ψ_s will be close to the inversion value $2\psi_B$, we can express the exponential in Eq. (2.4) in term of V_{gs} since $V_{gs} - V_t = m(\psi_s - 2\psi_B)$ and finally $V_{gs} = V_t + m(\psi_s - 2\psi_B)$. Therefore Eq. (2.3) can be rewritten as:

$$I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\epsilon_{si} q N_a}{4\psi_B}} \left(\frac{kT}{q}\right)^2 e^{q(V_{gs}-V_t)/mkT} (1 - e^{-qV_{ds}/kT}), \tag{2.10}$$

leading to the expression for the *subthreshold slope* (STS):

$$STS = \left[\frac{\partial \log_{10} I_{ds}}{\partial V_{gs}}\right]^{-1} = \frac{mkT}{q} \ln(10) \left[\frac{\text{mV}}{\text{dec}}\right]. \tag{2.11}$$

Since STS, as one can see from the definition, is the inverse of the actual I-V relation slope, higher values mean lower slope values, and hence with the rising of temperature the I-V slope will get more gentle.

2.1.4 Threshold voltage

Another important temperature effect is measurable on the threshold voltage V_t of the transistor:

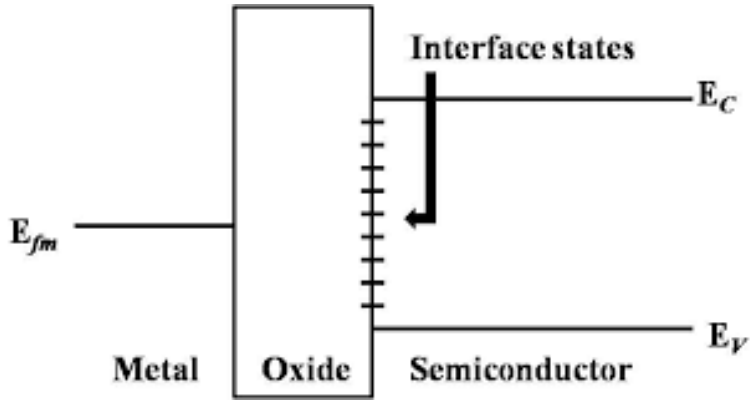


Figure 2.2: Schematic energy-band of a MOS structure, illustrating the presence of interface states.

$$V_t = \frac{E_g}{2q} + \psi_B + \frac{\sqrt{4\epsilon_{si} q N_a \psi_B}}{C_{ox}} \tag{2.12}$$

by differentiation we obtain:

$$\frac{dV_t}{dT} = -\frac{1}{2q}\frac{dE_g}{dT} + \left(1 + \frac{\sqrt{\epsilon_{si} q N_a \psi_B}}{C_{ox}}\right)\frac{d\psi_B}{dT} \tag{2.13}$$

with $\mathrm{d}\psi_B/dT$:

$$\frac{d\psi_B}{dT} = \frac{d}{dT}\left[\frac{kT}{q}\ln\frac{N_a}{\sqrt{N_c N_v}e^{-E_g/2kT}}\right] = \\ -\frac{k}{q}\ln\left(\frac{\sqrt{N_c N_v}}{N_a}\right) - \frac{kT}{q\sqrt{N_c N_v}}\frac{d\sqrt{N_c N_v}}{dT} + \frac{1}{2q}\frac{dE_g}{dT} \tag{2.14}$$

Since N_c and N_v are proportional to $T^{3/2}$ we have $d(N_c N_v)^{1/2}/dT = \frac{3}{2}(N_c N_v)^{1/2}/T$ and substituting Eq. (2.14) into Eq. (2.13) we obtain:

$$\frac{dV_t}{dT} = -(2m-1)\frac{k}{q}\left[\ln\left(\frac{\sqrt{N_c N_v}}{N_a}\right) + \frac{3}{2}\right] + \frac{m-1}{q}\frac{dE_g}{dT}. \tag{2.15}$$

We know from the literature that $dE_g/dT \approx -2.5 \times 10^{-4}$ eV/K at room temperature, so dV_t/dT is always negative and typically around -1 mV/K, meaning that the on-state is reached earlier with rising temperature.

2.1.5 Interface states and interface trapped charge

Many of the solids properties like the band structure and the presence of a forbidden gap arise from the lattice periodicity. This periodicity terminates at the silicon-oxide

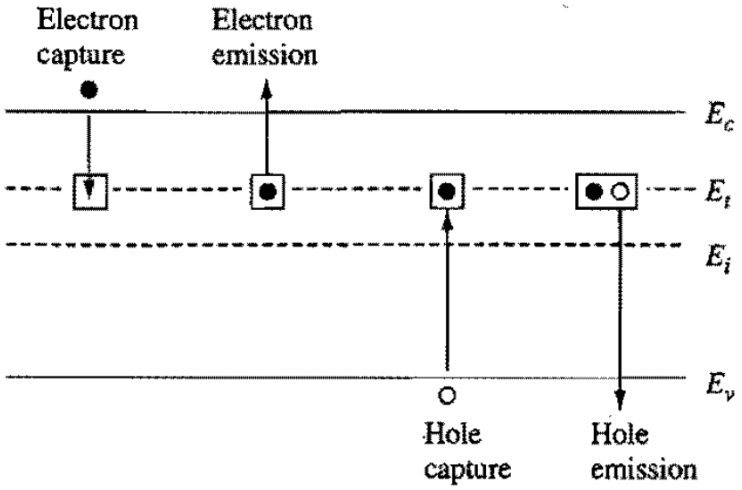


Figure 2.3: Schematic illustration of electron and hole capture and emission processes at a trap center located at energy level E_t.

interface, leaving dangling bonds at the end of the lattice that create new localized states with energy in the forbidden gap of silicon [15] as schematically shown in Fig. 2.2. Their occupancy probability is regulated by the surface-state energy relative to the Fermi level through the Femi-Dirac statistics.

Interface states have a fixed energy with respect to the energy-band edges at the interface. If, for example, the conduction band is bent and its edge is shifted downwards by 0.5 eV the energy of the localized states at the interface will be shifted along by the same value. In a planar MOSFET at flat bands, with the drain bias set to 0, once temperature and doping are fixed, the Fermi level E_f is fixed and flat along the direction perpendicular to the silicon-oxide interface, as are the bands shown in Fig. 2.2. When gate bias is changed, the bands bend and these states, that move along with the band, get closer or farther to the Fermi level, changing their occupation probability.

The presence of fixed charges obviously affects the electrostatics of the system, in particular the *I-V* characteristic. When the gate bias is raised, the induced negative charges will populate the trap states instead of the conduction band hence reducing the current flow. The *I-V* relation will be more stretched or, in other words, the STS will be higher. This effect can also be explained looking at Eq. (2.11): we have already shown that the parameter m is the ratio between the substrate capacitance and the total capacitance between the substrate and the gate, i.e. $C_s/C_{tot} = (C_s+C_{ox})/C_{ox}$. When traps are added, their capacitance contribution must be added to the numerator, obtaining a higher STS value, and hence a more gentle rise of the curve:

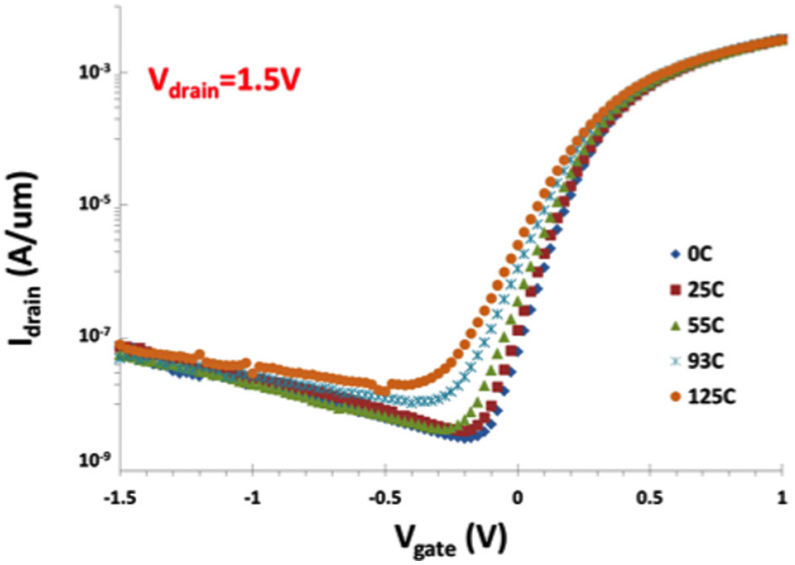


Figure 2.4: I_{drain} vs. V_{gate} at different temperatures for a 90 nm n-channel MOSFET device (from [16]).

$$STS = \left[\frac{\partial \log_{10} I_{ds}}{\partial V_{gs}}\right]^{-1} = \frac{kT}{q}\frac{C_{ox} + C_s + C_{traps}}{C_{ox}} \ln(10) \left[\frac{\text{mV}}{\text{dec}}\right]. \tag{2.16}$$

Trapped electrons and holes cannot contribute readily to the current, but have to make a transition to the conduction or valence band first; interface states can reduce the conduction current in MOSFETs by trapping electrons and holes. Furthermore, the trapped electrons and holes can act like charged scattering centers, located at the interface, for the mobile carriers in a surface channel, and thus lower their mobility [17]. Moreover, interface states can act like localized generation-recombinations centers for the *Shockley-Read-Hall Recombination* process, that is a defect-assisted process shown in Fig. 2.3. Recombination happens when an electron is first trapped from the conduction band and then it is subsequently recombined with a captured hole. Vice versa, traps can emit a hole and then emit an electron and act as generation centers.

The density of interface states and thus of traps is function of the silicon substrate orientation and a strong function of the fabrication process [18]. A postmetallization or a final anneal in hydrogen at high temperatues are quite effective in minimizing the density of interface traps and are commonly used in modern device fabrication.

2.1.6 Arrhenius analysis and activation energy

It is meaningful to perform an Arrhenius analysis every time there is a temperature-activated process which exhibits a relation like:

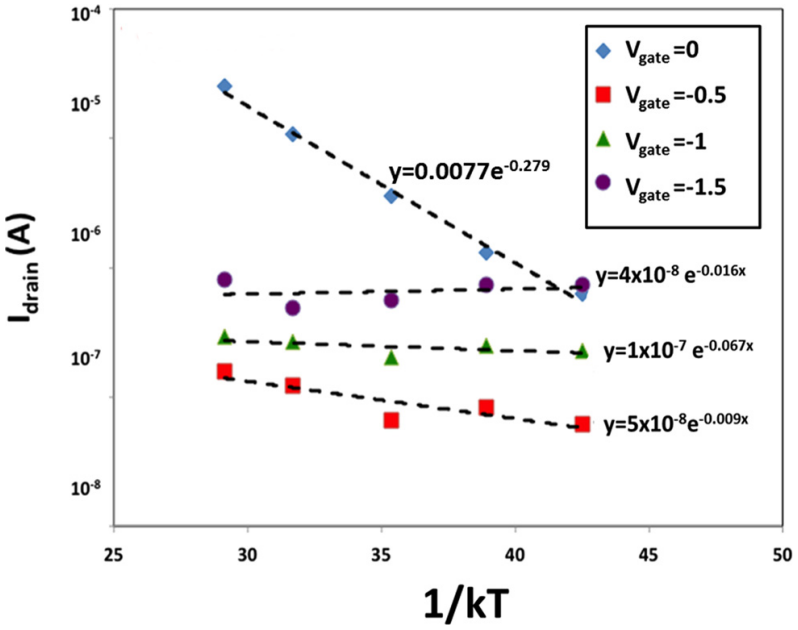


Figure 2.5: $\log I_{drain}$ vs $1/kT$ extracted from the I-V in Fig.2.4 (from [16]).

$$X = X_0 e^{-E_a/kT} \tag{2.17}$$

In this context the activation energy represents the energy barrier that have to be overcome for the physical process under analysis to take place.

The starting point for the Arrhenius analysis is to measure or simulate I-V relations at different temperatures and sample all of them at chosen biases. For every bias the sampled points will be plotted in a semilogarithmic plot against $1/kT$ obtaining the Arrhenius plot of the current. An example is shown in Fig. 2.5, computed from the I-V curves illustrated in Fig. 2.4. If the current relation is of the type of Eq. (2.17), the logarithm will yield:

$$\ln X = \ln X_0 - \frac{E_a}{kT} \tag{2.18}$$

hence obtaining straight lines in which the slope will be $-E_a$ and the intercept $\ln X_0$.

In our framework it is easy to recognize in Eq. (2.4) the form of Eq. (2.17). Accounting for the exponential term in the intrinsic concentration, that is $n_i = \sqrt{N_c N_v} e^{-E_g/2kT}$, we can rewrite the formula as:

$$I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\epsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right)^2 \left(\frac{\sqrt{N_c N_v}}{N_a}\right)^2 e^{\frac{q(\psi_s - E_g/q)}{kT}}. \tag{2.19}$$

It is important to point out the fact that the coefficient of the exponential presents

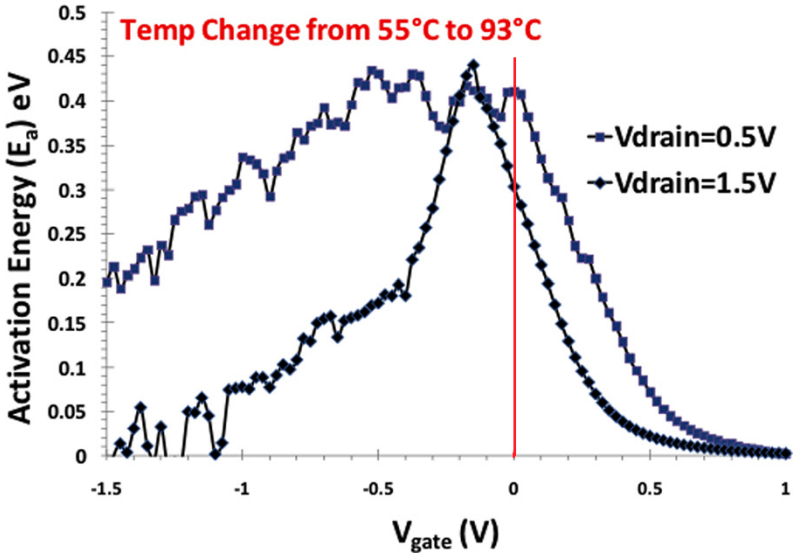


Figure 2.6: Activation energy vs V_{gate} for different V_{ds} (from [16]).

temperature dependeces, while in the Arrhenius analysis the exponential coefficient is usually assumed to be temperature independent. Moreover, since the temperature dependence of E_g is very weak (at room temperature it shrinks by 0.25 meV/K when temperature is increased), the activation energy in the subthreshold region becomes $E_a = E_g - q\psi_s$, regulated by the silicon voltage drop ψ_s. The relation between ψ_s and the gate voltage V_{gate} in the subthreshold region for a planar n-channel MOSFET is known to be linear and this is confirmed by the activation energy plot in Fig. 2.6 in the range of V_{gate} between 0.4 V and -0.1 V that corresponds to the subthreshold region as illustrated in Fig. 2.4.

From the activation energy plot one can understand which is the physical process under analysis from its temperature dependence. For $V_{gate} < -0.1$ V in Fig. 2.6, one can see activation energies that correspond to different GIDL mechanisms depending on the drain bias applied [16]. For $V_{gate} > 0$ V the activation energy decreases toward 0 eV, showing that the above-threshold regime is weakly temperature-dependent as one can see from Fig. 2.4.

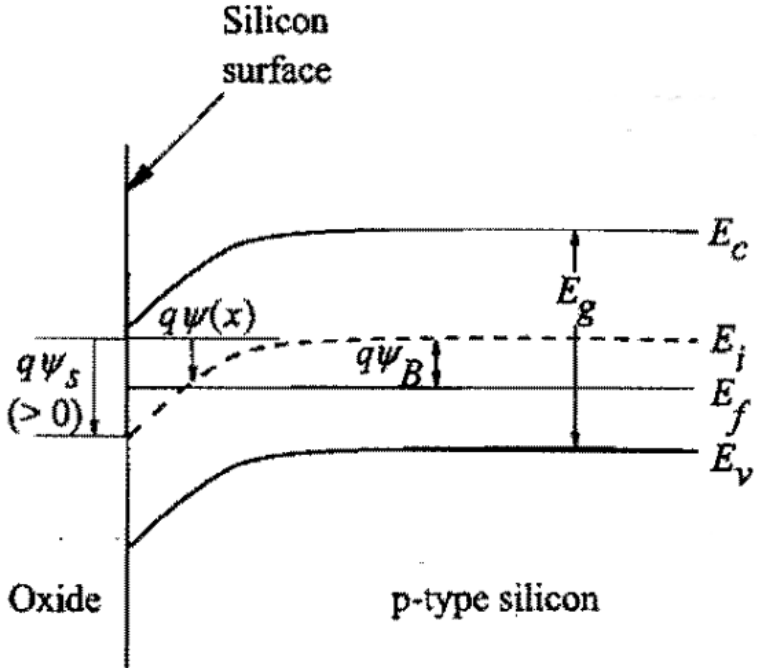


Figure 2.7: Energy bands in the substrate showing the silicon voltage drop ψ_s (from [13]).

2.2 Polycrystalline devices

2.2.1 Grain boundary traps and bulk traps

As written in Sec. 1.3.1 the last memory devices have a polycristalline structure due to their fabrication process. Polysilicon has a granular structure, in which every grain is monocrystalline with a random orientation. Grains are separated by grain boundaries that act as surfaces presenting the features just discussed in Sec. 2.1.5, i.e. every grain boundary is highly defective with the trap states localized in space at the grain boundaries with energy within the forbidden energy gap.

Different measurement methods have been proposed to determine the energies at which traps are located, like the field-effect conductance analysis method [19] [20] (results shown in Fig. 2.8) or by measuring low-frequency capacitance–voltage characteristics and using an extraction algorithm [21] (results shown in Fig. 2.9). It is very important to have precise information about the *trap density of states* since this strongly affects the behavior of polysilicon devices like *thin film transistors* (TFT) and cylindrical *gate-all-around* (GAA) devices.

In Figs. 2.8 and 2.9 a *double exponential* trend is highlighted by means of red straight lines. Since similar distributions have been found in many polyslicon devices, a double

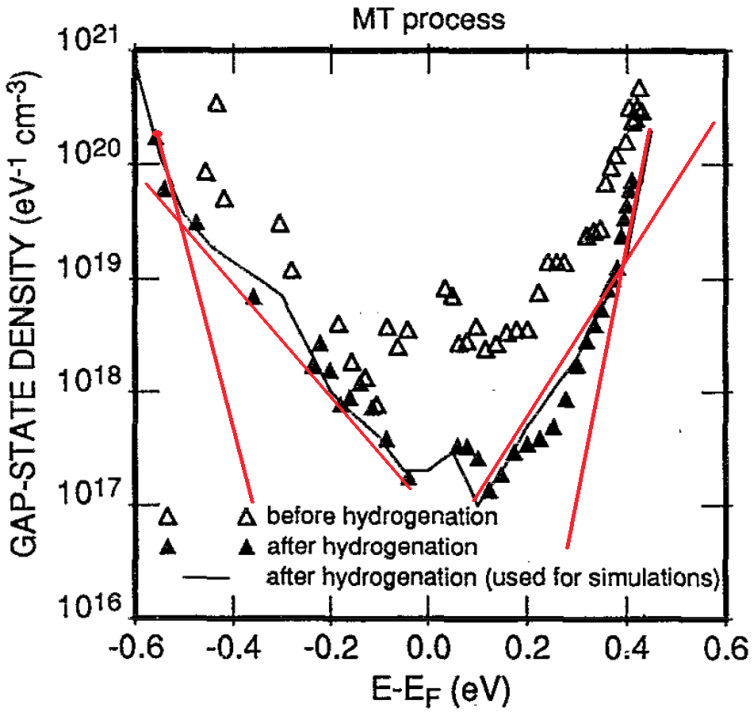


Figure 2.8: Effective density of trap states in the gap showing the differences before and after the hydrogenation process. A double exponential trend is observed as indicated from the red straight lines (from [19]).

exponential distribution like the one shown in Fig. 2.10 has been widely adopted in simulations for polysilicon devices. These two exponential distributions are the so-called *tail traps* and *deep traps* states respectively with the higher peak concentration and lower decay constant for the former and vice versa for the latter. In the upper part of the gap they have an *acceptor* behavior (i.e. they are negative when filled), while in the lower part thay have a *donor* behavior (i.e. they are positive when empty). So for the mathematical form of the acceptor traps (the donor ones are similar but starting from the valence band (Fig. 2.10)):

$$N_t = g_t \times e^{\frac{E-E_c}{\gamma}} + g_d \times e^{\frac{E-E_c}{\lambda}} \tag{2.20}$$

Where $g_t > g_d$ and $\gamma < \lambda$. These distribution will be used throughout all the thesis work.

The occupancy probability of these traps follows the Fermi-Dirac statistics, just like the occupancy probability of the interface traps discussed in Sec. 2.1.5. If the traps are uniformly distributed in space, an integration over all the energies in the gap can be performed to find the spatial trap density:

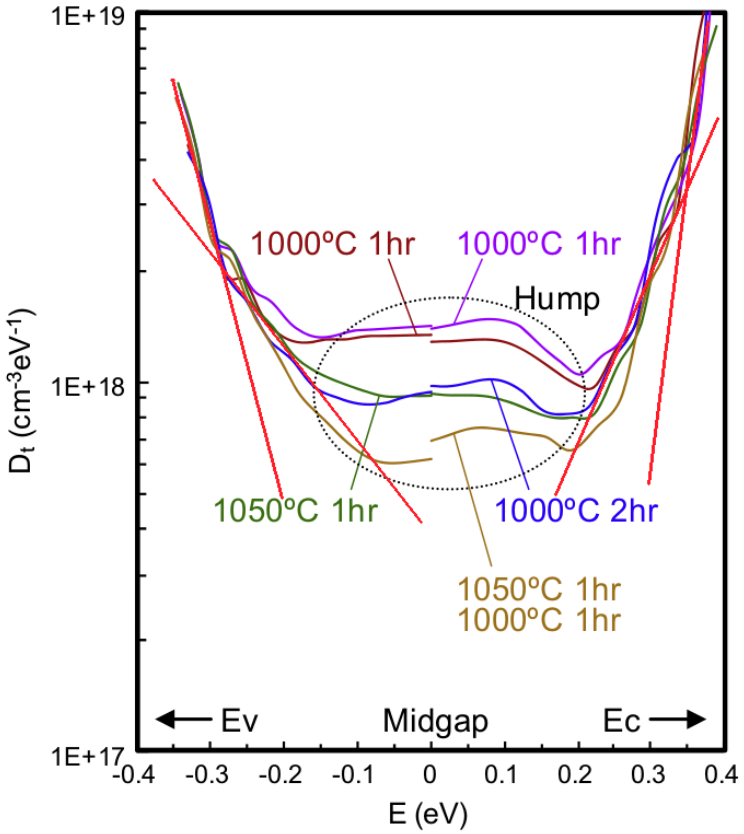


Figure 2.9: Effetive density of trap states in the gap showing the effects of different temperature and annealing time on the humps at midgap. A double exponential trend is observed as indicated from the red straight lines (from [21]).

$$D_t = \int_{E_v}^{E_c} N_t(E) f_D(E) dE. \tag{2.21}$$

After this integration, the problem of dealing with the space localization of the traps, i.e. the position of the grain borders remains. To develop suitable models for the polysilicon devices, a general approach is needed to avoid measurements on every device to find the spatial grain distributions.

2.2.2 Temperature dependence of traps population

It is not possible to obtain a closed analytical form for Eq. (2.21), that must be solved numerically. The product between f_D and the density of states for electrons and the one for traps shows huge differences due to their different localization in the energy spectrum. More specifically, the Fermi level, which is the axis of simmetry of f_D, can be located at

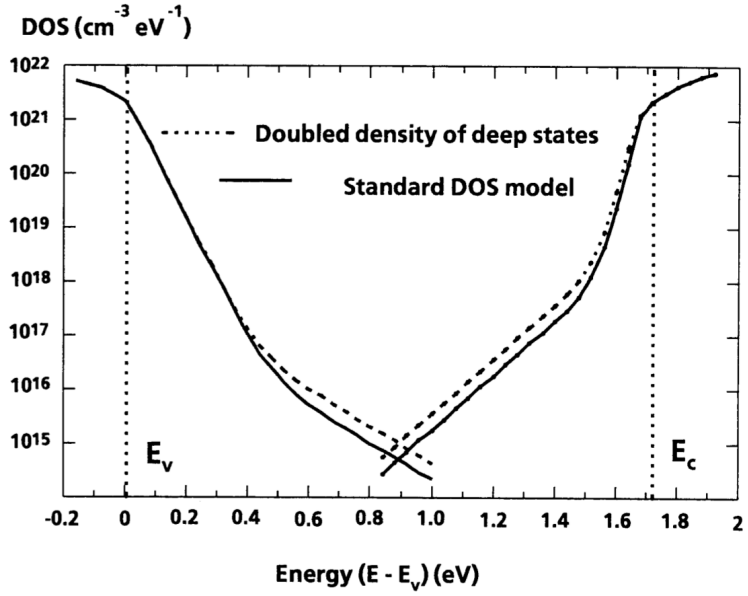


Figure 2.10: Energy distribution of localized trap states in the gap (from [22]).

the conduction band edge in the most extreme case, while the density of states of the free electrons takes values different from 0 only for energies higher than the energy value of the conduction band edge. Therefore, half of the Fermi-Dirac distribution function will be located at energies in which the free electrons DOS is 0 and so will be their product. The trap density of states, instead, covers almost all the energy gap as f_D does and so their product will take non-zero values for all energies in the gap.

The previously-mentioned integral over all the energies of the bandgap has been computed in order to obtain the spatial density of occupied trap states D_t. After this operation, D_t can be integrated over all the volume to obtain the absolute number of occupied traps. The integral in Eq. (2.21) has been computed for:

$$N_t = g_d \times e^{\frac{E-E_c}{\lambda}} \tag{2.22}$$

with $g_d = 5 \times 10^{19}$ cm^{-3} and for $\lambda = 160$ meV and $\lambda = 50$ meV.

The integration has been made only for acceptor-like traps, because the donor-type ones affect only the lower half of the gap which is difficult for the Fermi level to reach in real physical applications, i.e. they usually are well below the Fermi level and therefore they are usually occupied and thus neutral. For further semplification, only deep traps were implemented, because the tail ones have a role only at very low activation energies (i.e. when the Fermi level is very near the conduction band edge) and the focus was on the subthreshold region that corresponds to higher activation energies.

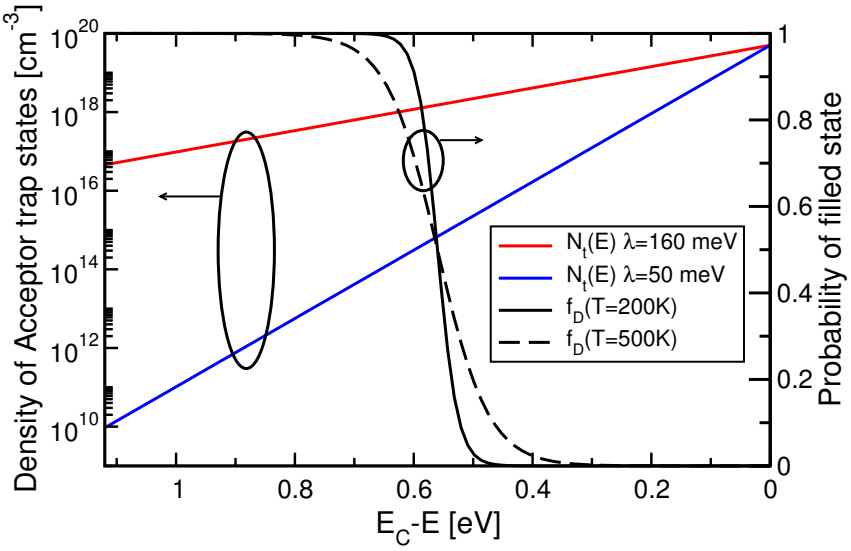


Figure 2.11: Plot of $N_t(E)$ for the two different values of λ and f_D for T=200 K and T=500 K with E_f located at midgap.

The relative difference between the spatial density at 500 K and at 200 K, computed as $D_t(500)/D_t(200)$, depending on the distance $E_c - E_f$ is reported in Fig. 2.12 for two different λ values.

With the help of Fig. 2.11 it is easier to interpret Fig. 2.12. Temperature has the only effect of stretching the Fermi-Dirac distribution function, the transition from 1 to 0 taking place in a range $\simeq 2kT$ wide, much smaller than the energy gap of silicon. When λ becomes smaller, the distribution gets steeper and the temperature effect on the occupation of the traps states gets higher. This fact happens because the total number of occupied traps is mainly determined by all the occupied traps below the energy at which f_D is 1. When the temperature rises, the point at which f_D is 1 shifts towards lower energies, hence the energy range in which all the traps are occupied shrinks. If the trap distribution is steeper, the order of magnitude of $N_t(E)$ decreases faster going towards lower energies and therefore the shrinking of the energy range in which all traps are occupied will have a greater impact on the total number of occupied trap states.

Coming to E_f one can see that the lower the Fermi level, the higher the temperature impact on the occupation of the trap states. That is because when E_f is near the conduction band the vast majority of the states in the gap are occupied and so the variation induced by the change of temperature is less visible on a big baseline of occupied states. Vice versa when the Fermi level is near the valence band the variation is more evident because the total number of occupied traps is lower.

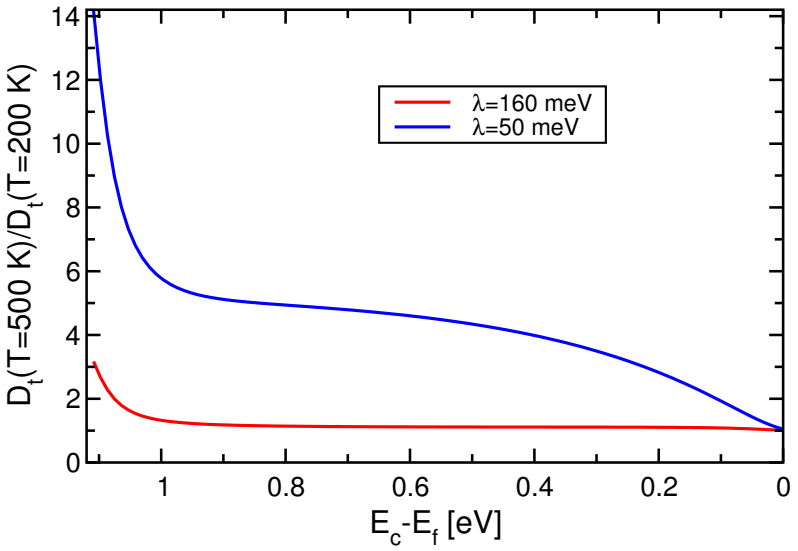


Figure 2.12: Results of the integration in Eq. 2.21 depending on $E_c - E_f$ for different λ values.

2.2.3 Subthreshold slope

In order to analyze the current transport temperature dependence on polysilicon devices we report here results obtained on *Thin-Film-Transistor* (TFT) with different grain boundary concentrations that is here related to the *field-effect-mobility* (FE) at room temperature (the lower FE, the higher the grain boundary concentration).

Speaking about the subthreshold slope with respect to the monocrystalline case, we have to add the capacitance related to the traps at the grain boundaries. In the monocrystalline case in Sec. 2.1.3 the capacitance related to interface traps was neglected because it was assumed to be sufficiently low. To take it into account we have to modify the value of $m = 1 + C_s/C_{ox}$ with:

$$m = 1 + \frac{C_s + C_{it}}{C_{ox}} \tag{2.23}$$

where C_{it} is the capacitance related to the traps located at the interfaces. Therefore we have:

$$SS = \frac{kT}{q} \ln(10) \left[1 + \frac{C_s + C_{it}}{C_{ox}}\right] \left[\frac{\text{mV}}{\text{dec}}\right] \tag{2.24}$$

This is consistent with Fig. 2.13 that shows how higher grain boundaries concentrations (i.e. lower FE), which have higher trap concentrations and in turn higher C_{it}, are found to have higher subthreshold slopes.

Moreover, higher grain boundaries concentrations (or trap densities) lead to higher

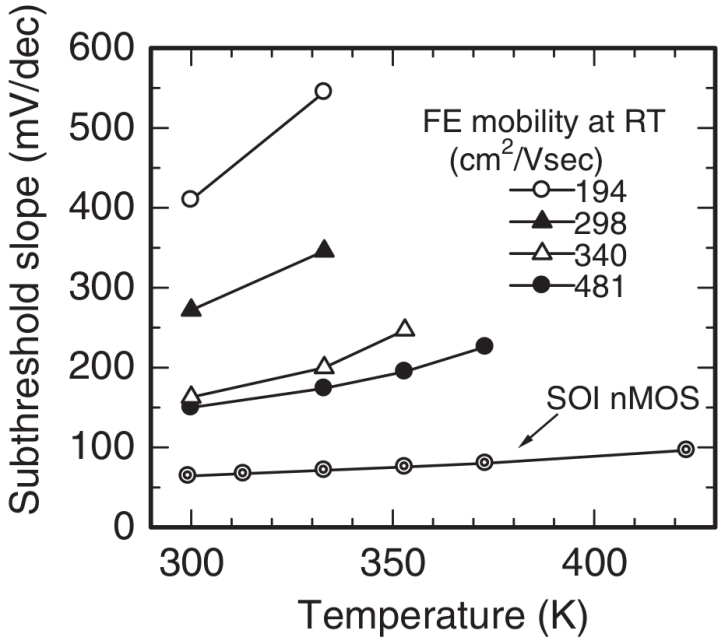


Figure 2.13: Temperature dependence of subthreshold slope in TFTs with different field-effect mobilities at room temperature (from [23]).

temperature dependences. This can be explained by the fact that if E_f is fixed, when the temperature is raised the total number of occupied traps (or accessible states) gets higher as shown in the previous section, bringing higher C_{it} and in turn higher SS.

Polysilicon devices are then found to have higher subthreshold slopes and higher temperature dependences of the subthreshold slopes due to the capacitance added by the presence of traps states.

2.2.4 Threshold voltage

The typical drain current I_D and transconductance g_m in a TFT are shown in Fig. 2.14a as a function of the gate voltage with temperature as a parameter. The effect of the higher temperature is to shrink the subthreshold region by moving to higher bias the beginning of the subthreshold region and to lower bias the above-threshold region. In devices without trap states dV_t/dT is known to be negative, it is important to point out, as shown in Fig. 2.14b, that in the polysilicon case the derivative is higher with respect to the case without trap states.

In order to explain what is happening we have to consider that we are no more considering a constant E_f with temperature as we have done in the previous cases.

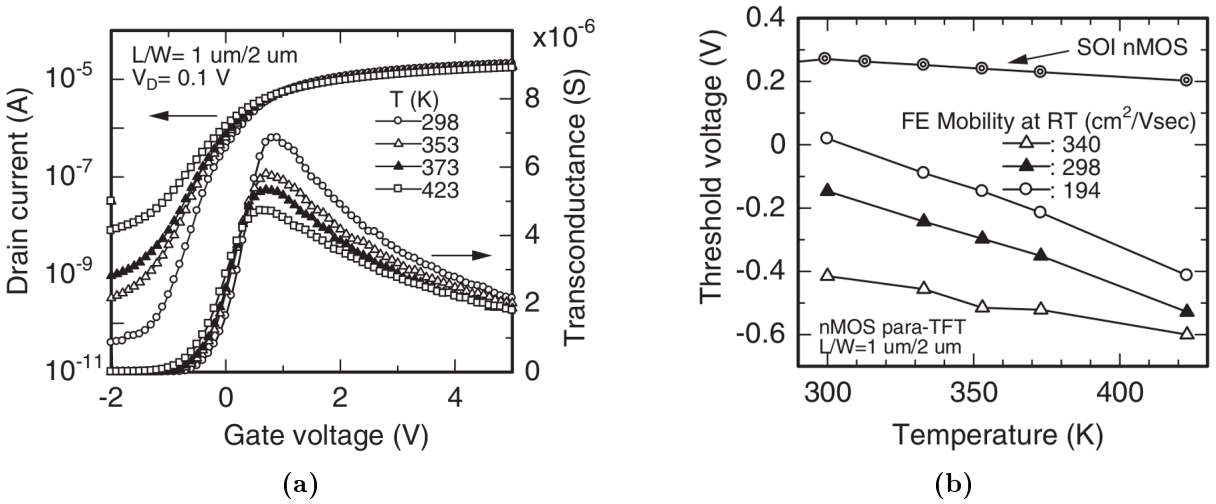


Figure 2.14: (a) Typical drain current and transconductance as function of gate voltage in TFT with temperature as parameter (b) Temperature dependence of threshold voltage in TFTs with different field-effect mobilities at room temperature. SOI nMOS data are also shown as a reference (from [23]).

We then refer to the bands depicted in Fig. 2.15. Fig. 2.15 (a)-(c) shows the bands at $V_G = V_T$ along with the surface potential barrier ϕ_{B1} induced by the charge trapped at the grain boundary for $T = T_1$ (similar diagrams are illustrated in Fig. 2.15 (d)-(f) at $T = T_2 > T_1$). In Fig. 2.15 (a) and (d) it is clearly shown that $q\phi_f(T_1) > \phi_f(T_2)$, i.e the Fermi level rises with increasing temperature at fixed bias: we are thus analyzing a different situation from Sec. 2.2.3 in which the Fermi level was fixed while temperature changes.

The rise of the Fermi level means that the distance between E_f and E_i decrease with rising temperature (Fig. 2.15 (b) and (e)). From the same figures one can see that the fact that this distance is reduced turn into less charge trapped in the interface states. As a result the potential barrier created by the charge trapped at the grain boundary is lower (Fig. 2.15 (c) and (f)).

It is known that the threshold condition in TFTs is related to the lowering of the grain boundary barriers. First, the inversion condition is reached within every single grain and then the potential barriers are lowered until a threshold value at which the device pass to the on-state is reached [24]. Therefore the threshold condition will be met earlier with increasing temperature and, because of its relation with the trap density, this dependence is enhanched with higher trap densities.

Polysilicon devices are then expected to have a higher threshold voltage dependence on T with respect to monocrystalline silicon devices. This dependence gets higher with

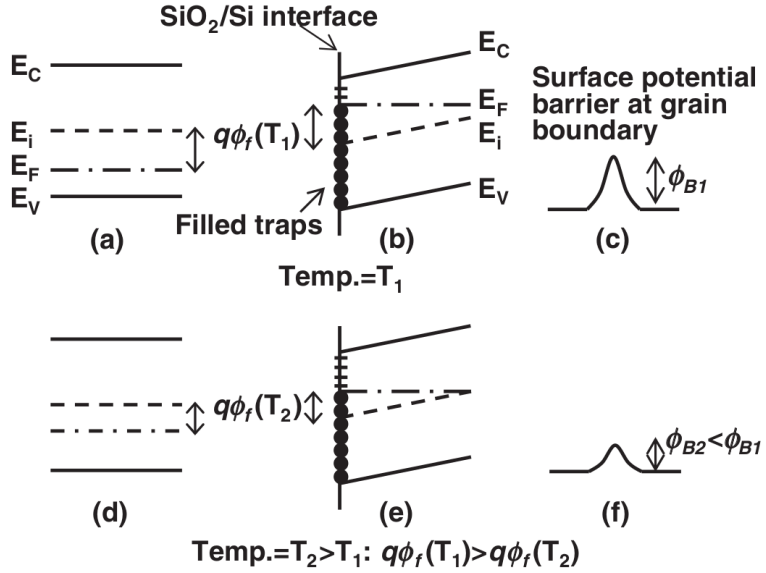


Figure 2.15: (a, d) Schematic band diagram of p-type silicon, (b, e) band diagram showing depth near surface including GB-related interface traps in nMOS-TFT at $V_G = V_T$, and (c, f) surface potential barrier at GB at $V_G = V_T$. (a)–(c) at $T = T_1$, and (d)–(f) at $T = T_2 > T_1$ (from [23]).

the rising of grain boundaries and trap densities.

2.2.5 Activation energy

In order to present the changes induced in the activation energy by the presence of the grain boundaries we analyze the performance of two TFTs, one that substained an hydrogenation process while the other is unhydrogenated. As mentioned in Sec. 2.1.5 the hydrogenation process is useful to reduce the total number of trap states, especially the density of traps energetically located at midgap (examples are shown in Figs. 2.8 and 2.9).

The activation energy is reported in Fig. 2.16. Activation energies are extracted for a drain bias of 0.1 V and for a temperature range from 0 to 100 C. They are extracted following the Arrhenius analysis process reported in Sec. 2.1.6.

We have already said that charge trapping at grain boundaries gives rise to potential barries which impede the transport of carriers across them. Without potential barriers at the grain boundaries, the activation energy for the drain current depends only on the difference between the energy gap and the silicon voltage drop (surface potential) as shown in Sec. 2.1.6. If a barrier forms at the grain boundary, then the activation energy is approximately given by the sum of the energy difference mentioned above, plus the

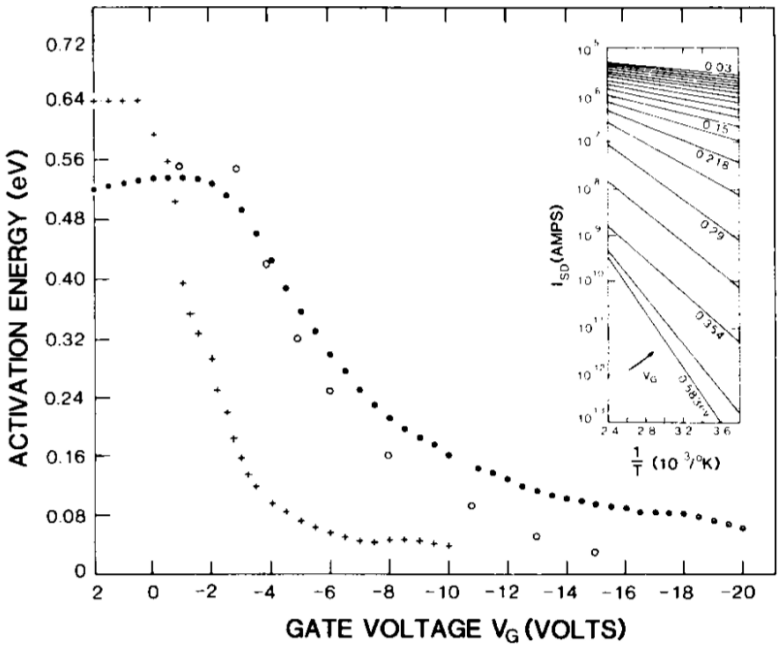


Figure 2.16: Activation energy versus gate voltage for 3 different TFT's. Filled circles-are for unhydrogenated TFT with an approximate grain size of 1200 A. Crosses are for hydrogenated TFT with the same grain size. Open circles are from [25] for a TFT with a grain size of 1200 A. Data for each device were generated from curves similar to those shown in the inset (from [26]).

barrier height. This sum assumes that the carriers must surmount this barrier by some emission process, that has been demonstrated elsewhere [27] to be thermionic emission, which depends exponentially on the barrier height. Hence the addition of a term in the activation energy.

Therefore, the activation energy is expected to have a slower decrease with increasing gate voltage as the number of trap states gets higher. That can be explained because higher concentration of trap states and thus higher localized charges trapped in these states give rise to higher potential barriers to be overcome and in turn higher energy contributions to the activation energies derived from such barriers. This is confirmed by Fig. 2.16 where unhydrogenated TFT, that has higher trap densities, shows a much slower activation energy decrease with respect to the hydrogenated TFT.

This also means that the Fermi level is pulled toward the band more slowly, because while it approaches the band edge, not only mobile carriers are induced in the band but also fixed charges have to be trapped in the localized states. Higher gate voltage is then needed to obtain the same current of a TFT with lower trap concentration.

We can conclude that the presence of high trap concentrations at grain boundaries

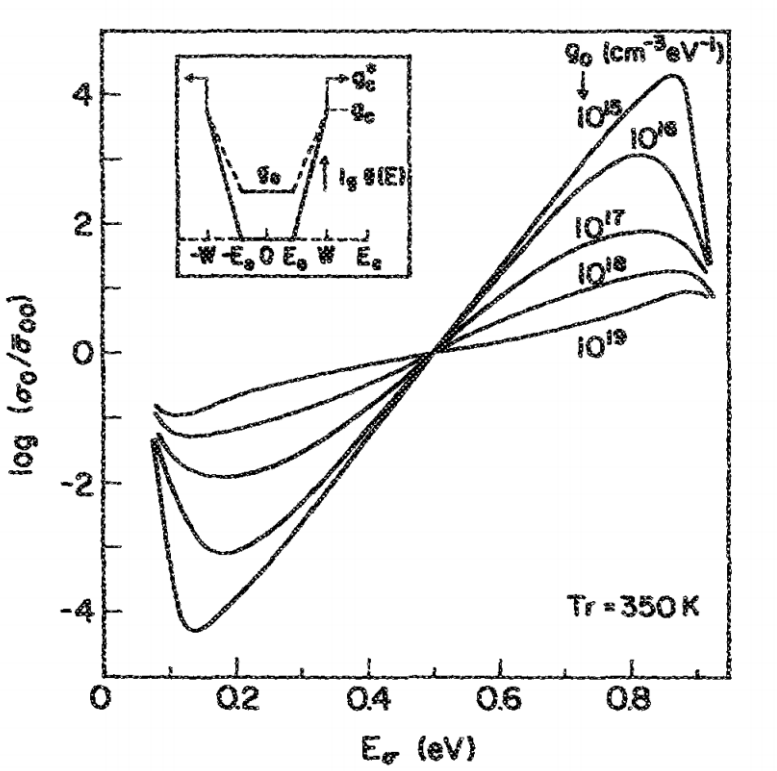


Figure 2.17: Semilogarithmic plot of normalized conductivity prefactor vs activation energy for the symmetric *g(E)* shown in the inset with different g_0 values. In the inset $E_c = 0.5$ eV, $W = 0.4$ eV, $E_0 = 0.1$ eV and g_c^*, g_c, g_0 are $10^{21}, 10^{20}, 10^{15} - 10^{19}$ cm^{-3}eV^{-1} respectively. (from [29]).

strongly affects the activation energy of the source to drain current because the presence of such traps gives rise to potential barriers that have to be overcome by thermionic emission, adding a term to the activation energy. The slope of the E_a vs V_G curve will therefore be lower.

2.3 Meyer-Neldel Rule

2.3.1 Origin and formulation

The Meyer-Neldel rule (MNR), also known as *compensation rule* named after the scientists that first observed it in 1937 [28], is an empirical relation between the prefactor and the activation energy of a thermally-activated process.

The rules states that if a process X obeys the equation:

$$X = X_0 e^{-E_a/kT} \tag{2.25}$$

X_0 and E_a obey the equation:

$$\ln X_0 = X_{00} + E_a/E_{MN} \tag{2.26}$$

where X_{00} is a constant and E_{MN}, defined as the Meyer-Neldel energy, is the the inverse of the slope of the prefactor X_0 in a semilogaritmic plot against the activation energy. The Meyer-Neldel ennergy is an important parameter of the rule, the different physical explanations of the phenomena that follow the rule aim to find a relationship between this energy and the other physical parameters of the system under analysis.

This rule is valid in a large quantity of materials and processes, and several microscopic explanations have been proposed in the last decades, some of which are briefly summarized hereafter.

2.3.2 Proposed MNR explanations

1. *Microscopic Explanation:* in a work published in 1990 [30] it is proposed that the MNR can be understood as arising naturally for assisted processes in which the energy exchanged is much larger than that of a typical excitation, so that the total required energy is provided by multiple excitations. Semiconductor current conduction is one of these processes since it is far more likely that the energy required for an electron to populate the conduction band comes from multiple phonons instead of from only one large phonon. The number of ways of assembling these phonons will obviously increase as the total energy barrier increases. The prefactor has then to be proportional to the number of ways of assembling these excitations to obtain the MNR.

   The mentioned article can be seen as a phenomenological approach widely taken by many aythors. It considers, as in the Eyring theory [31], that in a reaction the rate is equal to $X_0 e^{-\delta G/kT}$, where δG is the difference between the free energy of the state at the peak of an activation barrier and that of the initial state and X_0 is a constant. Since:

   $$\delta G = \delta H - T\delta S, \tag{2.27}$$

   where H is the enthalpy, S the entropy, T the temperature, it is evident that:

$$X = X_0 e^{\delta S/k} e^{-\delta H/kT}. \tag{2.28}$$

Thus the MNR follows if the activation entropy δS is proportional to the activation enthalpy δH. As previously said, if the activation energy is large compared to typical excitations of the system, it becomes necessary to assemble many excitations for the reaction to take place. As the entropy is proportional to the logarithm of the numer of different ways of assembling this excitations, δS will increase as δH increases. The MNR will follow if the two are proportional to each other.

In the mentioned article the authors discuss as an example in which the energy needed to surmount the enthalpy barrier to a physical process δH is obtained from multiphonon annihilation. In a simple Einstein model of the phonon spectrum, with each quantum energy E_E, $n = \delta H/E_E$ phonons must be annihilated in the excitation. Assuming N phonons lie within the interaction volume from which they can be annihilated, the dimensionless entropy change associated with the excitation over the barrier is the natural logarithm of the number of ways of assembling n out of N interacting phonons. Thus:

$$\delta S/k = \ln\left[\frac{N!}{n!(N-n)!}\right] \approx N \ln\left[\frac{N}{N-n}\right] + n \ln\left[\frac{N-n}{n}\right], \tag{2.29}$$

where the second expression is obtained from the application of Stirling's approximation. For $n \ll N$:

$$\delta S/k \approx n \ln \frac{N}{n} \approx \frac{\delta H}{E_E} \ln N \tag{2.30}$$

from which the desired proportionality is obtained. In fact combining Eqs. (2.28) and (2.30) one obtains:

$$X = X_0 e^{\delta H(\ln N)/E_E} e^{-\delta H/kT} \tag{2.31}$$

where, if N is independent of δH, we obtain exactly the MN rule with $E_{MN} = E_E/\ln N$.

2. *Fermi statistical shift:* in a work published in 1988 [29] the MNR in tetrahedral amorphous system is interpreted as naturally arising from the Fermi statistical shift, i.e. the fact that, when the temperature is changed, the Fermi level moves inside

the energy gap to respect the charge law conservation. A couple of years before it was demonstrated [32] that the Fermi statistical shift could reproduce the MNR for specific gap density of states spectrum *g(E)*. The aim of the mentioned work [29] was to show that the fact that the conductivity σ respects the MNR is the natural consequence of the presence of a deep well in the density of states *g(E)* inside the gap.

The demonstration starts by considering a system with steep walls in *g(E)* spectrum inside the gap. When temperature is raised the Fermi level always moves towards the center of the gap due to the Fermi statistical shift. In this situation also the Fermi temperature derivative dE_f/dkT changes considerably with temperature and is quite sensitive to the position of the Fermi level.

The conductivity can be expressed as $\sigma = \sigma_{00} exp(-(E_C - E_f)/kT)$, where σ_{00} is a constant. If define $E_C = 0$ for convenience, we can plot:

$$\ln \frac{\sigma}{\sigma_{00}} = \frac{E_f(kT)}{kT} \tag{2.32}$$

and the the activation energy E_a can be written as:

$$E_a = \frac{d\sigma}{d(1/kT)} = -E_f + kT\frac{dE_f}{dkT}. \tag{2.33}$$

We can finally write the conductivity as:

$$\sigma = \sigma_0 e^{-E_a/kT} \tag{2.34}$$

where:

$$\sigma_0 = \sigma_{00} e^{dE_f/dkT} \tag{2.35}$$

We can see from Eqs. (2.33), (2.34) and (2.35) that if dE_f/dkT linearity to E_f holds than we have dE_f/dkT linearity to E_a as required from the rule. In the mentioned work this linearity between the Fermi level and its derivative is demonstrated to be a consequence of the Fermi statistical shift in systems where *g(E)* presents steep walls.

Therefore the MNR rule arises from the shift of the Fermi level in the gap which is caused from the steep walls of the *g(E)* spectrum. The model shows also a relation between E_{MN} and the density of gap states *g(E)* (Fig. 2.17).

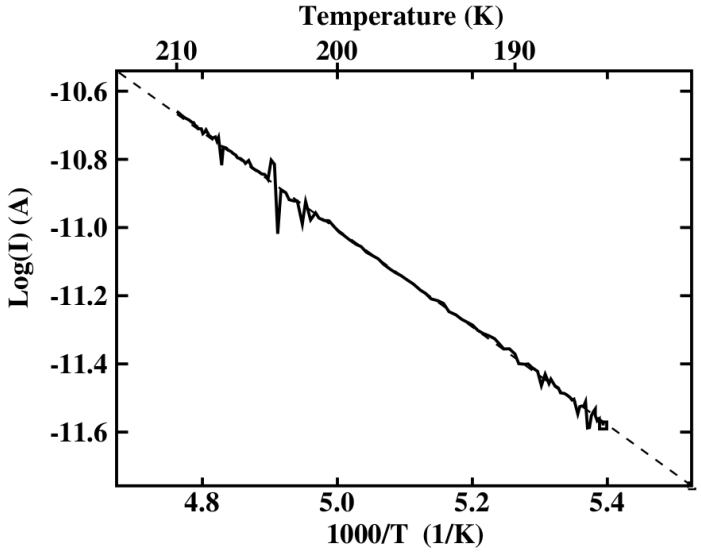


Figure 2.18: Arrhenius plot of the current with bias $V_{ds} = -0.5V$ and $V_G = -10.5V$ (from [33]).

3. *Trap states:* it has been shown that, whenever traps, distributed exponentially in energy, are governing the conduction in electrical materials, a Meyer-Nelder observation is expected [33]. The demonstration uses the model of Shur and Hack [34]. Measurements and observations are performed on a TFT FET of α-sexithiophene ($\alpha - T6$) in the linear region (Arrhenius plot shown in Fig. 2.18). The traps follow a model with the traps density that presents an exponential decrease as shown in Sec. 2.2.1.

   An evidence of the MNR is the presence of a temperature, known as the isokinetic temperature T_{MN}, where the dependence of the current on bias disappears. From the model of Shur and Hack it is possible to derive a relation between current and traps density, i.e:

$$I_{ds} = \frac{q\mu_0 W}{L} f(T, \lambda/k) \left[C_{ox}(|V_g - V_t|)\right]^{2\lambda/kT-1} V_{ds} \tag{2.36}$$

   where $f(T, \lambda/k)$ is:

$$f(T, \lambda/k) \approx N_V e^{-E_f/kT} \frac{kT\epsilon}{q} \left(\frac{\pi kT/\lambda}{2\pi\epsilon\lambda T g_{F0}}\right)^{\lambda/kT}. \tag{2.37}$$

   In these two equations μ_0 is the band mobility, λ the exponential decay constant of the traps distribution as shown in Eq. (2.22), N_V the effective band density of states and g_{F0} the density of deep localized states at the Fermi level. The relation

between the current and the traps density is contained in the parameter λ.

From Eq. (2.36) one can verify the presence of the isokinetic temperature $T_{MN} = 2\lambda/k$. Hence the authors state that there is a relation between the Meyer-Neldel parameter T_{MN} and the trap density distribution, such that measuring the MN parameter through the current could yield information about the trap distribution. They also state that MNR may arise in other system, from different conditions and thus that the presence of traps is not a necessary condition for finding the rule.

We can conclude that there is no general agreement on the physical reasons for this general rule which regulates a large number of quite different systems, the first has a general approach speaking of elementary excitations, while the second two explanations reported link it to the structure of the density of states in the gap.

2.3.3 Example of Meyer-Neldel analysis on TFTs

The interest in this rule applied to semiconductor conduction arose because, as presented in the Sec. 1.4, new device structures need polysilicon in place of the crystalline silicon. In such structures, high concentrations of traps at interfaces and grain boundaries are present and many evidence of relations between the presence of trap states and the rule have been found. Here we report a very recent MNR analysis [35] performed on three different TFTs, one amorphous and two polycrystalline with different doping (one n-type and one p-type) with the aim of outlining the analysis that will be performed on all the devices throughout the thesis work.

In the reported work the effective medium approximation [22] is assumed, taking into account only acceptor traps with a double exponential distribution like Eq. (2.20):

$$g_{TA}(E) = g_{cd} \times e^{\frac{E-E_C}{E_d}} + g_{ct} \times e^{\frac{E-E_C}{E_t}} \tag{2.38}$$

the values of the various parameters for the 3 different TFTs are reported in Fig. 2.19. The spatial traps densities N_t and N_d, respectively tail and deep traps, can be obtained by integrating over all the energies of the gap as in Eq. (2.21) where the only parameter is E_f.

	a-IGZO	MILC	ELA
g_{cd} ($cm^{-3}eV^{-1}$)	4.4×10^{17}	4.8×10^{19}	2.3×10^{19}
E_d (eV)	0.17	0.09	0.062
g_{ct} ($cm^{-3}eV^{-1}$)	6×10^{17}	1×10^{20}	8×10^{19}
E_t (eV)	0.016	0.01	0.029
μ_b ($cm^2/V/s$)	100	80	82
ψ_B (V)	0.1	0.14	−0.01
V_{FB}(V)	−2.4	−1.6	1.1
N_C (cm^{-3})@300K	4.8×10^{18}	2.86×10^{19}	-
N_V (cm^{-3}) @300K	-	-	2.66×10^{19}
$\varepsilon_s/\varepsilon_0$	12.3	11.7	11.7
SS(V/dec)	0.23	1.48	0.48
V_{th}(V)	2.1	6.3	−3.1

Figure 2.19: Model parameters and fundamental electrical parameters for three kinds of TFTs (from [35]).

The density of free electrons n can be obtained by integrating the DOS distribution of the conduction band $g_n(E)$:

$$n(E_f) = \int_{E_C}^{\infty} g_n(E) f_D(E) dE \tag{2.39}$$

where:

$$g_n(E) = \frac{2N_C}{\pi^{1/2} kT^{3/2}} \sqrt{E - E_C} \tag{2.40}$$

with N_C from Fig. 2.19.

Based on the gradual channel approximation, the 1-D Poisson's equation along the channel depth (x) can be expressed as:

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{q}{\epsilon_s}(N_d + N_t + n) \tag{2.41}$$

where ψ is the potential and ϵ_s is the dielectric constant of the semiconductor channel. Referring to Fig. 2.20 we define ψ_s as the potential drop between the Si-SiO_2 interface and the substrate where the bands are flat. As already explained, when the bands bend with the gate bias, E_f remains constant, but the values of N_t, N_d and n change because they depend on the distance between E_f and E_C (or E_V for the holes of p-channel TFT). Mathematically we have $\psi(x) = (E_f - E_i - q\psi_B)/q$, where $q\psi_B$ is the difference between E_f and E_i at the back channel. Using $F(\psi) = \left[2\int_0^{\psi}\left(\frac{\partial^2\psi}{\partial x^2}\right)d\psi\right]^{1/2}$ with F electric field, and substituting E_f with $\psi(x)$, trapped and free charges can be obtained as a function of the potential. Following the Pao-Sah model, the sheet density of trapped (Q_d and Q_t for

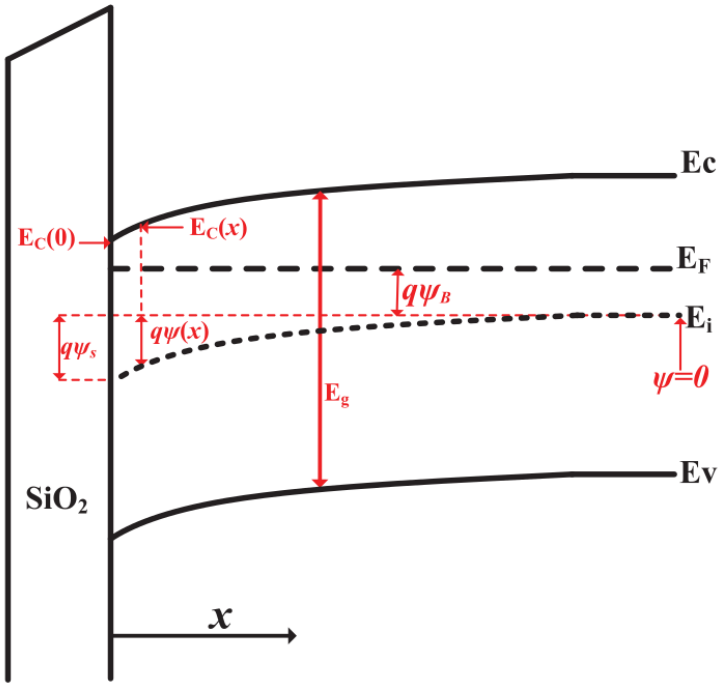


Figure 2.20: Energy bands diagram of TFTs along the channel depth and the definition of energy bands parameters (from [35]).

deep and tail traps) and free charges Q_f in the channel can be obtained as:

$$Q_d(\psi_s) = q\int_0^{\psi_s} \frac{N_d}{F(\psi)}d\psi \tag{2.42}$$

$$Q_t(\psi_s) = q\int_0^{\psi_s} \frac{N_t}{F(\psi)}d\psi \tag{2.43}$$

$$Q_f(\psi_s) = q\int_0^{\psi_s} \frac{n}{F(\psi)}d\psi \tag{2.44}$$

the sum of the densities above is the total sheet charge density in the channel $Q_i(\psi_s)$.

Using Gauss's law and knowing that the voltage in the channel is $V_{gc} = V_{gs} - V_{cs}$, where V_{gs} is the applied gate voltage and V_{cs} is the local channel potential, one can obtain Q_f as a function of V_{gc} or V_{cs}, although these are not explicit functions and have to be solved numerically for any given V_{gc}. So the drain current is finally derived as:

$$I_d = \mu\frac{W}{L}\int_0^{V_D} Q_f(V_{cs})dV_{cs} \tag{2.45}$$

where μ is the band mobility for electrons. The derived model fits very well with the measurements.

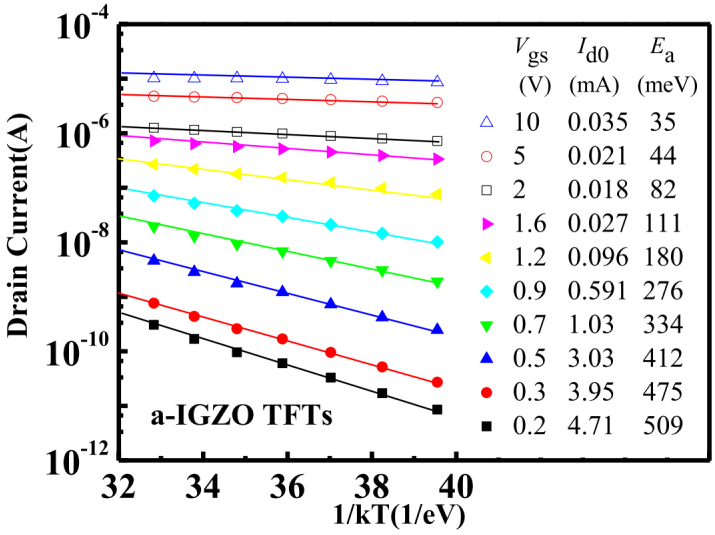


Figure 2.21: Arrhenius plot of I_{ds} for different V_{gs} values for a-IGZO TFTs. The value of the slope E_a and the intercept I_{d0} are also shown (from [35]).

As presented in Sec. 2.1.6, to perform the Arrhenius analysis after obtaining the I-V relations at different temperatures one samples all of them at chosen biases. For every bias the sampled points will be plotted in a semilogarithmic graph against $1/kT$ obtaining the Arrhenius plot of the current as shown in Fig. 2.21. If the current relation is of the type of Eq. (2.25) the logarithm will yield:

$$\ln I_{ds} = \ln I_{d0} - \frac{E_a}{kT} \tag{2.46}$$

hence obtaining straight lines in which the slope will be $-E_a$ and the intercept $\ln I_0$.

If one assumes that the prefactor I_0 takes a form like:

$$I_{d0} = I_{d00} e^{E_a/E_{MN}} \tag{2.47}$$

applying another natural logarithm and making a graph will lead to a form like Eq. (2.26), such that the inverse of the slope in the graph will be the Meyer-Neldel energy E_{MN} that is a constant that defines the relation, Fig. 2.22 illustrates the Meyer-Neldel plot for the three analyzed TFTs.

Referring to the model here reported some approximations can be made to the analytical formulas in order to find meaningful relations between some physical parameters and E_{MN} considering that the Meyer-Neldel rule holds in the subthreshold region.

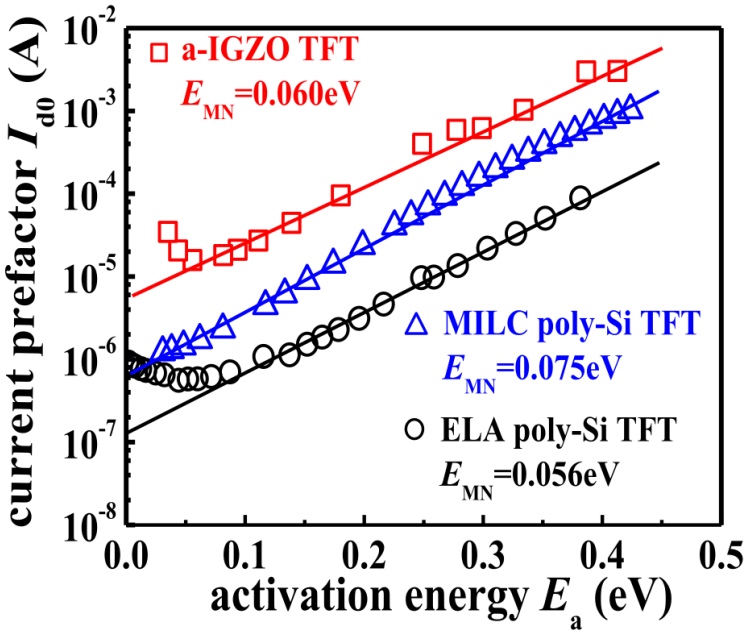


Figure 2.22: Dependence of I_{d0} on E_a for the three analyzed TFTs in log scale. The values of the respective E_{MN} are shown. The lines are the linear fits for the extraction of the Meyer-Neldel energy (from [35]).

1. With the 0 K approximation, i.e. all states above E_f are empty and all below are occupied, N_d can be simplified as $N_d \simeq g_{cd} \times E_d \times e^{\frac{-(E_C - E_f)}{E_d}}$.

2. With the Maxwell-Boltzmann approximation the electron density can be approximated with $n = N_C \times e^{\frac{-(E_C - E_f)}{kT}}$.

Noting that E_c is a function of x (Fig. 2.20), the sheet density of trapped charge as in Eq. (2.42) can be obtained by integrating N_d along the channel depth x over the channel thickness t_{ch}. Using the approximation just presented one obtains:

$$Q_d = q \int_0^{t_{ch}} g_{cd} \times E_d \times e^{\frac{-(E_C(x) - E_f)}{E_d}} dx \tag{2.48}$$

and defining $f(x) \equiv e^{-(E_C(x) - E_f)/E_d}$, according to the mean value theorem for integration there always exist $x_1 \in (0, t_{ch})$ so that $f(x_1)$ can be taken out of the integral, obtaining:

$$Q_d = q g_{cd} E_d f(x_1) t_{ch} \quad x_1 \in (0, t_{ch}) \tag{2.49}$$

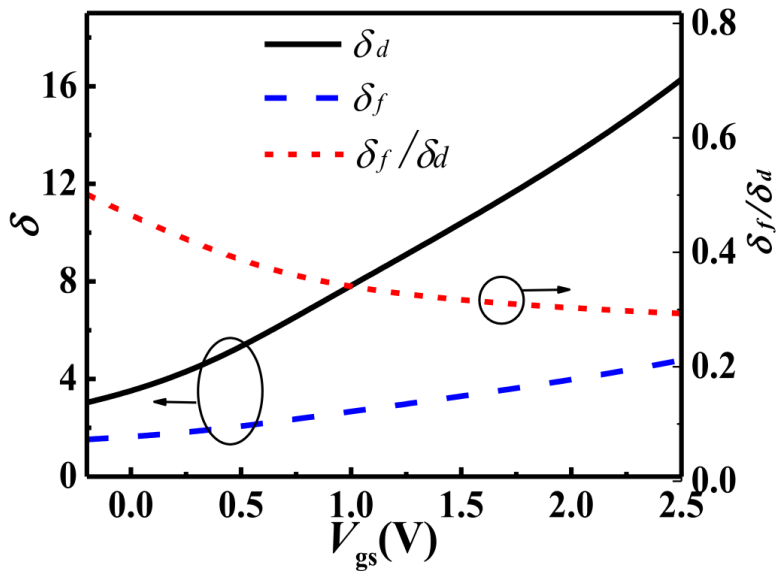


Figure 2.23: Calculation result of δ_d, δ_f and ratio for a-IZGO TFTs in the subthreshold region. (from [35]).

Similarly one can integrate n over the channel thickness, and define a function $g(x) \equiv e^{-(E_c(x)-E_f)/kT}$ such that with the mean value theorem for integration one obtains:

$$Q_f = qN_C g(x_2) t_{ch} \quad x_2 \in (0, t_{ch}) \tag{2.50}$$

It is now useful to define two functions:

$$\delta_d = \frac{E_C(x_1) - E_f}{E_C(0) - E_f} \quad \delta_d > 1 \tag{2.51}$$

and

$$\delta_f = \frac{E_C(x_2) - E_f}{E_C(0) - E_f} \quad \delta_f > 1 \tag{2.52}$$

Knowing that for small V_D Q_f is uniform along the channel length so that Eq. (2.45) can be reduced into $I_d = \mu(W/L)V_d Q_f$, and with further considerations and approximations about the ratios between Q_f, Q_d and Q_t and some algebra reported in [35] one can finally write I_d in the following form:

$$I_d = I_{d0} e^{-\delta_f (E_C(0) - E_f)/kT} \tag{2.53}$$

$$I_{d0} = I_{d00} e^{\delta_d (E_C(0) - E_f)/E_d} \tag{2.54}$$

$$I_{d00} = \mu C_{ox} V_D \frac{W}{L} \frac{N_C}{g_{cd} E_d} \left[V_{gs} - V_{FB} - \left(\frac{E_g}{2q} - \psi_B - \frac{E_C(0) - E_f}{q} \right) \right]. \tag{2.55}$$

One should note that this is an expression consistent with the Pao-Sah model and it is a MNR rule form, where the key parameters are:

$$E_a = \delta_f (E_C(0) - E_f) \tag{2.56}$$

and the MN energy E_{MN} is:

$$E_{MN} = \delta_f E_d / \delta_d \tag{2.57}$$

The consistency between the Pao-Sah model and the Meyer-Neldel rule has then been demonstrated. The Meyer-Neldel energy is then related to three parameters: E_d, δ_d and δ_f. The first one has a simple meaning, is the slope of the deep traps DOS distribution while the other two are more complicated. They brought information related to the band bending. From Eqs. (2.51) and (2.52) one can see that they depend on the gate voltage, this may take to infer that the Meyer-Neldel energy is not constant in this model, but E_{MN} depends on their ratio and in Fig. 2.23 is shown that their ratio is a slowly varying function.

To conclude, a Meyer-Neldel relation has been obtained starting from the Pao-Sah model. Meyer Neldel energy is related to the trap DOS distribution and to the ratio between trapped and free charges and how this affects the bands bending. In the next chapters a similar analysis will be performed on different devices in order to understand if the rule stems from the traps contribution or other contributions are present.

Chapter 3

Meyer-Neldel rule in monocrystalline devices

In this chapter the presence of the MNR in monocrystalline devices without trap states will be deeply investigated. This choice has been made to verify the hypothesis that the MNR is related to the presence of trap states in the gap. In the beginning the physical models implemented are illustrated, followed by the introduction to the nanowire MOSFET that has been simulated and the approximations adopted. The Arrhenius analysis will then be exposed and the underlying physical processes will be depicted. Moreover the results of the simulations will be compared with the results of analytical formulas from the literature. At the end of the chapter the same analysis will be performed on the well know planar MOSFET to clarify the role of the cylindric geometry.

3.1 Nanowire MOSFET

3.1.1 TCAD simulation setup and physical models

The results illustrated from now on were achieved using the tools provided by *Synopsys Sentaurus TCAD* [36] and with the aid of *MathWorks MATLAB*.

First the simulated string structure is defined with *Sentaurus Structure Editor* (SDE), then a mesh, for the finite element method, is generated with *Sentaurus Mesh* (SNMESH), the simulations were finally performed using *Sentaurus Device* (SDEVICE). The use of a commercial product such as the Sentaurus TCAD suite simplified the implementation of the physical models described in the following, allowing a quick setup of the project environment, which would not have been possible with an in-house solution.

Figure 3.1: Cylindrical nanowire MOSFET simulated by TCAD. Figure not in scale.

All the devices presented in this chapter are simulated without trap states: neither at the $Si - SiO_2$ interface, neither distributed in the bulk to simulate the grain boundaries' traps, i.e the *effective medium approximation* that will be explained in the next chapter. The aim of neglecting all the trap states is to verify in the first place if the MNR is present only in devices with localized states in the gap as presented in the previous chapter.

Along with the trap states other physical effects will be neglected in the analysis:

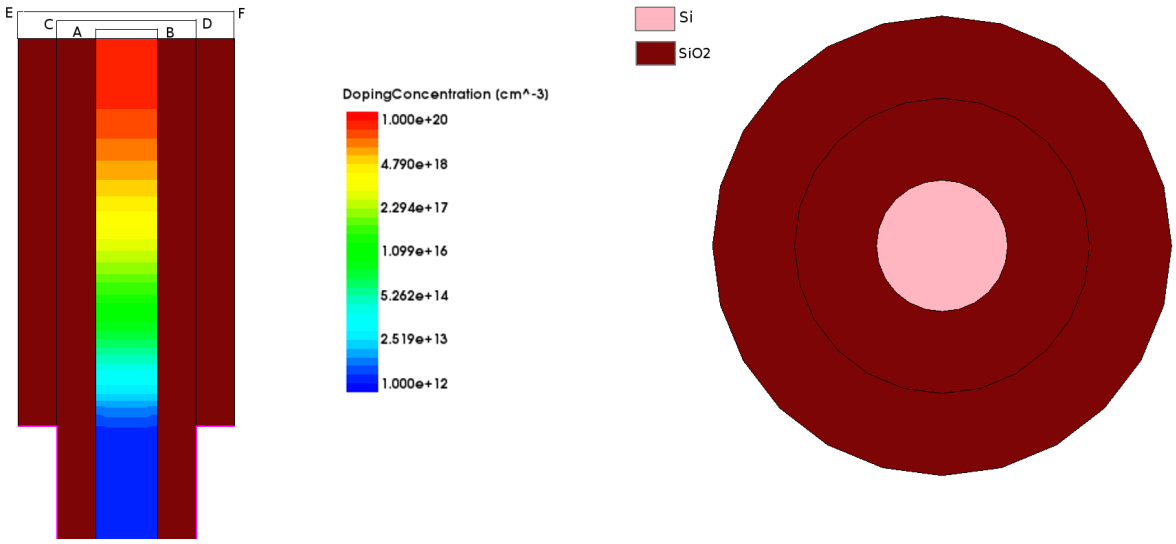
1. *Bandgap narrowing*: the silicon bandgap has been assumed constant and equal to $E_g = 1.124$ eV; its dependences on temperature and on heavy doping have been neglected.
2. *Incomplete ionization*: all dopants are assumed to be ionized at every temperature.
3. *Quantum mechanics*: quantum confinement effects are neglected.
4. *Mobility temperature dependence*: electron mobility is assumed constant.

The current conduction is based on a *drift and diffusion* model, Fermi-Dirac distribution is always implemented.

3.1.2 Geometrical structure

In Fig. 3.1 the geometrical structure of the nanowire transistor is reported, the cylinder height is 1.15 μm, in Fig. 3.2 a vertical and horizontal cross sections are illustrated along with the silicon doping concentration and dimensions.

The two heavy doping implants (one is shown in Fig. 3.2 (a)) are made to avoid parasitic resistances towards the source or the drain conctacs which are located at the two bases of the silicon cylinder.



(a) **(b)**

Figure 3.2: (a) Magnification on the heavy doping implant of a cross section of the cylinder in Fig. 3.1 with a plane passing through the axis of the cylinder. The doping profile is also shown. AB=16 nm, CD=36 nm, EF=56 nm. (b) Horizontal cross section of the cylinder in Fig. 3.1 with a plane perpendicular to the axis of the cylinder crossing one of the two heavy doping implants at the beginning and end of the cylinder.

The layer for data storage has not been included in this work, the gate contact is shown with a purple line in Fig. 3.2 and covers all the lateral surface of the cylinder allowing an optimal electrostatic control on all the semiconductor channel. In memory devices as presented in Sec. 1.3 there are multiple WLs to access different cells, here the device works as a long cylindrical capacitor. The cylinder height ensure that the drain bias applied would not change significantly the elctrostatics in the channel region.

3.1.3 Charge investigation and impact of the drain bias

In Sec. 2.3.3 it was shown that, as a first approximation, the drain current depends on the dimensions of the device, the mobility, the drain bias applied and the free charge carriers in the channel. Here we want to verify in first place the absence of the MNR in monocrystalline devices. Since the current depends mostly on the free charges present in the channel we have chosen to perform this first check on the simulated $Q_f - V_G$ characteristics.

In Fig. 3.3 free charges in the channel for 7 temperatures between 200 K and 500 K with different drain biases are shown. For easier anaytical comparison with formulas from the literature, charges are reported as densities per unit length (unit 1/cm). It is

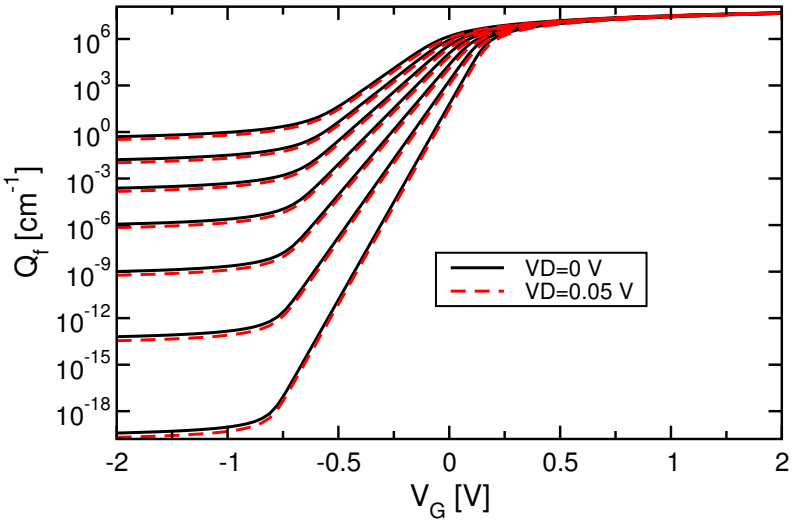


Figure 3.3: Free charges in the channel Q_f vs. V_G, with and without bias applied on the drain. Temperatures from 200 K to 500 K with steps of 50 K.

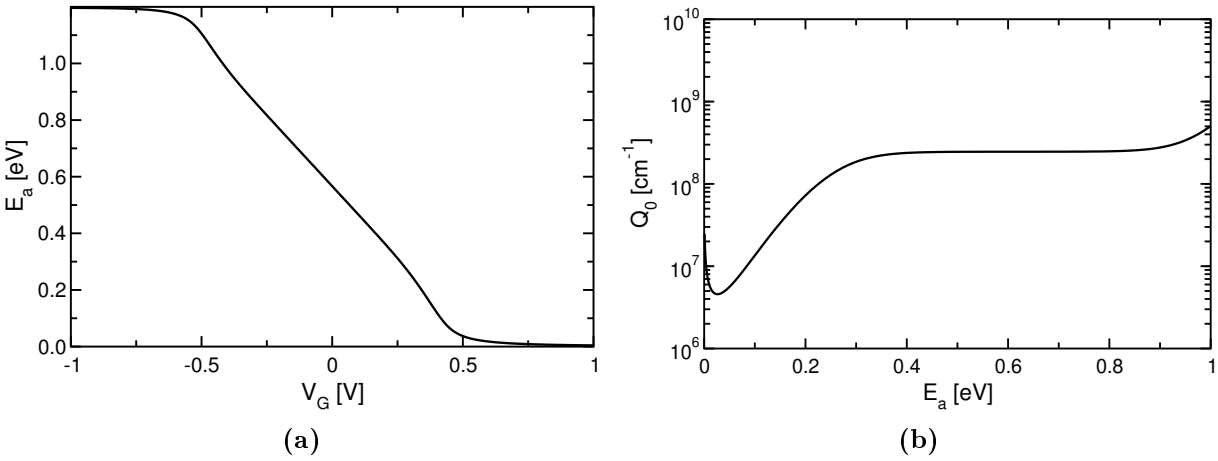


Figure 3.4: (a) Activation energies extracted from $Q_f - V_G$ in Fig. 3.3 relations without drain bias applied (b) Charge density prefactor extracted from $Q_f - V_G$ relations in Fig. 3.3.

easy to see how a small bias on the drain does not modify the density of free charges in the channel. This was expected, as written above, because the string is long enough (1.15 μm) to mitigate the effect of this bias.

Following the Arrhenius analysis illustrated in the previous chapter, the activation energy is extracted by the means of an Arrhenius plot along with the intercept in $1/kT = 0$, from which also the prefactor Q_0 is extracted.

Fig. 3.4b shows the prefactor Q_0, one can see at first glance that in a wide region the prefactor is linear in a semilogarithmic plot, hence it has an exponenatial relation with the activation energy for the process, that is the condition to follow the MNR. The fact

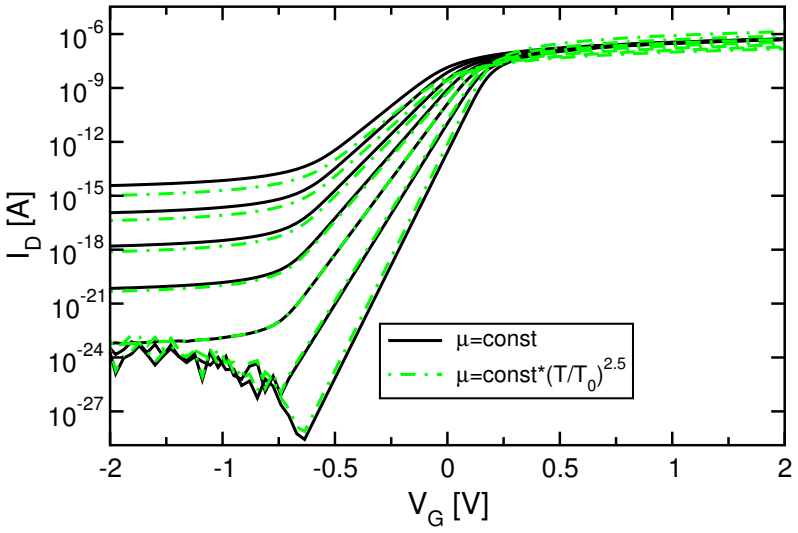


Figure 3.5: I_D vs. V_G, with $V_D = 50$ mV for the nanowire shown in Fig. 3.1. Temperatures from 200 K to 500 K with steps of 50 K.

that a device without trap states follows the MNR is an absolute novelty, no similar case can be found in the literature.

3.1.4 Current investigation and impact of the mobility

In order to escape the doubt that the MNR found in the $Q_f - V_G$ relation is only an error related to the fact that we analyzed the free charges in the place of the current, the same Arrhenius analysis, performed on the $I_d - V_G$ relation, is hereafter reported.

In Fig. 3.5 the $I_d - V_G$ relation is reported for 7 temperatures between 200 K and 500 K with and without the temperature dependence of the mobility as indicated by the graph labels ($T_0 = 300\ K$). The main differences are found in the off state and in the above threshold region which are not of interest for the MNR investigation.

With the temperature dependence, a *zero temperature coefficient* point in which all the curves with different temperature take the same value is present. After that point, with the rising of temperature the current decreases due to the mobility degradation. We therefore expect the activation energy to change sign for the $\mu = f(T)$ curves.

In Fig. 3.6a it is shown that the activation energy reaches negative values for a T-dependent mobility as expected. The two curves have a slightly different absolute value but the shapes of the curves are very similar, moreover the interesting part for

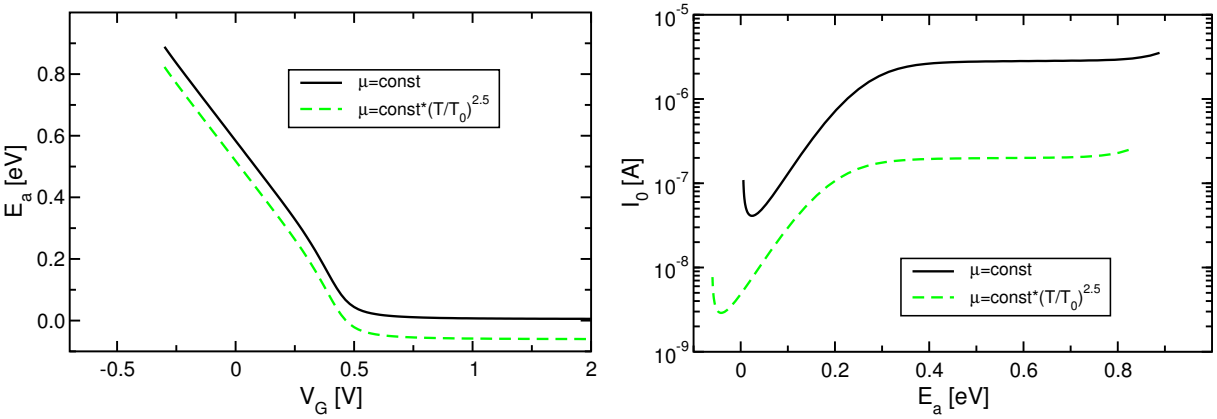


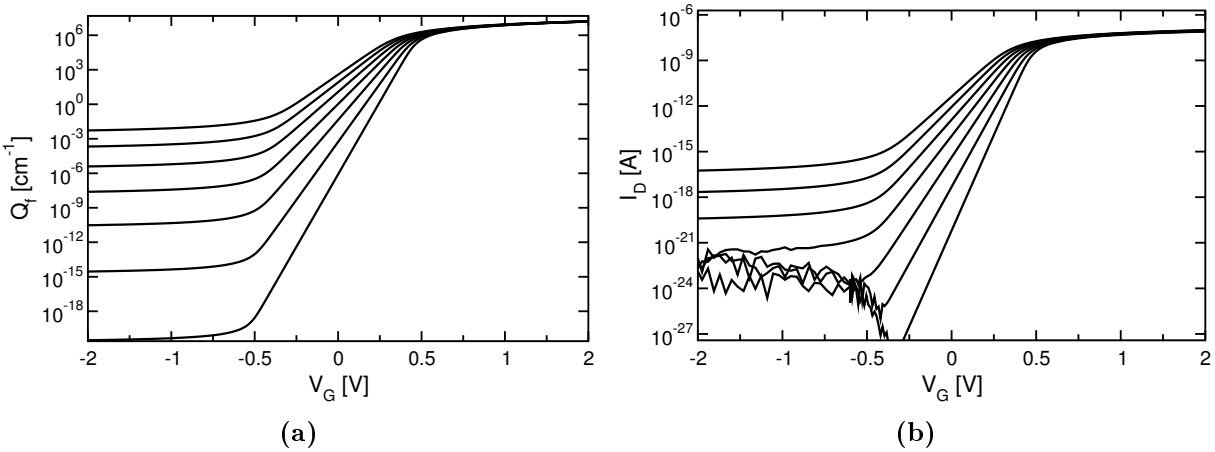
Figure 3.6: (a) Activation energy extracted from the $I_D - V_G$ curves in Fig. 3.5 by an Arrhenius analysis. (b) Current prefactor extracted from the $I_D - V_G$ curves in Fig. 3.5 by an Arrhenius analysis.

the Meyer-Neldel analysis is the subthreshold region which ends before the separation between the two curves become significant.

In Fig. 3.6b we can see again the mark of the MNR. Hence we can be sure that the MNR behavior shown in Fig. 3.4b is not an error related to the use of the charge in the place of the current. Moreover we can see in the figure that the curve with the T-dependent mobility is horizontally shifted in activation energies and has a vertical shift of approximatively one order of magnitude. Apart from these two rigid shifts the shape of the curves is identical, hence the assumption of constant mobility would not affect the investigation of the rule.

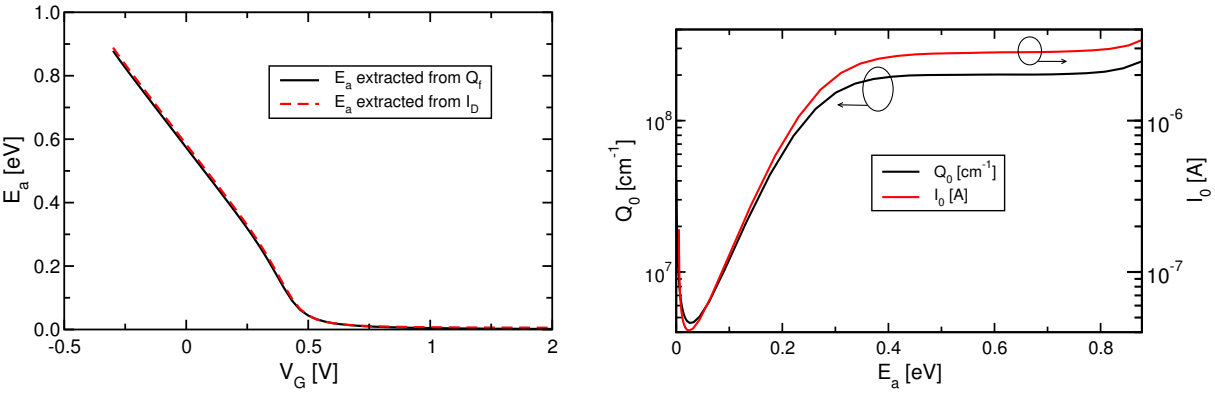
Once we have shown that both the current and charge prefactors show a MNR behavior we perform a comparative analysis to make sure of the abscence of differences between them. This last validation step is required since we decided to work with the charges relation throughout all the thesis work. This decision has been made to allow easier comparison with analytical formulas from the literature, because the current relation is usually obtained by integration of the charge in the channel as a function of the source-drain bias. Therefore using the charge for the analysis will save one integration step and in turn the analytical formulation is expected to be easier and to have closed form.

In Fig. 3.7 a comparison between free charges and current as a function of the gate bias is shown. We can see that apart from the order of magnitude the curves are quite similar, hence we expect similar activation energies and prefactors.



(a) (b)

Figure 3.7: (a) Free charges Q_f vs. V_G (b) Current I_D vs. V_G. Temperatures from 200 K to 500 K with steps of 50K.



(a) (b)

Figure 3.8: (a) Activation energies E_a extracted from 3.7a and from 3.7b (b) Prefactors Q_0 and I_0 extracted from 3.7a and from 3.7b.

In Fig. 3.8 activation energies and prefactors extracted from Fig. 3.7 are reported. Fig. 3.8a shows almost identical activation energies for the two processes while the prefactors differ from several orders of magnitude. Even with this magnitude difference the left and right y-axis, which are different, both cover 2 orders of magnitude and thus the represented shapes are very similar and both show the exponential region with almost the same slope.

We are now sure that the Meyer-Neldel analysis can be performed on the free charges present in the channel. Moreover, we can neglect the effect of the drain bias because of the chosen device. The already shown little mobility impact is not to take into account while investigating the charge because the latter does not depend on it.

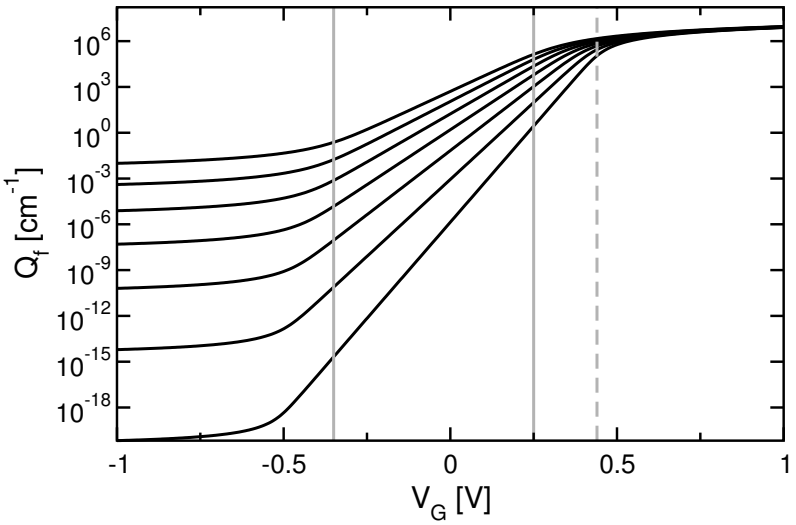


Figure 3.9: Q_f as a function of V_G. Between the two continuous grey line the region in which all temperatures follow the pure subthreshold relation is highlighted, i.e. a pure exponential relation, that is linear in this semilogarithmic plot. The dashed grey line signs the end of the subthreshold relation for the lowest temperature. Temperatures from 200 K to 500 K with steps of 50 K.

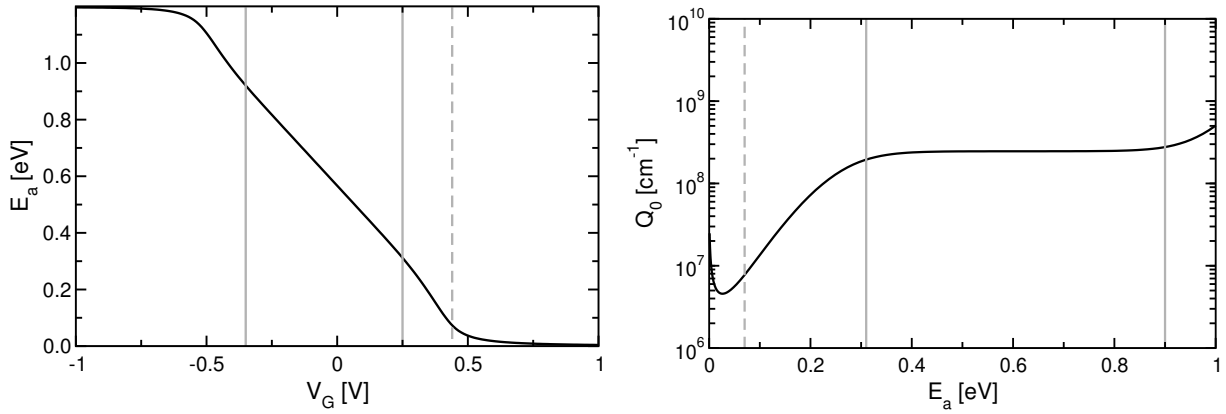


Figure 3.10: (a) Activation energy extracted from $Q_f - V_G$ relations in Fig. 3.9 with the voltage spans highlighted (b) Charge density prefactor extracted from $Q_f - V_G$ relations in Fig. 3.9. The grey lines corresponds to the activation energies at the intersections with the grey lines in Fig. 3.10a.

3.1.5 Prefactor regions correlations

Fig. 3.9 shows the same $Q_f - V_G$ relations of Fig. 3.7a in which two zones are highlighted. Between the two continuous grey line, from -0.35 V to 0.25 V, there is the region in which the highest temperature is in the pure subthreshold region, i.e. the $Q_f - V_G$ relation is purely exponential. Since the highest temperature is the last that enters the subthreshold region and the first leaving it, between these two grey lines all

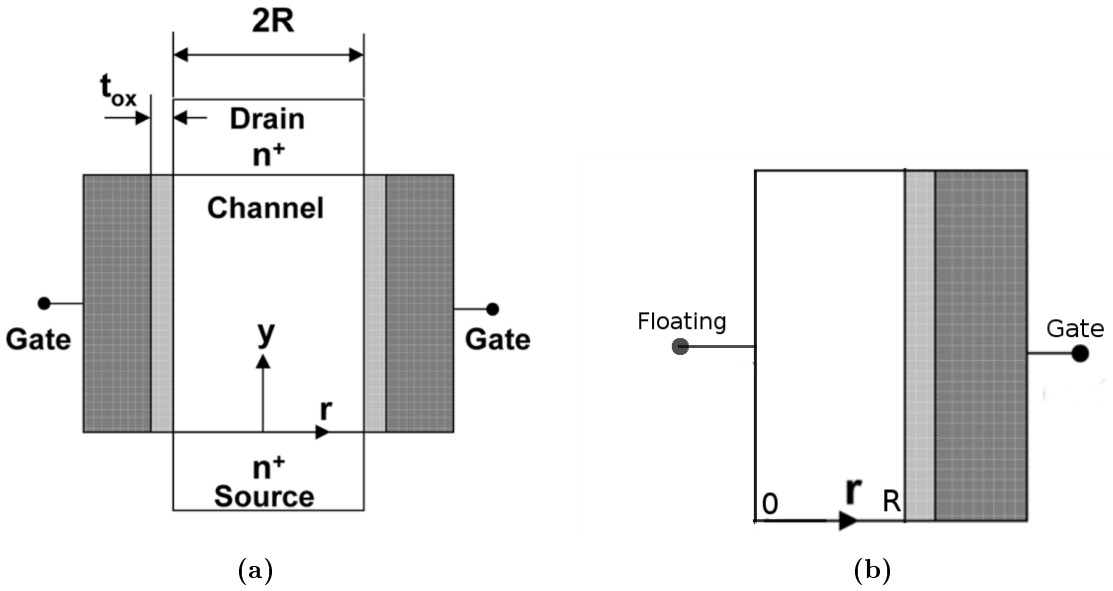


Figure 3.11: (a) Schematic illustration of a vertical cross section of the device. (b) Schematic illustration of the device's half in which the Poisson' equations is solved.

the curves follow an exponential relation.

The dashed grey line at $V_G = 0.44$ V marks the end of the pure subthreshold region for the lowest temperature, between this line and the continuous one at $V_G = 0.25$ V we have that some temperatures have already left the pure subthreshold trend while others are still following the pure exponential relation that characterizes the above mentioned region. We will refer to this region as the *transition region* between the pure subthreshold region and the above-threshold one.

Fig. 3.10 shows the corresponding zones in the activation energy and prefactor plots. Fig 3.10a reports the same vertical lines of Fig. 3.9 and illustrates that, as previously said, the activation energy is perfectly linear in the subthreshold region and deviates from this trend in the transition region. The E_a values at which the vertical grey lines intercept the curve are 0.9, 0.31 and 0.07 eV, reported in Fig. 3.10b by means of vertical grey lines. It can be seen that in the pure subthreshold region the prefactor is constant, while there is a good agreement between the exponential region of the prefactor, i.e. the region in which the MNR is respected, and the transition region reported in Fig. 3.10.

The nanowire MOSFET is a so-called *floating substrate* device. This means that, with the exceptions of the source and drain contacts at the two cylinder bases, there is only the gate bias imposed to the channel. The Poisson's equation is usually solved

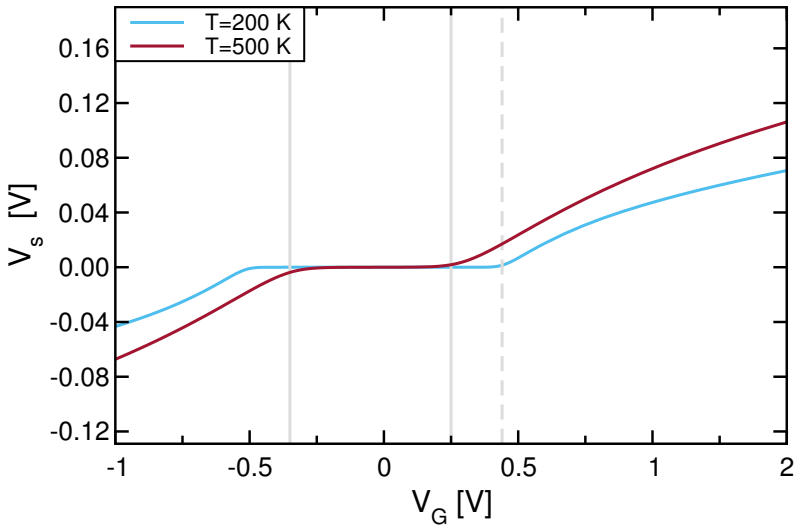


Figure 3.12: Voltage drop V_s between the center of the cylinder and the silicon oxide interface for the lowest (200 K) and highest (500 K) temperatures.

exploiting the cylindrical simmetry; in the radial direction the device can be considered as a 1-D device with the gate bias imposed on one side and left floating on the other side (Fig. 3.11b). The electron quasi-Fermi level is imposed by the source and drain biases, set to 0 V in our case. The situation is schematically depicted in Fig. 3.11.

We define the silicon voltage drop V_s as the potential difference between $r = R$ and $r = 0$ (Fig. 3.11b). From the electrostatic point of view the switch from the pure subthreshold region to the transition region seems to be a mark of the bands' bending. In Fig. 3.12 one can see that the pure subthreshold region corresponds to the region in which V_s remains constant for all the temperatures, while the entering in the transition region is marked by the point in which $V_s(T = 500\ K)$ starts rising. The transition region ends when V_s stops to be constant for the lowest temperature.

The fact that the voltage drop V_s remains constant over a wide range of gate voltages is a mark of the combined effect of the floating device structure and the small silicon volume. Assuming that the flatband voltage is 0 V and that the device is undoped and gate, source and drain biases are set to 0 V, the electrons quasi-Fermi level is located at midgap and the bands are flat since no charge is present. When the gate bias is raised, the bands will start to bend only when the electrons coming from source and drain conctacts will be enough to create a non-neglectable electric field and hence potential drop along with bands bending. Till that moment the charge amount must rise with rising gate bias but the bands must remain flat. Therefore there is a rigid shift of the bands to take the electron quasi-Fermi level closer to the conduction band and increase the electron density.

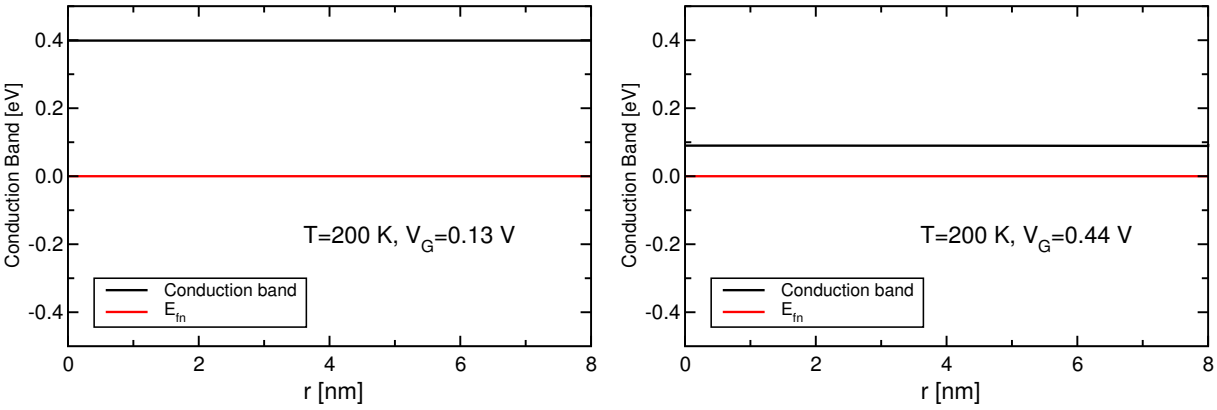


Figure 3.13: (a) Energy bands plot in the radial direction for gate bias $V_G = 0.13$ V. The structure is composed by 8 nm of silicon followed by the silicon oxide insulator. (b) Energy bands plot in the radial direction for gate bias $V_G = 0.44$ V. In both figure it can be noted that the conduction band is flat but with the rising of the gate bias the conduction band get closer to the elctrons quasi-Fermi level.

This process is confirmed by Fig. 3.13a in which we can see thatthe conduction band is flat at $V_G = 0.13$ V and Fig. 3.13b, where the gate bias is raised to 0.44 V, and the band is still flat, but the Fermi level is closer to the conduction band. At $V_G = 1.325$ V (Fig. 3.14) the electron quasi-Fermi level (always fixed at 0 V by the source and drain contact) is inside the conduction band and the induced charge bends the conduction band.

We have just shown a correlation between the MNR and the transition region of the Q_f-V_G characteristic. The entering in the transition region happens when the highest temperature stops to follow a pure exponenatial relation, which occurs, as shown in Fig. 3.12, when V_s starts to take values different from 0 V and hence the bands bend. This bending takes place when the charge brought by the electrons induced by the rising gate bias creates a strong enough electric field. This is an evidence for the correlation of the MNR with the *self-consistency* of the bands: as soon as the induced electrons have an effect on the energy bands, the system does not follow the pure exponential relation of the subthreshold region. When this happens at different biases for different temperatures, because with lower temperatures come lower induced charge, we have the transition region and in turn we found a charge prefactor that follows the MNR.

3.1.6 Frozen-band analysis

At the end of the previous subsection a correlation between the presence of the transition region and the following of the MNR was presented. In order to find more evidence of this theory we want to introduce the concept of *frozen bands*.

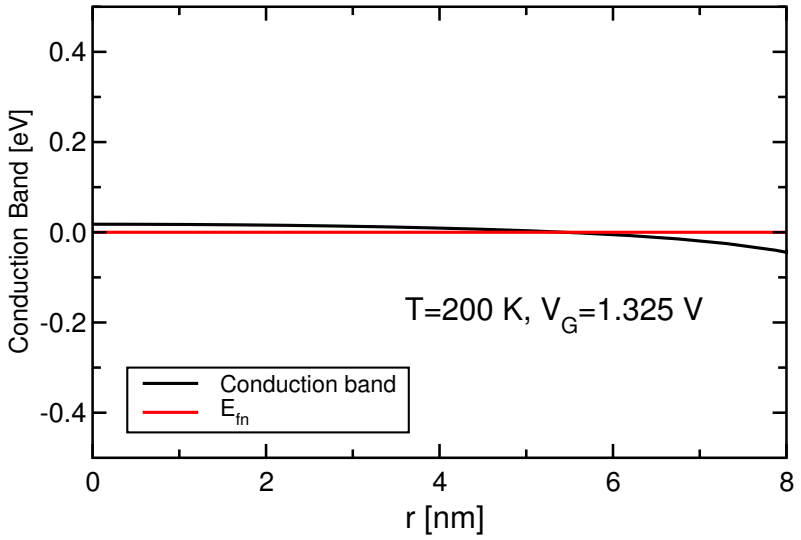


Figure 3.14: Energy bands plot in the radial direction for gate bias $V_G = 1.325$ V. The structure is composed by 8 nm of silicon followed by the silicon oxide insulator. The induced charge is high enough to bend the band.

In our region of interest the Fermi-Dirac distribution is well approximated by the Maxwell-Boltzmann distribution since the Fermi level is far enough from the conduction band. The induced charges are regulated by this equation [13]:

$$n = N_C \times e^{-(E_C - E_f)/kT}, \tag{3.1}$$

where the temperature dependences are contained in the prefactor N_C which has a $T^{3/2}$ dependence, in the denominator of the exponential and in the numerator. Actually, different temperatures will have, in general, different energy bands and hence different numerators.

With *frozen bands* we mean that we will compute the band profile for the lowest temperature (200 K) and use it to compute the induced charges also for the higher temperatures. In this way we are turning off the self-consistency of the Poisson equation, the band profile will not be affected by the charge induced by the rise of temperature. With the rising of temperature the population of the electronic states will change (N_C and T in Eq. (3.1)) while the band profile in the numerator will be frozen at the lowest temperature. The result is shown in Fig. 3.15. It should be noted that in the pure subthreshold region there is a perfect agreement for each temperature. This means that in the pure subthreshold region the bands are identical for each temperature and induced electrons have a totally negligible effect.

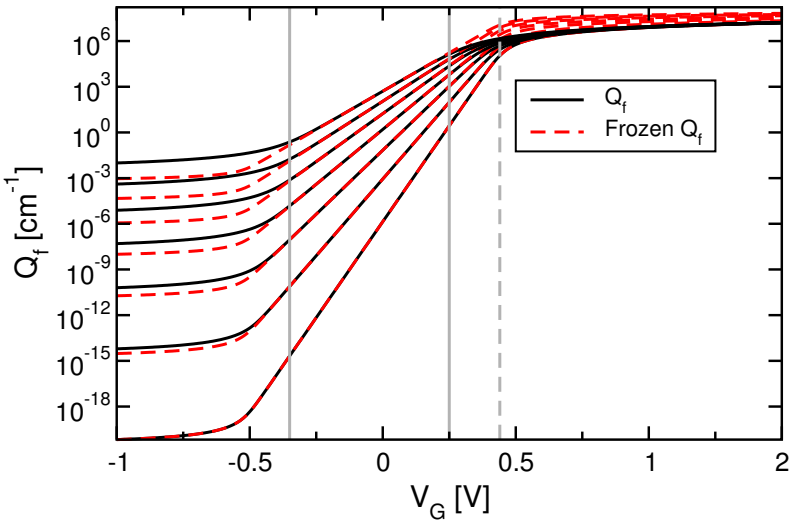


Figure 3.15: Comparison between the free charges calculated with and without self-consistency. Temperature from 200 K to 500 K with steps of 50 K.

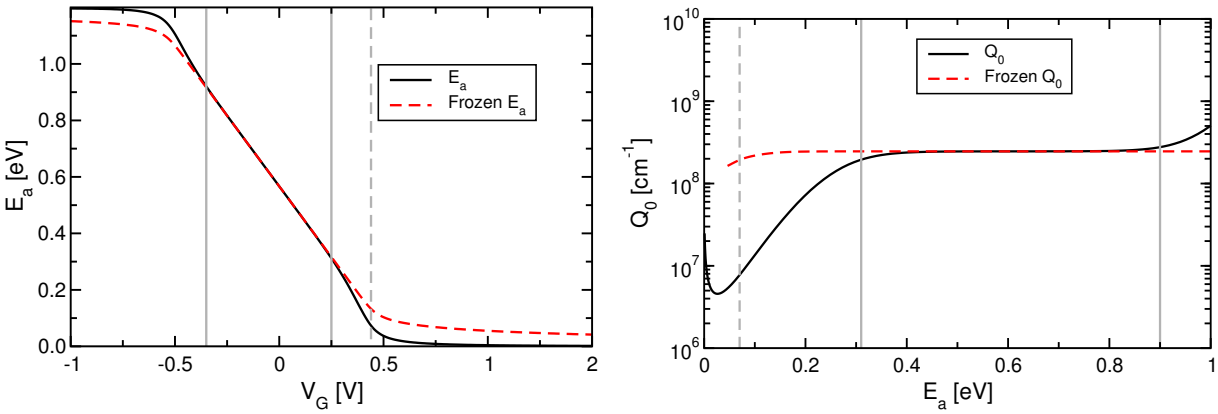


Figure 3.16: (a) Activation energies extracted from Q_f-V_G relations in Fig. 3.15 (b) Prefactors extracted from Q_f-V_G relations in Fig. 3.15.

By performing this analysis we can search for another evidence of the correlation between the MNR and the transition region. In the frozen bands case we expect the absence of the transition region and hence of the MNR. That is because if its presence is brought by the fact that the induced charge at different temperatures starts to bend the bands at different biases, with the frozen bands we use the same bands for all temperatures and in turn the leaving of the pure exponential relation must occurr at the same bias for each temperature.

The activation energies and prefactors extracted from the Q_f-V_G relations in Fig. 3.15 are shown in Fig. 3.16. Our hypothesis on the effect of the frozen bands on the MNR turned out to be true: the exponential relation between the prefactor and the activation

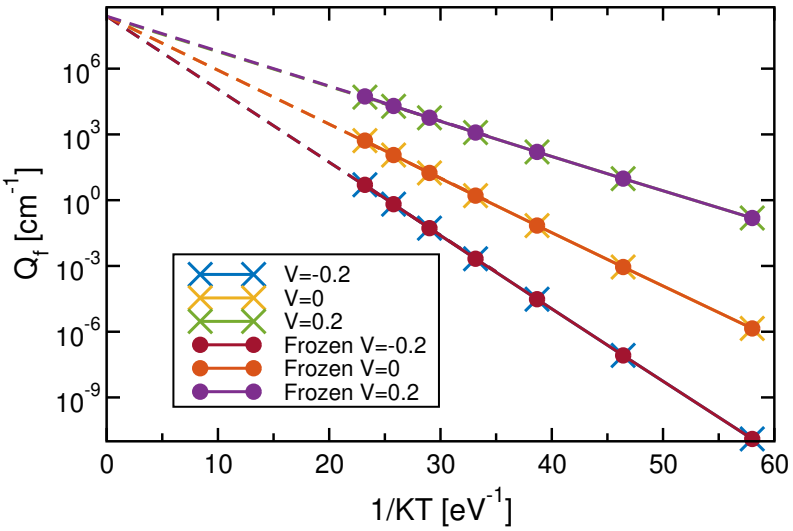


Figure 3.17: Arrhenius plot of the charges presented in Fig. 3.15 in the pure subthreshold region. The charge obtained with and without the self-consistency are identical and follow a pure exponential relation as one can see from the perfect fit with the dashed lines. The vertical grey line marks the intercept from which the prefactor is extracted.

energy is absent in the prefactor extracted from the frozen curves as shown in Fig. 3.16b. As a final remark one should note that also in the prefactor calculated from the frozen relation there is a small activation energy interval, around the dashed grey line in Fig. 3.16b, in which the prefactor decreases. An explanation of this fact is presented in the next subsection by means of Arrhenius plots.

We have then found a proof of the fact that the MNR is present in the charge, and in turn in the current, relations of a monocrystalline device and it is related to the self-consistency of the energy bands. The fact that higher temperatures bring higher induced charges makes the Q_f-V_G relations at different temperatures to deviate from the pure exponential relation at different biases. The voltage span that starts when highest temperature leaves the pure exponential relation and ends when the lowest temperature does the same thing has been called *transition region* and whitin it the induced free charges and thus the current follow the MNR.

3.1.7 Arrhenius plots

In the previous subsections we have presented a correlation of the MNR with a transition region in which the $Q_f - V_G$ relations with the higher temperatures have stopped to follow a pure exponential relation while the lower temperatures are still following it.

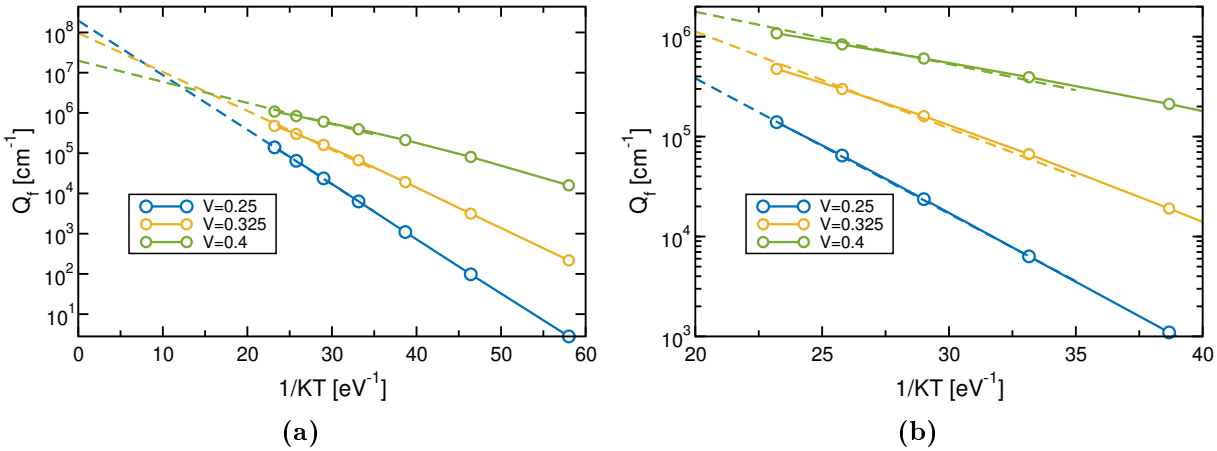


Figure 3.18: (a) Arrhenius plot of the transition region of the charges in Fig. 3.15 along with exponential fitting dashed lines (b) Magnification of the Arrhenius in Fig. 3.18a to show the deviation from the exponential trend.

In this subsection we present the Arrhenius plots for the analysis just reported with focus on two different regions: the pure subthreshold and the transition one. It will turn out that the MNR in this system may derive also from the extraction method, questioning its validity. Moreover, an investigation on the decrease of the prefactor obtained from the frozen charges will be shown.

Fig. 3.17 shows the Arrhenius plot for the charges in Fig. 3.15 in the subthreshold region with and without the self-consistency (dashed lines are the exponential fits). The figure shows another proof of the fact that the mobile charges do not have any effect on the bands in the pure subthreshold region since the charge values obtained with and without the frozen bands are identical. Moreover, Fig. 3.17 confirms that in the pure subthreshold region the relation between the charge and the activation energy is purely exponential since there is no deviation from the exponential fit which is linear in the semilogarithmic plot.

Fig. 3.18 shows instead that outside the pure subthreshold region the charge relations deviates from the exponential trend. The light blue curve at $V_G = 0.25$ V is located at the boundary between the pure subthreshold and transition region, the fit shows that at this bias the exponential relation is still valid for every temperature (each empty circle in the plot represents one temperature). The other two lines, whose biases are located in the transition region, show how, with the rise of temperature, the deviation from the exponential trend is more and more evident.

Without the self-consistency we expect the exponential relation to hold also at high

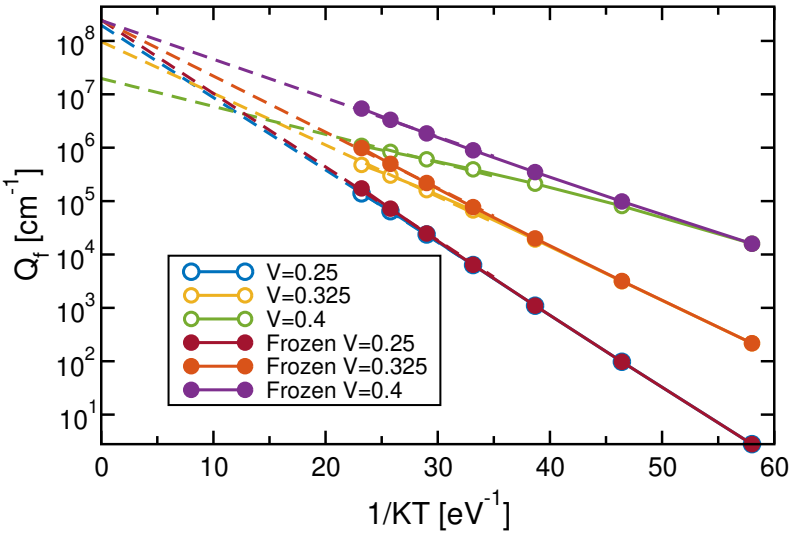


Figure 3.19: Arrhenius plot of the charges presented in Fig. 3.15 in the transition region with and without the frozen bands. In the transition region the frozen bands have a noticeable effect. The vertical grey line marks the intercept from which the prefactor is extracted.

temperatures. Fig. 3.19 confirms our hypothesis. The point located at $1/kT = 58$ eV^{-1} is obviously the same with and without the frozen bands, it is the reference point for the band calculation. The more the biases are afar from the pure subthreshold region and the higher the temperatures, the bigger the difference between the case with the frozen bands ad the self-consistent one. From the graph one can also see that the prefactor of the frozen case, i.e. the intercept of the dashed fitting lines with the vertical grey line, remains constant. The self-consistent one, instead, decreases. This decrease, which is linear on this semilogarithmic plot is the exponential decrease related to the MNR.

We can then infer that the self-consistency of the bands and the extraction method can lead to the MNR. In Fig. 3.18 it is evident that we performed an exponential fit with points that were deviating from a pure exponential relation. These little deviations, due to the semilogarithmic plot, turn out in a linear decrease of the logarithm of the prefactor and finally in the exponential decrease that we linked to the MNR.

The fact that evidence of MNR can be related to experimental errors and extraction methods is already present in the literature [37] [38]. Here we want to point out that in MNR analysis of semiconductor devices, the risk of relating the MNR to such effects is relevant since the vast majority of devices presents I-V characteristics with exponential trends in some voltage ranges that are abandoned when a threshold condition is met and this threshold condition varies with temperature, as in the case that we have here analyzed.

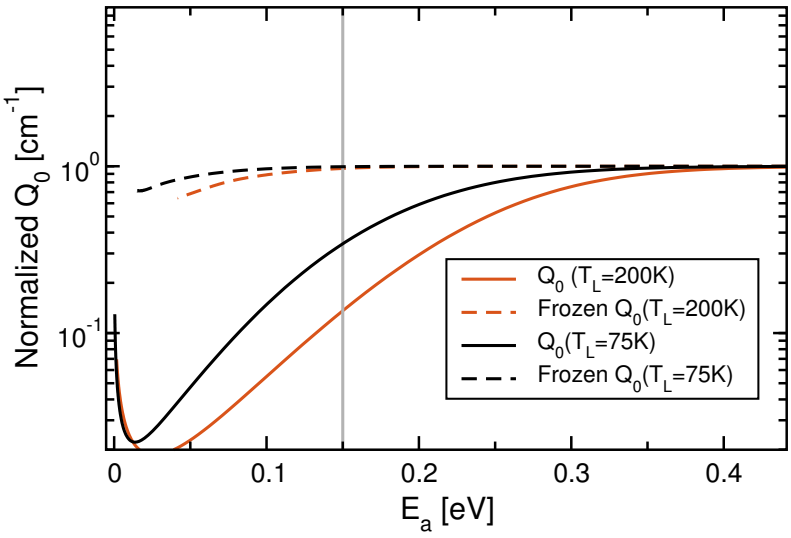


Figure 3.20: Comparison between the prefactors extracted with and without the frozen bands for two different temperature ranges. The vertical grey line marks the begin of the descending region for the frozen prefactor extracted from the 200-500 K temperature range. The prefactor extracted from the 75-500 K temperature range is still flat at the same activation energy.

With reference to the previously mentioned descending trend of the prefactor extracted from the charges obtained with the frozen bands, we have hypothesized that it is related with the bias at which, for the reference temperature, the self-consistency is no longer neglectable. In other words we expect the descending region of the frozen prefactor to take place at activation energies corresponding to the bias at which, for the reference temperature, the induced charges start to have an effect on the bands. This fact happens because when the bands start to bend the rise of the gate bias will induce less charges, and then the point located at $1/kT = 58$ eV^{-1} will take a slightly smaller value with respect to the case in which the bands are blocked. Hence the straight line toward $1/kT = 0$ eV^{-1} will give a smaller prefactor.

To verify this hypothesis we decided to run a simulation with the bands frozen at 75 K. We expect that the prefactor with the frozen bands at 75 K will start to decrease at higher biases and thus lower activation energies with respect to the one computed with the frozen bands at 200 K, since with lower temperatures come lower induced charges and then the point at which the self-consistency will be no longer neglectable will come at higher biases.

Fig. 3.20 seems to confirm our hypothesis. The prefactor extracted with the bands computed at 75 K starts to descend at lower activation energies with respect to the one with the bands calculated at 200 K. All the prefactors plotted in the figure are normalized with respect to the value that they take in the pure subthreshold region where they are

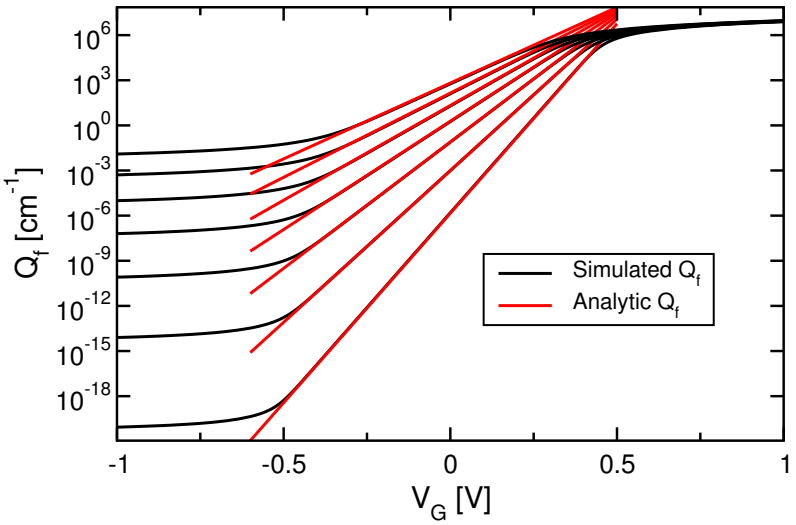


Figure 3.21: Comparison between the Q_f-V_G relations computed by the simulator and the Q_f-V_G relations obtained with Eq. (3.2). Temperatures from 200 K to 500 K with steps of 50 K.

flat.

3.1.8 Analytical and numerical comparison

To conclude the analysis on this nanowire device we present here a comparison between the results of our simulations and an analytical model found in the literature, both for the induced free charges and the prefactors. Moreover, the Meyer-Neldel energy E_{MN} extracted from the exponential region of the prefactor of our simulations will be compared with the Meyer-Neldel energy obtained in the TFT analysis presented at the end of the second chapter.

It has been demonstrated [39] that the free charges in the subthreshold region of an undoped nanowire MOSFET follow:

$$Q = Q_0 \exp\left(\frac{V_{GS} - V_0 - V_{fn}}{V_{th}}\right) \tag{3.2}$$

here $V_0 = \Delta\varphi + (kT/q)\ln(8/\delta R^2)$ and $V_{th} = kT/q$; where $\Delta\varphi$ is the work-function difference between the gate and instrinsic silicon, $Q_0 = (4\epsilon_{si}/R)(kT/q)$ and $\delta = q^2 n_i / kT\epsilon_{si}$.

Combining the definitions written above we can rewrite the equation to clarify the temperature dependences in the subthreshold region:

$$Q = \frac{qR\sqrt{N_C N_V}e^{-E_g/2kT}}{2} exp\left(\frac{q(V_{GS} - \Delta\varphi - V_{fn})}{kT}\right) \tag{3.3}$$

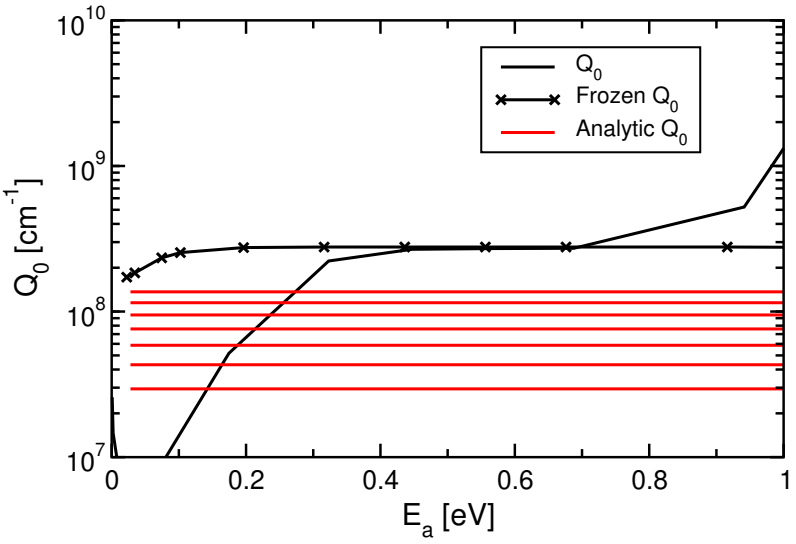


Figure 3.22: Comparison between the prefactors extracted from the frozen and self-consistent charges relations and the ones obtained from the analytical formulation in Eq. (3.2).

the temperature dependence is hence contained in the N_c and N_v terms in $n_i = \sqrt{N_C N_V} e^{-E_g/2kT}$ and in the exponential term.

Fig. 3.21 shows a perfect agreement between the free charges values obtained in our simulations and the analytical model. We can rewrite Eq. (3.3) including the energy gap dependence in the exponential term obtaining:

$$Q = \frac{qR\sqrt{N_C N_V}}{2} exp\left(\frac{q(V_{GS} - \Delta\varphi - V_{fn} - E_g/2q)}{kT}\right) \tag{3.4}$$

In this equation the prefactor $qR\sqrt{N_C N_V}/2$ contains a temperature depedence in $\sqrt{N_C N_V}$ that goes with $T^{3/2}$ and will affect our analytical comparison, because the prefactor of the MNR is expected to be temperature-independent. The analytical activation energy turns out to be: $E_a = q(V_{GS} - \Delta\varphi - E_g/2q)$, where we neglected the electron quasi-Fermi potential because it is constant and equal to 0 in our device, as imposed by the source and drain biases.

Fig. 3.22 shows the comparison of the prefactors extracted from the Arrhenius analysis with and without the frozen bands and the 7 different prefactor values calculated with Eq. (3.4), one for each temperature. The absolute value differs by one order of magnitude at maximum, we think that this error depends on the fact that the prefactors with the Arrhenius analysis are extracted assuming that they do not contain any temperature dependence. The fact that the analytical prefactors are constant for all the activation energies is coherent with the correlation between the decrease of the prefactor and the

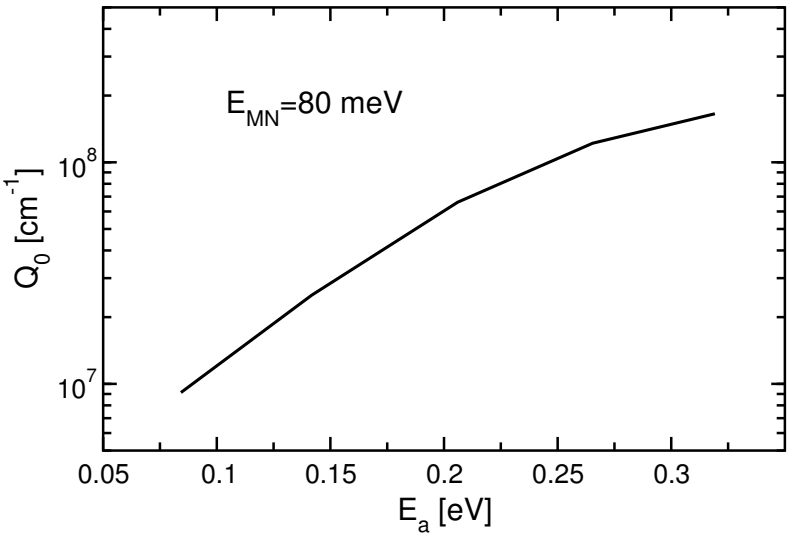


Figure 3.23: Prefactor extracted from the transition region of the Q_f-V_G relations in Fig. 3.9 along with the the inverse of the slope in the semilogarithmic plot E_{MN}, i.e. the Meyer-Neldel energy.

fact that the relation stop to follow the exponential trend. One can see from Fig. 3.21 that the red curves are perfect straight lines, because the model reported in Eq. (3.2) is valid only in the pure subthreshold region (i.e. it is purely exponential) and not in the transition one. The fact that the analytical prefactor is flat like the frozen one is another evidence of the fact that the presence of the MNR in this device may depends on the extraction method that assumes an exponential relation which is not present at certain biases for some temperatures under analysis. One may argue that the fact that the transition region is absent in the analytical model depends on the approximation introduced to derive Eq. (3.2). This will be investigated with the planar MOSFET analysis, whose relations contain little approximations, that will be exposed in the next section.

To conclude the analysis of the MNR in a nanowire MOSFET we report the extracted Meyer-Neldel energy E_{MN}, which is the inverse of the slope of the prefactor in a semilogarithmic plot against the activation energy. Fig. 3.23 shows the value of 80 meV, similar to the ones shown in Fig. 2.22 at the end of chapter 2. In that work the extracted E_{MN} value was also correctly predicted by an analytical model that linked it to the trap states distribution, while we obtained the same value in a device with any kind of state in the gap. We will search for explanation of this discrepancy either by performing the same analysis on a planar device more similar to the TFT analyzed in the cited work, either by including trap states in our analysis as it will be done in chapter 4.

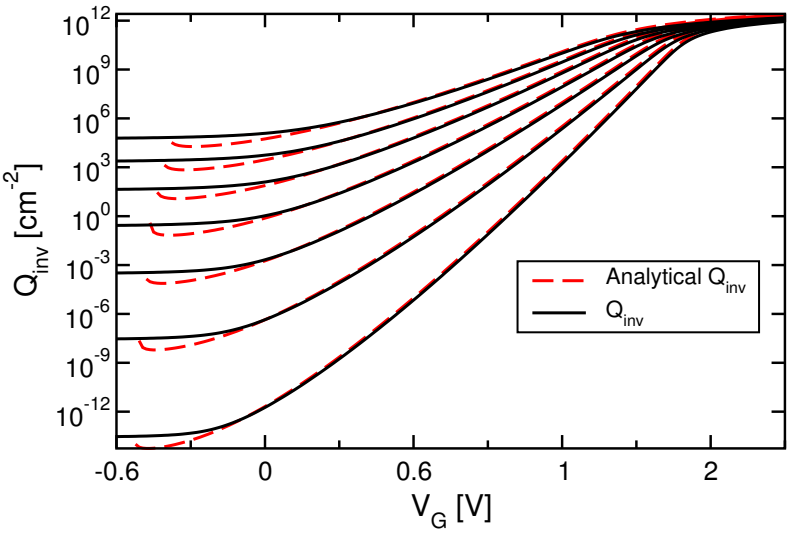


Figure 3.24: Q_{inv}-V_G relations for 7 temperatures from 200 K to 500 K with steps of 50 K along with the comparison with the values obtained from Eq. (3.5).

3.2 Long N-channel planar MOSFET

3.2.1 Device and environment

The simulated device is a standard long N-channel planar MOSFET, like the one presented at the beginning of chapter 2. The channel length is 1 μm, the substrate is heavily p-doped with a concentration $N_d = 10^{18}$ cm^{-3}. The oxide and substrate thickness are respectively 7 nm and 1 μm. With this thickness we are sure that the substrate will not be fully depleted for any gate bias and then the device will follow the standard long channel MOSFET relations. Source and drian bias are always set to 0 V, hence the device works like a capacitor as the nanowire MOSFET presented in the previous analysis.

We have assumed the same approximations taken in the nanowire MOSFET analysis presented in the previous section. We will hence base our analysis on the inversion charge present in the channel region in the place of the current, since the latter is proportional to the derivative of the former, on the temperature and on the mobility as one can see from Eq. (2.2). We will then neglect the mobility dependence and we will reduce the dependence on temperature, since the charge depends on T while the current on T^2.

3.2.2 Simulation results and analytical comparison

In Fig. 3.24 the simulated Q_{inv}-V_G relations are reported for 7 temperatures from 200 K to 500 K with steps of 50 K, along with the same relations obtained from the

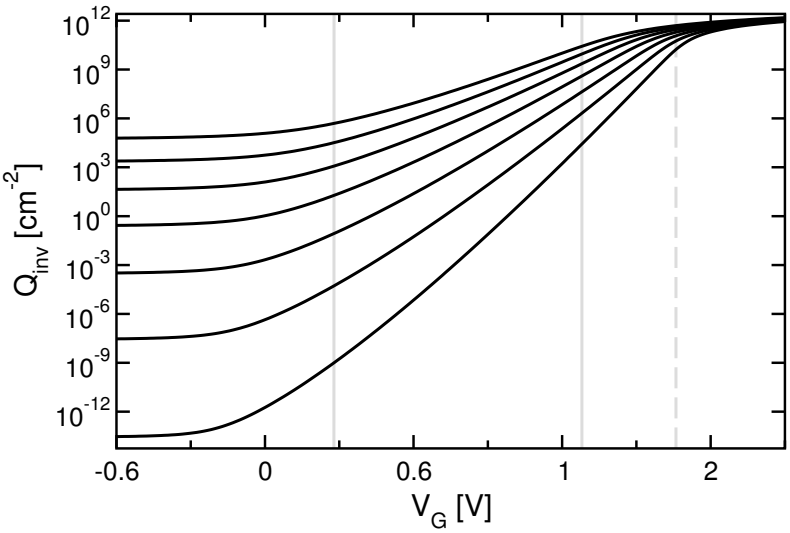


Figure 3.25: Q_{inv}-V_G relations for 7 temperatures from 200 K to 500 K with steps of 50 K, the veritcal grey lines highlight the pure subthreshod and the transition region.

analytical expression for the subthreshold region that we report here:

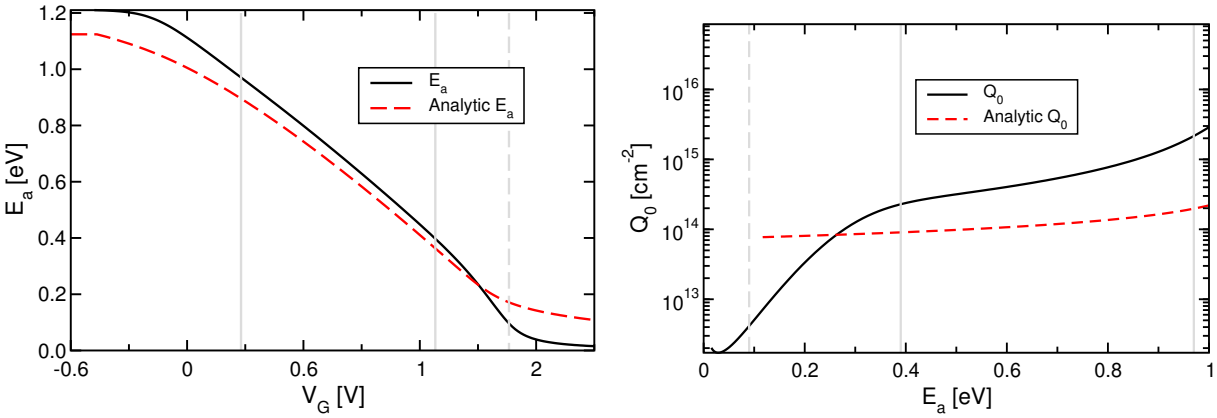
$$Q_{inv} = \sqrt{\frac{\epsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right) \left(\frac{n_i}{N_a}\right)^2 e^{q(\psi_s - V_{fn})/kT}. \tag{3.5}$$

As indicated from the units used, all the curves reported are carrier concentrations per unit area (1/cm^2). We can see that there is a perfect agreement from the beginning of the subthreshold region and also in the transition to the above-threshold region. The analytical Q_{inv} values are not obtained in a fully theoretical fashion. The values substituted in the analytical formula are taken from the results of the simulations.

An important parameter in this analytical analysis is ψ_s, i.e. the voltage drop in the substrate, since it appears both in the prefactor and in the activation energy. In fact, by rewriting Eq. (3.5) in order to make all the exponential dependencies explicit we obtain:

$$Q_{inv} = \sqrt{\frac{\epsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right) \left(\frac{\sqrt{N_c N_v}}{N_a}\right)^2 e^{\frac{q(\psi_s - E_g/q)}{kT}} \tag{3.6}$$

where the prefactor shows two main dependecies: one on $\psi_s^{-1/2}$ and one on the temperature, either explicit either with another factor T^3 in $\sqrt{N_C N_C}$. The numerator of the exponential is the activation energy for the process, we therefore have $E_a = E_g - q\psi_s$ that is coherent with the Arrhenius analysis presented in Sec. 2.1.6.



(a) (b)

Figure 3.26: (a) Activation energy extracted from the charge relations in Fig. 3.25 along with the computed analytical value (b) Prefactor extracted from the charge relations in Fig. 3.25 along with the computed analytical value for T=350 K. All the vertical grey lines are the same as in Fig. 3.25.

3.2.3 Pure subthreshold and transition regions

As presented in Sec. 3.1.5 we can identify two zones, the pure subthreshold and the transition one, as they were defined before. In Fig. 3.25 are reported the same Q_{inv}-V_G relations of Fig. 3.24 with the vertical grey lines that indentify the two different zones.

From Fig. 3.26 we can see again that the activation energy is linear in the pure subthreshold region and that the exponential trend of the prefactor is again correlated with the transition region as defined before. The vertical grey lines in Fig. 3.26a are located at the same bias as in Fig. 3.25. The vertical grey lines in Fig. 3.26b are located at the activation energies at which the vertical grey lines intersect the curve in Fig. 3.26a.

3.2.4 Analytical comparison and Arrhenius plots

To better understand the behaviors of activation energy and prefactor it is useful to look at the analytical comparisons shown in Fig. 3.26. Similar trends can be observed in the pure subthreshold zone. We know from Sec. 2.1.6 that the $E_a = E_g - q\psi_s$ and from Eq. (3.6) that the analytical prefactor follows a $\psi_s^{-1/2}$ trend. Hence the analytical prefactor will be described by $(E_g - E_a)^{-1/2}$ as shown in Fig. 3.26b. We believe that the differences between the two prefactors are mainly due on the fact that the analytical one is temperature-dependent (the shown one was computed with T=350 K) while the extracted one is independent of temperature and it performs a sort of average.

The fact that the analytical prefactor does not deviate from the trend in the transition

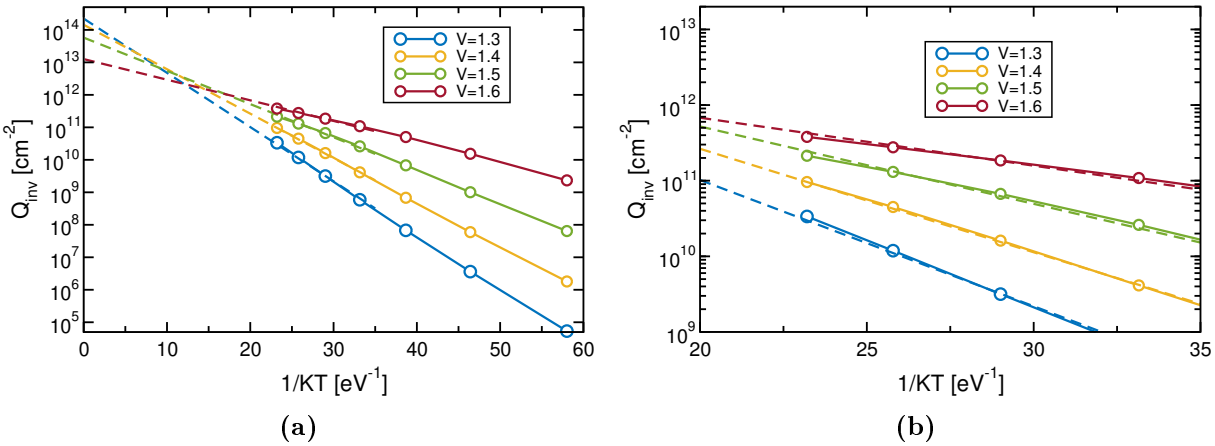


Figure 3.27: (a) Arrhenius plot of the charges in Fig. 3.25 in the transition region as indicated from the label. The vertical grey line marks the intercept for the extraction of the prefactor (b) Magnification on the highest temperature points of Fig. 3.27a in order to show the deviation from the exponential trend that makes the fit deviates and in turn the prefactor to decrease exponentially.

region suggests that the exponential trend of the prefactor is again due to the self-consistency and the extraction method as described in the nanowire MOSFET analysis. In this case the analytical formulation shows good agreement also in the transition region, this fact falsifies the idea that the absence of an exponential trend in the prefactor could be related to the approximations made. In Fig. 3.27a it is shown the Arrhenius plot for 4 gate bias values in the transition region. In Fig. 3.27b the magnification on the highest temperature points shows the deviation from the exponential trend that is responsible for the exponenial trend of the extracted prefactor. As a final remark we show in Fig. 3.28 that the extracted Meyer-Neldel energy for the transition region of the planar MOSFET is almost the same value of the nanowire MOSFET shown in Fig. 3.23.

We can conclude that the cylindric geometry and the the fact that the nanowire MOSFET is a floating substrate device have some effects on the prefactor: in the pure subthreshold region the prefactor extracted from the nanowire MOSFET is constant while the one extracted from a planar transistor shows a descending trend. The presence of a region in which the prefactor shows an exponential trend, instead, is not related to the specific device that is under analysis. We have concluded that it is related to the self-consistency of the energy bands and the transition region in which not every curve used for the Arrhenius analysis follow the same trend. This fact is dependent on the temperature range under analysis since a wider temperature range will have a larger transition region, vice versa for a narrower temperature range.

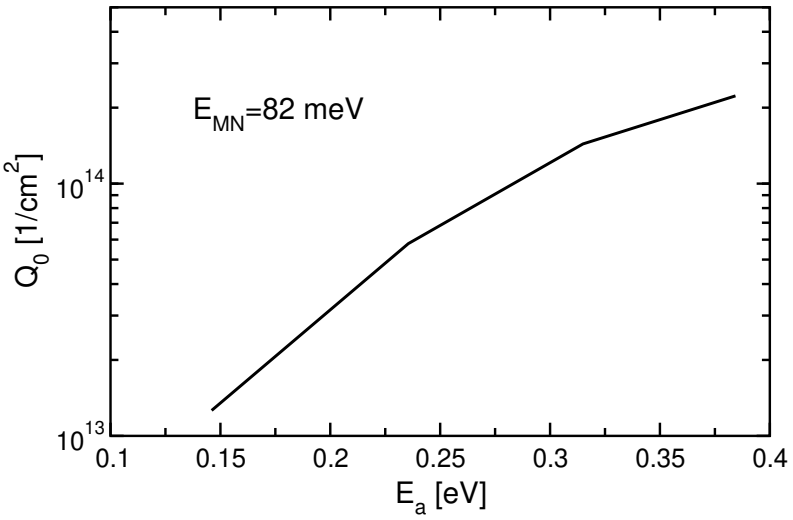


Figure 3.28: Prefactor extracted from the transition region of the Q_f-V_G relations in Fig. 3.25 along with the the inverse of the slope in the semilogarithmic plot E_{MN}, i.e. the Meyer-Neldel energy.

We pointed out that in the analysis of MOSFET devices this effect must be taken into account since a large quantity of devices present exponential trend in the subthreshold region that are gradually abandoned towards the above-threshold region.

Chapter 4

The impact of traps

The purpose of this chapter is to investigate the role of trap states in a Meyer-Neldel rule analysis. The analysis will be performed on the same cylindrical device presented in the previous chapter, with identical assumptions and approximations. The analysis will start with the investigation on how different trap distributions affect the $Q-V$ characteristic, followed by the effect of the trap states on the electrostatics, the energy bands, and the MNR. At the end of the chapter an overview on the phenomenon is given in which both effects are taken into account: the one presented in the previous chapter and the traps one.

4.1 Trap models

The simulated device is the nanowire MOSFET presented in Sec. 3.1 with the same assumptions and approximations. Trap states are implemented in the *effective medium approximation*, i.e. uniformely distributed in the whole silicon volume. It has beeen demonstrated [22] that polycrystalline devices can be modelled with an effective medium approximation in which the grains and their localization are neglected and the traps are considered uniformly distributed in the whole silicon volume. This approach is more and more valid as the number of grains in the volume considered gets higher. That is because the trapped charges at the grain borders form a potential barrier that conduction electrons have to overcome. As the number of potential barriers rise their effect can be mediated.

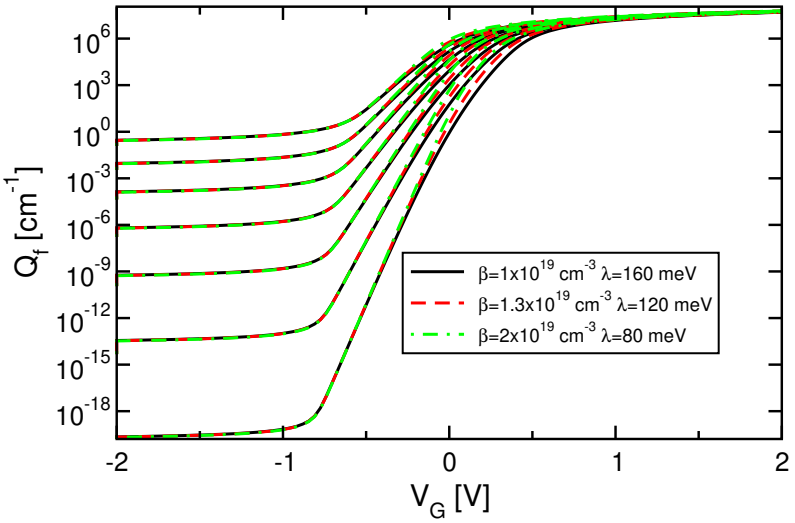
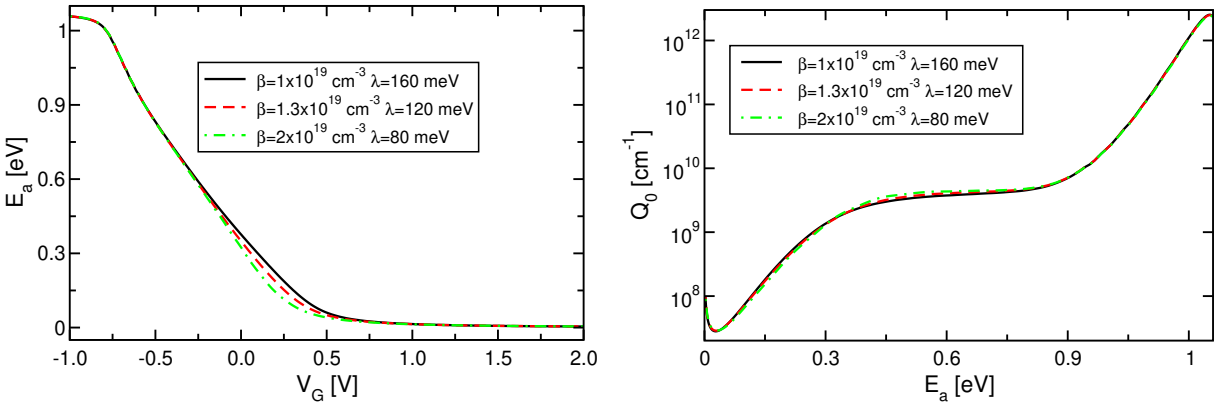


Figure 4.1: $Q_f - V_G$ relations for three different trap distributions.



(a) **(b)**

Figure 4.2: (a) Activation energies as a function of gate bias V_G for the $Q_f - V_G$ relations shown in Fig. 4.1. (b) Prefactors as a function of activation energy E_a for the $Q_f - V_G$ relations shown in Fig. 4.1.

As already mentioned, only deep trap states will be implemented, since they are the only trap states that affect the subthreshold region. The trap state distributions used are purely exponential, like the one in Eq. (2.20). The chosen peak concentrations (β from now on) are similar to the ones presented in the reported work at the end of Chapter 2 [35].

Fig. 4.1 shows the free charges Q_f as a function of the gate bias V_G for different β and decay constants (λ from now on). λ and β have been chosen such that their product is constant and hence the total number of traps in the gap is fixed for fixed volume.

Even if different trap distributions bring slightly different $Q_f - V_G$ relations as shown in Fig. 4.1, the activation energies and prefactors illustrated in Fig. 4.2 show negligible

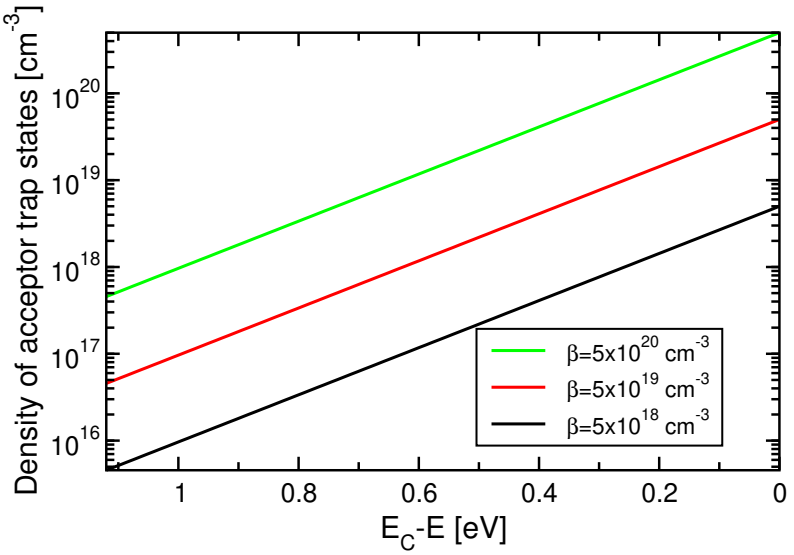


Figure 4.3: Trap state distributions for three different β. These distributions are used throughout this chapter for the Meyer-Neldel analysis.

differences, meaning that the relevant parameter is the total number of traps, and not λ and β. We have therefore chosen to perform the analysis only for $\lambda = 160$ meV and to vary β. In Fig. 4.3 the deep trap state distributions used in the thesis work are shown. We have arbitrarily chosen three β values leaving λ unaltered.

4.2 Simulation results

4.2.1 Free and trapped charges

Fig. 4.4 shows the $Q_f - V_G$ relations for different temperatures and for the three different trap concentrations shown in Fig. 4.3, along with the case without traps for comparison. We can see that the lowest trap concentration almost does not affect the characteristic with respect to the monocrystalline case. The effect of the trap states is more visible with the increase of the traps concentration and completely changes the $Q - V$ curve in the highest concentration case.

The fact that the trap effect gets stronger with the rising of the gate bias is coherent with the trap distribution analytical formulation: we can see from Fig. 4.3 that the number of states rises exponentially going towards the conduction band. With the rising of the gate bias, the electron quasi-Fermi level gets closer to the conduction band, increasing the number of available trap states. This effect is clearly visible in Fig. 4.4 looking at the green curves: for low gate bias the green lines overlap with the black lines obtained without any trap. When the gate bias is high enough, the number of occupied

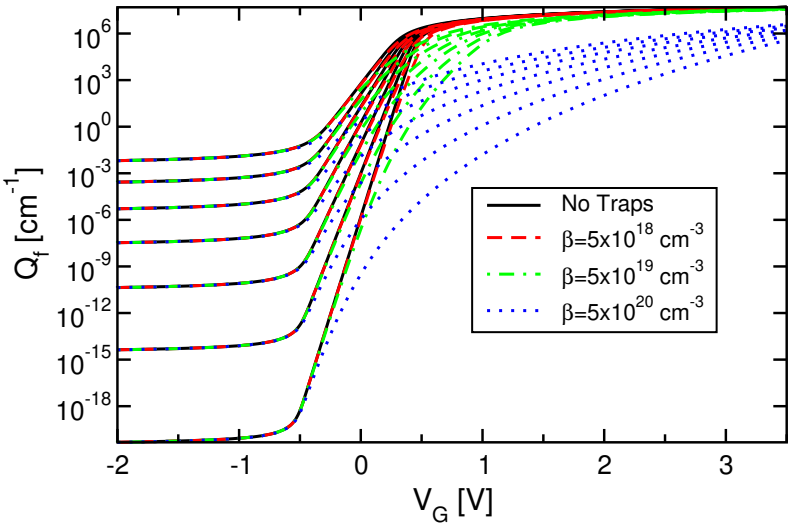


Figure 4.4: $Q_f - V_G$ relations for the three different trap distributions shown in Fig. 4.3. Temperature from 200 K to 500 K with steps of 50 K.

states reaches a value such that the trapped charge creates an electric field that bends the bands and modifies the $Q - V$ relation with respect to the case without traps.

This phenomenon is confirmed by Fig. 4.5. The figure shows, for the lowest temperature, the silicon voltage drop between the middle of the channel and the silicon-oxide interface as defined in Chapter 3. In that chapter the black curve of Fig. 4.5 was shown and it was explained that V_s was blocked to 0 for a wide range of gate bias because, due to the small silicon volume, the induced electrons were not enough to create an electric field that could bend the bands: this is the *blocked bands* region. One can easily see that in the highest trap concentration this region is absent. This happens because the trapped charge is enough to bend the bands even at low gate bias. For the other trap concentrations a blocked band region is present but the point at which the charges start to bend the bands is located at lower biases.

Another important feature that can be seen in Fig. 4.4 is that the free charges with and without traps converge to the same value as the gate bias increases. This suggests that for gate biases high enough the trapped charges are not relevant for the electrostatic. In order to confirm this we can look at Fig. 4.6. We can see that at a certain gate bias the trapped charges saturate, suggesting that the Fermi level reached the conduction band, while the free charges keep on rising reaching a point in which they are the only relevant charges from an electrostatic point of view.

From Fig. 4.6 we can find another proof of the temperature behavior of the trap

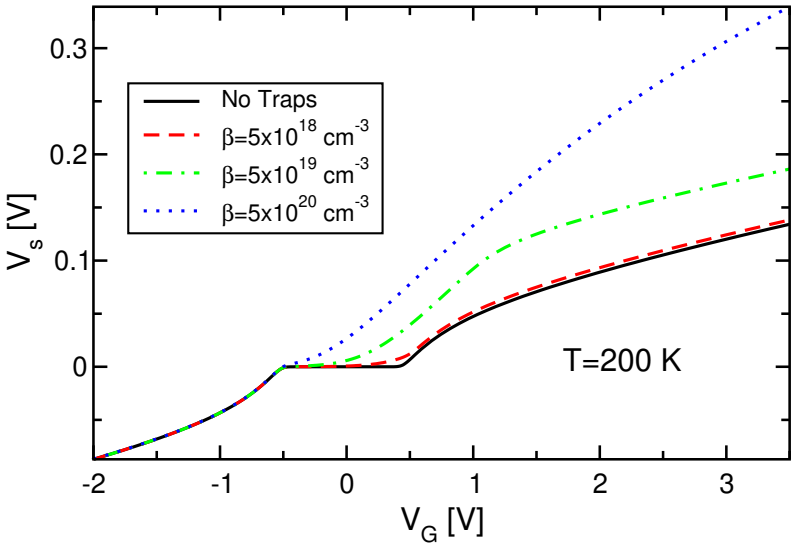


Figure 4.5: Silicon voltage drop V_s at T=200 K between the center of the cylinder and the silicon oxide interface as defined in Chapter 3. For the highest traps concentration there is no blocked bands region.

states: when free charges are irrelevant, in the subthreshold or blocked bands region, the temperature has barely any visible effect on the trapped charge, as pointed out from the theoretical analysis on the trap occupation in Chapter 2. Vice versa, when the free charges become relevant from the electrostatic point of view, different temperatures lead to different amounts of trapped charges. This happens because the free charges are strongly temperature-dependent and the total amount of charges (free and trapped) is substantially determined by the gate bias applied. If the total number of charges has to remain constant and the free charges rise with temperature, the trapped ones have to decrease.

4.2.2 Activation energy and prefactor

In Fig. 4.7 we can see that the curves split up and that their separation gets bigger with the rise of the trap concentration. Morevoer, we can also see that the decrease of the activation energy with the rising of the gate bias gets gentler when a higher number of traps is present. This happens because when the quasi-Fermi level moves inside the energy gap, it modifies the occupancy of the trap states, populating more and more states with electrons as it gets closer to the coduction band. The trapped charge will create the silicon voltage drop shown in Fig. 4.5, rising the potential in the floating silicon volume. Therefore, when trap states are present, we need higher biases to bring the quasi-Fermi level at the same distance from the conduction band that we get without

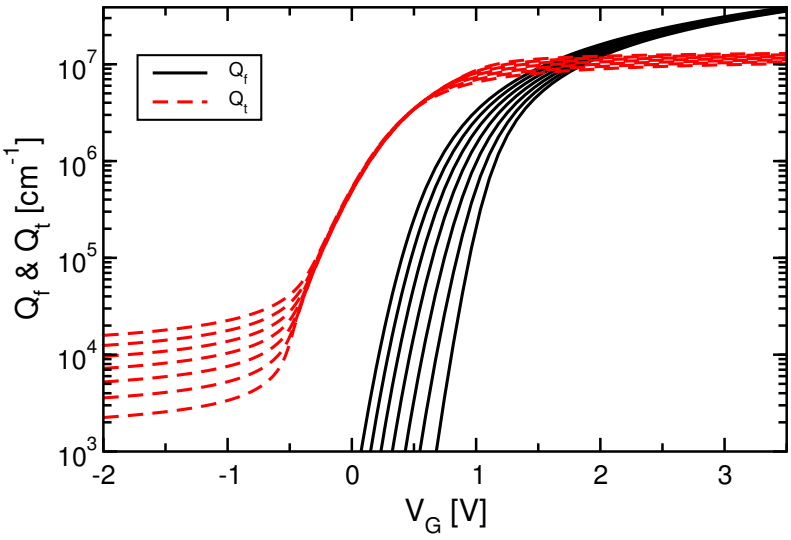


Figure 4.6: Free charges Q_f and trapped charges Q_t as a function of the gate bias V_G for 7 different temperatures from 200 K to 500 K with steps of 50 K.

traps, hence the slower decrease of the activation energy.

Fig. 4.8 shows the prefactors as a function of the activation energies in Fig. 4.7, extracted from the relations in Fig. 4.4. We can see that for the lowest trap concentration almost nothing changed from the monocrystalline case and hence we can ascribe the exponential region of the prefactor to self-consistency of the bands and the extraction method effect shown in the previous chapter. As the trap concentration rises (green curves), we can see that the exponential decrease is gentler with respect to the previous case. To understand what is happening it is useful to look first to the highest trap concentration case. The blue curve has only one slope from $E_a = 0.8$ eV towards 0 eV and the two *pure-subthreshold* and *transition* regions seem to be absent. Looking at Fig. 4.4 we can indeed see that all the curves seem to follow the same relation for each bias, the *transition region* in which some curve follows a pure exponential relation while others are deviating being absent. We infer that all the blue curves are in a *pseudo-subthreshold* region for every bias higher than -0.3 V (when they separate from the other curves) in the graph. The exponential region of the prefactor and hence the MNR, in this case, does not seem to be related to the self-consistency brought by free electrons and the extraction method. The absence of a transition region and therefore the traps related Meyer-Neldel relation of the prefactor will be verified in the following.

Coming back to the green curves we can observe an intermediate behavior, where two different competing phenomena are present: the self-consistency related one presented in the previous chapter and the emergence of a *pseudo-subthreshold* region in which the MNR

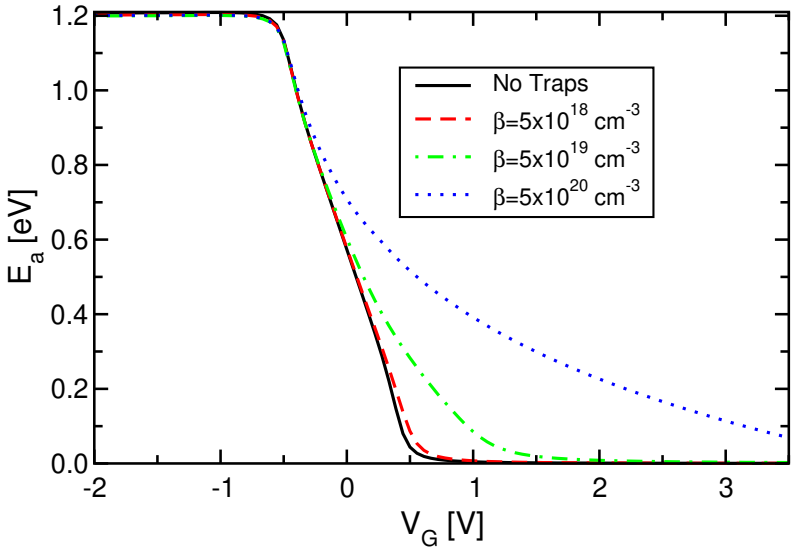


Figure 4.7: Activation energy E_a as a function of the gate bias V_G extracted from the relations in Fig. 4.4.

is respected. The first one gives the steep decrease of the prefactor of the monocrystalline case, the second mitigate this decrease towards gentler slopes like the one of the blue curves.

4.2.3 Arrhenius plots

In Fig. 4.9 the Arrhenius plot of the $Q_f - V_G$ relations for the lowest trap concentration is shown. From Fig. 4.9a it can be seen that the prefactors extracted from the yellow and green curves, that correspond respectively to $V_G = 0.39$ V and $V_G = 0.47$ V, show an exponential decrease, while Fig. 4.9b confirms that this exponential trend derives from the fact that the highest temperatures deviate from the exponential fit.

Fig. 4.10 shows the same Arrhenius plots for the highest trap concentration. The biases chosen for the Arrhenius plots are located in the region in which the activation energies and the charges deviate from the case without traps as one can see from Figs. 4.4 and 4.7. In Fig. 4.8 we can see that an exponential trend is present for all the biases shown in Fig. 4.10. From the magnification illustrated in Fig. 4.10b we can notice that all the points which correspond to the different temperatures follow the fit quite well, the extraction method related effect, due to the deviation from the trend of the highest temperatures is absent in this case.

We have shown the absence of the transition region in the highest trap concentration case. All the $Q_f - V_G$ characteristics follow the same relation, regardless of the tem-

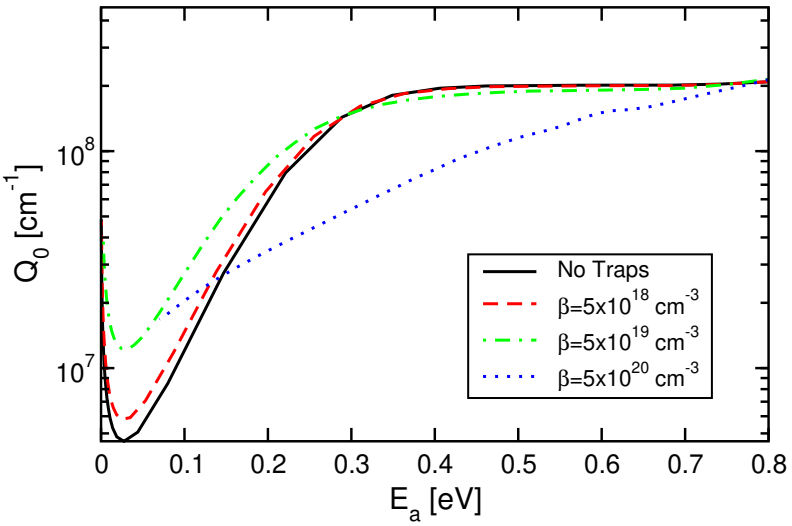


Figure 4.8: Prefactors Q_0 as a function of the activation energies E_a in Fig. 4.7 extracted from the relations in Fig. 4.4.

perature. This phenomenon takes place because the trap concentration is so high that the electrostatics is mainly governed by the trap states for all the biases under analysis. As we have shown by analytical formulas and simulations, the occupancy of the trap states has a very weak dependence on temperature and this is the reason why with the rising of temperature we do not find a deviation from the trend followed by the lower temperatures and hence a transition region that led to the MNR as in monocrystalline case.

The beginning of the *pseudo-subthreshold* takes place at the bias in which the trapped charges have an effect on the bands (Figs. 4.4 and 4.5). The descending trend of the prefactor also starts from the activation energy that corresponds to that bias and hence we can infer that the MNR shown by the blue curve in Fig. 4.8 is again related to a self-consistent effect of the bands. Differently from the case without trap states the electrostatic is here governed by the trapped charge: when enough trap states are populated by electrons, a non neglectable electric field is present and in turn the bands bend. Moreover, the transition region presented in the previous chapter is absent, in its place we can see only one trend that we have called *pseudo-subthreshold*.

4.2.4 Frozen band analysis

In order to find another proof of the presence of two different phenomena as argued in Sec. 4.2.2 about the green curves of Fig. 4.4, we can perform the frozen analysis as previously described in Sec. 3.1.6. In the frozen analysis the bands are calculated at a

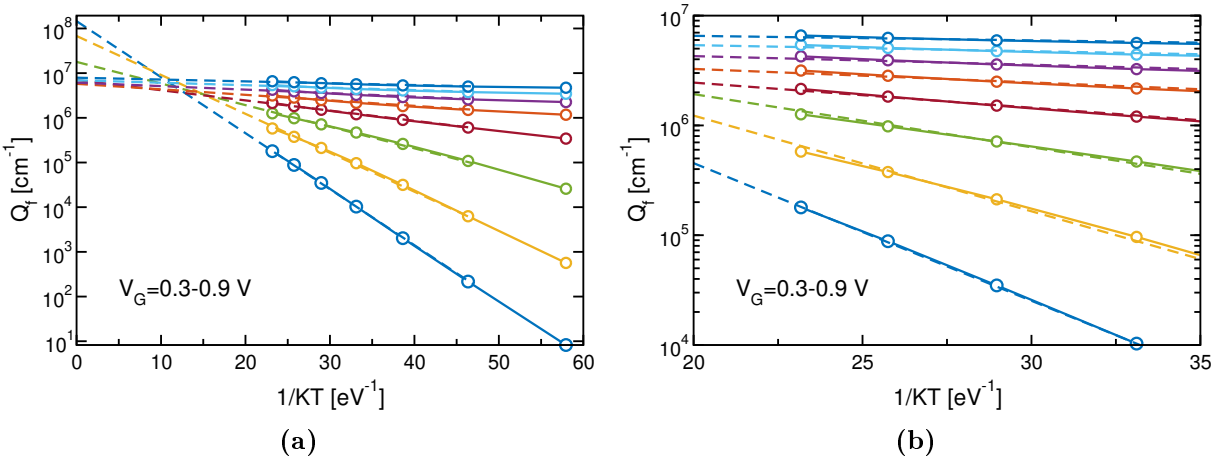


Figure 4.9: (a) Arrhenius plot of the $Q_f - V_G$ relations for the lowest trap concentration. (b) Zoom on the highest temperature of Fig. 4.9a highlighting the deviation from the fit of the yellow and green curves.

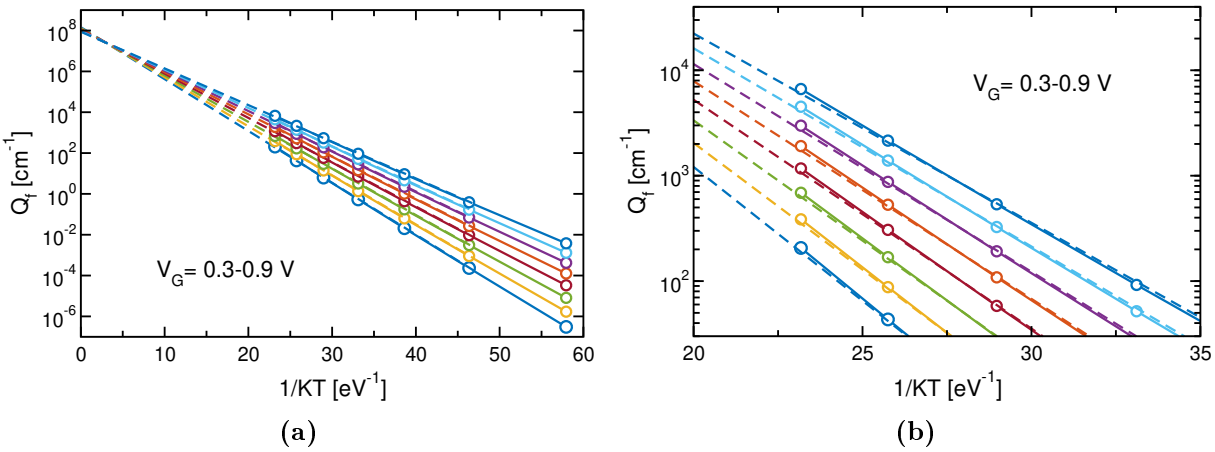


Figure 4.10: (a) Arrhenius plot of the $Q_f - V_G$ relations for the highest trap concentration. (b) Zoom on Fig. 4.9a highlighting the absence of curves deviating substantially from the fit.

reference temperature of 200 K and then they are used for all the temperatures. Since the population of the trap states almost does not depend on temperature, the effect of the trapped charges on the electrostatics is expected to be the same at 200 K and 500 K and hence we do not expect any differences between the frozen and self-consistent case when the electrostatic is controlled by the trapped charges. Moreover, since free charges are strongly temperature-dependent, we expect to shut down their effect on the bands and the related MNR behavior as shown in Chapter 3, but not the effect of trapped charges. Fig. 4.11 shows the $Q_f - V_G$ relations at 500 K with and without the frozen bands for all the trap concentrations and confirms our hypothesis. The fact that the highest trap concentration shows no differences between the frozen and self-consistent case gives us another proof of the fact that, in this case, the free electrons do not affect

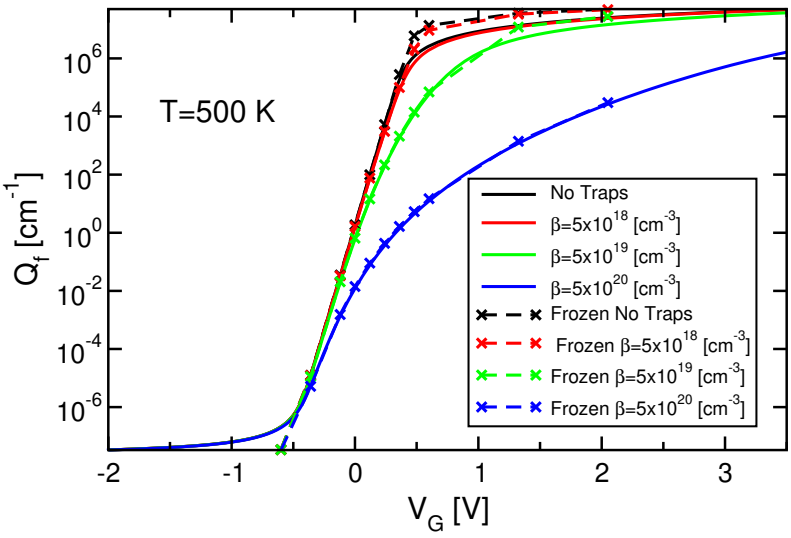


Figure 4.11: $Q_f - V_G$ relations at 500 K with and without the frozen bands for all the trap concentrations. The frozen bands are calculated with the lowest temperature as done in Sec. 3.1.7.

the electrostatics of the system even at high bias; the electrostatics is governed by the trapped charges.

Figs. 4.12 and 4.13 shows activation energies and prefactors computed with and without the frozen bands. We can see from Fig. 4.12 that there are no differences for the highest trap concentration as we expected. Going to lower trap concentrations the curves obtained with and without the frozen bands start to separate when the free charges induced are able to affect the electrostatics.

Fig. 4.13 illustrates the prefactors obtained with and without the frozen bands and confirms the presence of the MNR in systems with trap states and their relation with the self-consistency of the bands. The lowest trap concentration shows the same trend of the monocrystalline case, but with the rising of the trap concentration we can find exponential regions of the prefactors even with the frozen bands. This happens because with the frozen bands we shut down the effect of the free electrons on the bands but we could not do the same for the trapped ones, these electrons create an electric field that bends the bands and the MNR in the frozen case with trap states arise from this self-consistency.

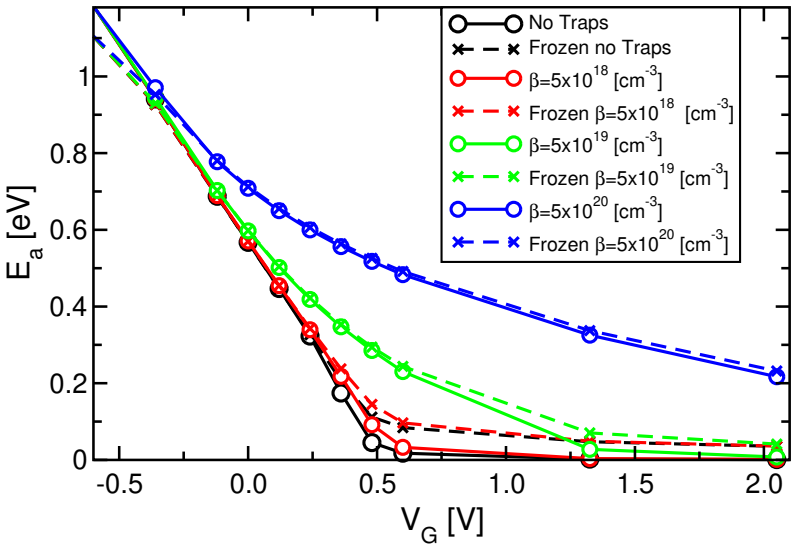


Figure 4.12: Activation energies extracted from the $Q_f - V_G$ relations in Fig. 4.11.

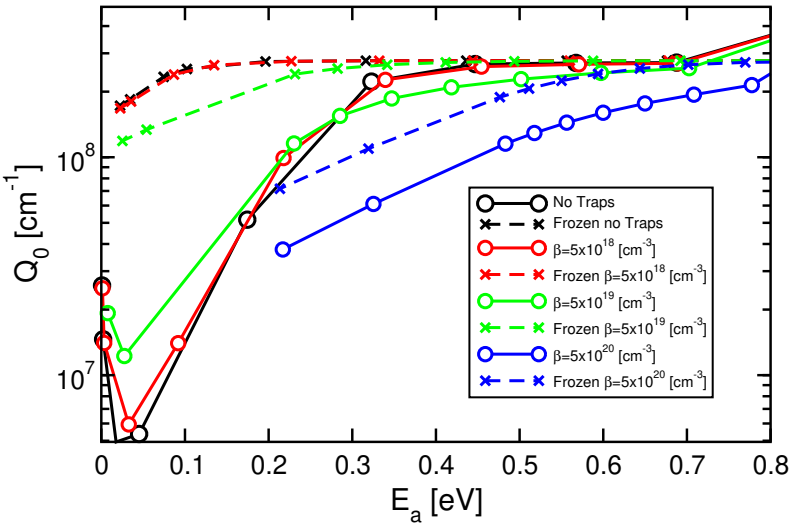


Figure 4.13: Prefactors extracted from the $Q_f - V_G$ relations in Fig. 4.11.

4.2.5 Meyer-Neldel energy

To conclude our analysis on the device with trap states we show here the obtained Meyer-Neldel energies with and without the frozen bands for the two highest trap concentrations; the lowest trap concentration has been neglected since it shows only little differences from the monocrystalline case. The results are shown in Fig. 4.14 and Fig. 4.15.

The fact that the MNR behaviors related to the band bending brought from the free and trapped electrons are different phenomena can be seen in Fig. 4.14: for this trap concentration ($\beta = 5 \times 10^{19}$ cm^{-3}) the band bending derives from both free and trapped electrons. When the bands are frozen we get rid of the effect on the bands of

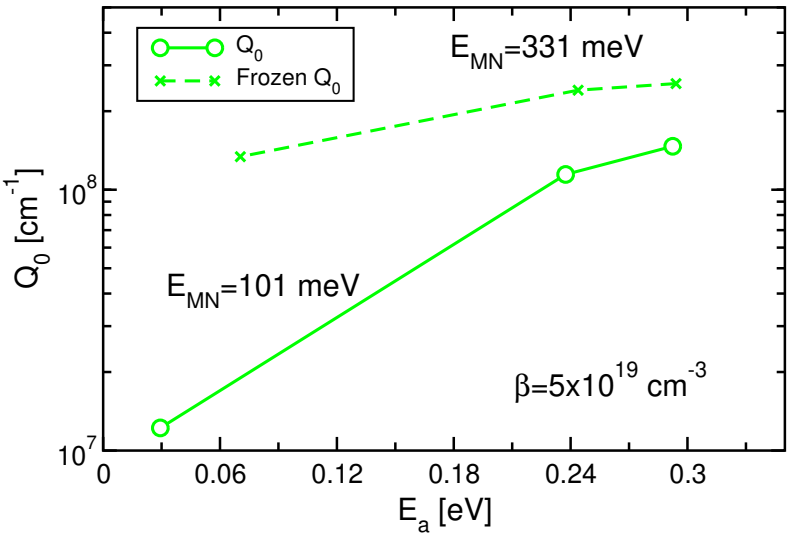


Figure 4.14: Exponential region of the prefactors with and without self-consistency extracted from the $Q_f - V_G$ relations for the trap concentration with $\beta = 5 \times 10^{19}$ cm^{-3}. The Meyer-Neldel energies E_{MN} are also shown.

the free electrons but not of the one related to trapped ones as described in the previous subsection, in fact the resulting E_{MN} for the frozen case (dashed line) is similar to the the ones of the $\beta = 5 \times 10^{20}$ cm^{-3} case (Fig. 4.15) in which the free charges have negligible effects. Vice versa, when the bands are not frozen and the bands bending come from both trapped and free electrons, the slope of the continuous curve in Fig. 4.14 is more similar to the ones shown in Chapter 3 related to the band bending caused by the free electrons, even if the trap presence makes the decrease less steep. The resulting E_{MN} for the continuous line in Fig. 4.14 is a sort of average. As a final evidence of our hypothesis we can see that in Fig. 4.15 the two curves have a very similar E_{MN} since the free electrons are negligible regardless of the frozen bands.

We can see that there is a correlation between the trap concentration and the Meyer-Neldel energy, even if we were not able to write an analytic formulation for this correlation as it has been done in the work reported at the end of chapter 2 [35]. The abscence of an analytical formulation in our trap analysis is related to the lack of closed analytical form for the free charges in a nanowire device with trap states.

We can also see that the E_{MN} that we believe to be related to bands bending caused by the trapped charges is very different from the one obtained in the work cited in Chapter 2, while the E_{MN} that we believe to be related to the self-consistency related to free electrons and the extraction method (the continuous green curve in Fig. 4.14) takes a very similar value. This fact may suggest that the E_{MN} reported in the cited work was not trap-

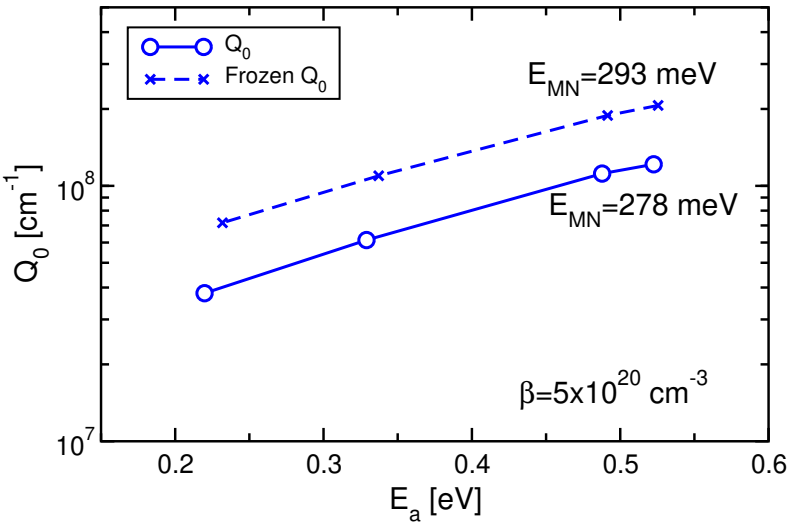


Figure 4.15: Exponential region of the prefactors with and without self-consistency extracted from the $Q_f - V_G$ relations for the trap concentration with $\beta = 5 \times 10^{20}$ cm^{-3}. The Meyer-Neldel energies E_{MN} are also shown.

related, even if the analytical analysis performed in the work predicts the measured value. Deeper investigation on this two competing effects is needed to better understand the origin of the MNR.

Conclusions

In this work we have investigated the Meyer-Neldel rule (MNR) in various MOSFET devices. The MNR is a relation linking the prefactor of an Arrhenius-like relation with the activation energy, and has been widely reported in the literature for amorphous and polycrystalline MOSFET devices. In particular, the Meyer-Neldel energy (E_{MN}, the key parameter of the MNR) has been claimed to be related to the energy distribution of traps within the energy gap.

Our results call this interpretation into question, based on thorough numerical simulation of the temperature behavior of MOSFETs. In particular, we show that even in a MOSFET devoided of any trap, the MNR can still be observed over a suitable bias (hence, activation energy) range. To provide an interpretation of this unexpected result, we conducted ad hoc simulations where the electrostatic of the device was calculated at a reference temperature of 200 K and then the computed bands were used to calculate the charges induced by the raising of temperature without letting them modify the bands accordingly to their presence, in fact shutting down the self-consistency of the bands. Since the population of the trap states almost does not depend on temperature, the effect of the trapped charges on the electrostatics is expected to be the same at every temperature, while free charges are strongly temperature-dependent. We have therefore shown that with these simulations we could not see the effect on the bands of the free charges induced by the raising of temperature, while almost nothing changed with temperature increase when enough trap states were present. In the monocrystalline case, with the aid of the method just described, the MNR was absent. Hence we have found an evidence of the correlation between the self-consistency of the bands and the MNR. This result is an absolute novelty since there is no similar case reported in the literature. The obtained E_{MN} for the two monocrystalline devices have been found to be very similar to the one found in the reported MNR analysis at the end of Chapter 2 that was performed on a polycrystalline TFT.

Since with the method described above we could get rid of the effect of the free

charges we were also able to get rid of the MNR related to their effect on the bands and then to isolate the effect of the trapped charges. We have then discovered that the presence of enough trap states led to the MNR. This fact was again correlated with the self-consistency of the bands: the MN behavior took place at biases for which the trapped charges were enough to bend the bands. The E_{MN} in this case was found to be very different from the one of the monocrystalline case.

We can conclude that our evidence seem to question the interpretation of the MNR that is currently found in the literature. Deeper investigation are needed to verify the relation between the MNR and the self-consistency of the bands and its presence in monocrystalline devices. Moreover, experimental measures are needed to confirm what has been found in the simulations.

Bibliography

[1] C. M. Compagnoni, A. Goda, A. S. Spinelli, P. Feeley, and A. L. Lacaita, "Reviewing the evolution of the nand flash technology: from historical trends to 3-d arrays," in *To appear in the Proc. IEEE*, 2017. 3, 4, 5, 8, 11, 13, 14

[2] C. M. Compagnoni, A. Spinelli, R. Gusmeroli, A. Lacaita, S. Beltrami, A. Ghetti, and A. Visconti, "First evidence for injection statistics accuracy limitations in nand flash constant-current fowler-nordheim programming," in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pp. 165–168, IEEE, 2007. 6

[3] K. Imamiya, H. Nakamura, T. Himeno, T. Yarnamura, T. Ikehashi, K. Takeuchi, K. Kanda, K. Hosono, T. Futatsuyama, K. Kawai, *et al.*, "A 125-mm/sup 2/1-gb nand flash memory with 10-mbyte/s program speed," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 11, pp. 1493–1501, 2002. 8

[4] M. Helm, J.-K. Park, A. Ghalam, J. Guo, C. wan Ha, C. Hu, H. Kim, K. Kavalipurapu, E. Lee, A. Mohammadzadeh, *et al.*, "19.1 a 128Gb MLC NAND-Flash device using 16nm planar cell," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, pp. 326–327, IEEE, 2014. 8

[5] K. Parat and C. Dennison, "A floating gate based 3d nand technology with cmos under array," in *Electron Devices Meeting (IEDM), 2015 IEEE International*, pp. 3–3, IEEE, 2015. 9

[6] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive analysis of random telegraph noise instability and its scaling in deca–nanometer flash memories," *IEEE Transactions on Electron Devices*, vol. 56, no. 8, pp. 1746–1752, 2009. 10

[7] D. James, "Recent advances in memory technology," in *ASMC 2013 SEMI Advanced Semiconductor Manufacturing Conference*, pp. 386–395, May 2013. 12

[8] R. Katsumata, M. Kito, Y. Fukuzumi, M. Kido, H. Tanaka, Y. Komori, M. Ishiduki, J. Matsunami, T. Fujiwara, Y. Nagata, *et al.*, "Pipe-shaped bics flash memory with

16 stacked layers and multi-level-cell operation for ultra high density storage devices," in *VLSI Technology, 2009 Symposium on*, pp. 136–137, IEEE, 2009. 15

[9] J.-D. Lee, J.-H. Choi, D. Park, and K. Kim, "Degradation of tunnel oxide by fn current stress and its effects on data retention characteristics of 90 nm nand flash memory cells," in *2003 IEEE International Reliability Physics Symposium Proceedings, 2003. 41st Annual.*, pp. 497–501, March 2003. 15

[10] M. Park, K. Suh, K. Kim, S. Hur, K. Kim, and W. Lee, "The effect of trapped charge distributions on data retention characteristics of lt;emphasis emphasistype="smcaps" gt;nand lt;/emphasis gt; flash memory cells," *IEEE Electron Device Letters*, vol. 28, pp. 750–752, Aug 2007. 15

[11] J. Jang, H.-S. Kim, W. Cho, H. Cho, J. Kim, S. I. Shim, J.-H. Jeong, B.-K. Son, D. W. Kim, J.-J. Shim, *et al.*, "Vertical cell array using tcat (terabit cell array transistor) technology for ultra high density nand flash memory," in *VLSI Technology, 2009 Symposium on*, pp. 192–193, IEEE, 2009. 16

[12] N. Mielke, H. Belgal, I. Kalastirsky, P. Kalavade, A. Kurtz, Q. Meng, N. Righos, and J. Wu, "Flash eeprom threshold instabilities due to charge trapping during program/erase cycling," *IEEE Transactions on Device and Materials Reliability*, vol. 4, pp. 335–344, Sept 2004. 16

[13] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*. Cambridge university press, 2013. 19, 27, 60

[14] R. M. Swanson and J. D. Meindl, "Ion-implanted complementary mos transistors in low-voltage circuits," *IEEE Journal of Solid-State Circuits*, vol. 7, pp. 146–153, Apr 1972. 20

[15] A. Many, Y. Goldstein, and N. B. Grover, "Semiconductor surfaces," 1965. 23

[16] A. Alnuaimi, A. Nayfeh, and V. Koldyaev, "Electric-field and temperature dependence of the activation energy associated with gate induced drain leakage," *Journal of Applied Physics*, vol. 113, no. 4, p. 044513, 2013. 24, 25, 26

[17] C. Sah, T. Ning, and L. Tschopp, "The scattering of electrons by surface oxide charges and by lattice vibrations at the silicon-silicon dioxide interface," *Surface Science*, vol. 32, no. 3, pp. 561–575, 1972. 24

[18] D. Aspnes *et al.*, "Properties of silicon," *EMIS Datareviews*, 1988. 24

[19] T.-J. King, M. G. Hack, and I.-W. Wu, "Effective density-of-states distributions for accurate modeling of polycrystalline-silicon thin-film transistors," *Journal of applied physics*, vol. 75, no. 2, pp. 908–913, 1994. 27, 28

[20] G. Fortunato, D. B. Meakin, P. Migliorato, and P. G. L. Combers, "Field-effect analysis for the determination of gap-state density and fermi-level temperature dependence in polycrystalline silicon," *Philosophical Magazine B*, vol. 57, no. 5, pp. 573–586, 1988. 27

[21] M. Kimura, "Extraction of trap densities in entire bandgap of poly-si thin-film transistors fabricated by solid-phase crystallization and dependence on process conditions of post annealing," *Solid-State Electronics*, vol. 63, no. 1, pp. 94–99, 2011. 27, 29

[22] M. Hack, J. Shaw, P. LeComber, and M. Willums, "Numerical simulations of amorphous and polycrystalline silicon thin-film transistors," *Japanese journal of applied physics*, vol. 29, no. 12A, p. L2360, 1990. 30, 42, 74

[23] T. Tsuchiya, T. Miura, T. Yamai, G. Kawachi, and M. Matsumura, "Influences of grain boundaries on temperature dependence of device characteristics and on hot carrier effects in low-temperature polycrystalline silicon thin film transistors containing large grains," *Japanese journal of applied physics*, vol. 46, no. 3S, p. 1312, 2007. 33, 34, 35

[24] P. M. Walker, H. Mizuta, S. Uno, Y. Furuta, and D. G. Hasko, "Improved off-current and subthreshold slope in aggressively scaled poly-si tfts with a single grain boundary in the channel," *IEEE Transactions on Electron Devices*, vol. 51, no. 2, pp. 212–219, 2004. 34

[25] B. Faughnan, "Subthreshold model of a polycrystalline silicon thin-film field-effect transistor," *Applied physics letters*, vol. 50, no. 5, pp. 290–292, 1987. 36

[26] B. A. Khan and R. Pandya, "Activation energy of source-drain current in hydrogenated and unhydrogenated polysilicon thin-film transistors," *IEEE Transactions on Electron Devices*, vol. 37, no. 7, pp. 1727–1734, 1990. 36

[27] R. Degraeve, M. Toledano-Luque, A. Arreghini, B. Tang, E. Capogreco, J. Lisoni, P. Roussel, B. Kaczer, G. Groeseneken, J. Van Houdt, *et al.*, "Characterizing grain size and defect energy distribution in vertical sonos poly-si channels by means of a resistive network model," in *Electron Devices Meeting (IEDM), 2013 IEEE International*, pp. 21–2, IEEE, 2013. 36

[28] W. v. Meyer and H. Neldel, "Relation between the energy constant and the quantity constant in the conductivity–temperature formula of oxide semiconductors," *Z. tech. Phys*, vol. 18, no. 12, pp. 588–593, 1937. 37

[29] M. Kikuchi, "The meyer–neldel rule and the statistical shift of the fermi level in amorphous semiconductors," *Journal of applied physics*, vol. 64, no. 10, pp. 4997–5001, 1988. 37, 39, 40

[30] A. Yelon and B. Movaghar, "Microscopic explanation of the compensation (meyer-neldel) rule," *Physical review letters*, vol. 65, no. 5, p. 618, 1990. 38

[31] S. Glasstone, H. Eyring, and K. J. Laidler, *The theory of rate processes*. McGraw-Hill, 1941. 38

[32] B.-G. Yoon, C. Lee, and J. Jang, "Effect of the statistical shift on the anomalous conductivities of n-type hydrogenated amorphous silicon," *Journal of applied physics*, vol. 60, no. 2, pp. 673–676, 1986. 40

[33] P. Stallinga and H. L. Gomes, "Trap states as an explanation for the meyer–neldel rule in semiconductors," *Organic Electronics*, vol. 6, no. 3, pp. 137–141, 2005. 41

[34] M. Shur and M. Hack, "Physics of amorphous silicon based alloy field-effect transistors," *Journal of Applied Physics*, vol. 55, no. 10, pp. 3831–3842, 1984. 41

[35] Z. Han and M. Wang, "Analysis of the meyer-neldel rule based on a temperature-dependent model for thin-film transistors," *IEEE Transactions on Electron Devices*, vol. 64, no. 1, pp. 145–152, 2017. 42, 43, 44, 45, 46, 47, 75, 85

[36] S. D. U. Guide, "Version m-2016.12, synopsys," *Inc., Mountain View, CA*, 2016. 49

[37] W. Gillin and D. Dunstan, "Diffusion in semiconductors," *Computational materials science*, vol. 11, no. 2, pp. 96–100, 1998. 64

[38] G. Kemeny and B. Rosenberg, "Compensation law in thermodynamics and thermal death," *Nature*, vol. 243, no. 5407, p. 400, 1973. 64

[39] F. Lime, O. Moldovan, and B. Iñiguez, "A compact explicit model for long-channel gate-all-around junctionless mosfets. part i: Dc characteristics," *IEEE Transactions on Electron Devices*, vol. 61, pp. 3036–3041, Sept 2014. 66