POLITECNICO DI MILANO
DEPARTEMENT OF PHYSICS
DOCTORAL PROGRAM IN PHYSICS

INTEGRATED COMPLEMENTARY GRAPHENE INVERTERS IN ANALOG AND DIGITAL ELECTRONICS

Doctoral Dissertation of:
Massimiliano Bianchi

Supervisor:
Prof. Roman Sordan

Tutor:
Prof. Lamberto Duò

The Chair of the Doctoral Program:
Prof. Paola Taroni

2014 – XXVI Doctoral Cycle

Abstract

CMOS technology, from its introduction in 1963, has found applications in analog and digital electronics and nowadays it represents the main building block of modern electronics. The performances of CMOS devices depend entirely on the characteristics of the silicon metal-oxide-semiconductor field-effect transistors (MOSFETs). So far the scaling of the dimensions of the transistors has allowed to reach the needed improvements required by the market. However, the MOSFET scaling is approaching a limit in which a further shrinking of the devices will not be possible anymore. For that reason it will be necessary to introduce new materials in order to continue with the increase of the performances of electronic devices.

From its discovery graphene has been considered as the material that could replace silicon in future CMOS technology. Graphene is the name given to a flat monolayer of carbon atoms tightly packed into a two-dimensional honeycomb lattice. Scientific and technological interests in graphene have rapidly grown because of its extraordinary electronic properties, such as high-mobility of charge carriers which can travel thousands of inter-atomic distances without scattering. Andre Geim and Konstantin Novoselov were the first to perform groundbreaking experiment on graphene sheets for which they were awarded the Nobel Prize in Physics in 2010. Graphene has significant potential for electronic applications but, unfortunately, its applicability is currently limited by the fact that it is a semi-metal. For that reason graphene field effect transistors (GFETs) cannot be turned-off and that mainly limit use of graphene in niche applications of analog electronics.

In this thesis the effects of short channel length and width on the performances of the Si MOSFETs are analyzed and a direct comparison with GFETs is made. It is shown that it is difficult to reach a saturation region of the drain current in GFETs. This influences the voltage gain which has usually been small in graphene circuits preventing signal amplification. GFETs must exhibit an over-unity intrinsic voltage gain in order to allow the realization of more complex

circuits. This thesis is focused on the realization of realistic graphene digital circuits. All steps that allow to reach that goal are described and a comparison with Si CMOS is made.

Graphene voltage amplifiers were fabricated starting from mechanically exfoliated graphene flakes. The use of very thin top-gate dielectrics allowed the realization of graphene complementary inverters with a high voltage gain (>10 dB) operating at room temperature. Such devices could be used as the main building block of both analog and digital graphene electronics. For instance, the fabricated graphene inverters could be used in audio applications since it is able to maintain an over-unity voltage gain in the audio frequency range.

In graphene digital logic gates it is also important to have an over-unity voltage gain in order to match the input and output signals. Without signal matching it is not possible to cascade different logic gates and realistic digital circuits cannot be realized. In order to allow the scalability of the logic gates, CVD graphene was used. It is shown that it was possible to match the input and output signals at room temperature for the first time ever. This also allowed to cascade different logic gates, which has not been demonstrated at any temperature so far.

The obtained results were the pre-requisite for the realization of realistic graphene digital circuits, out of which the most important classes are the ring oscillators (ROs). ROs represent the perfect device to test the ultimate performances limits of a given technology. The first graphene ROs operating at gigahertz frequencies have demonstrated. Performance increased by reducing gate length of the GFETs in ROs was also demonstrated.

However it is shown that graphene digital circuits cannot compete directly with Si CMOS even though the same gate delays are obtained in both cases. Graphene transistors have large static power dissipations compared to Si CMOS which is a low-power consumption technology. Graphene instead could find applications in extremely high-frequency devices where static power dissipation is not an issue. Graphene could replace InP in high-frequency electronics due to its higher mobility but for this to happen further technological advances are needed.

Contents

Chapter 1

Field-Effect Transistors

1.1 Introduction

The metal-oxide-semiconductor field-effect transistor (MOSFET) is the most important member of the family of the field-effect transistors and it represents also the main building block of VLSI technology in microprocessors and semiconductor memories. The MOSFET is an unipolar or majority-carrier device since the current is transported mainly by carriers of only one polarity (electrons in n-channel devices and holes in p-channel devices).

It was proposed for the first time by J. E. Lilienfeld in the early 1930s, but it became practical much later in the early 1960s and at first only n-type transistors were produced. The revolution in semiconductor industry began in 1963, thanks to Frank Wanlass, when the complementary MOS (CMOS) technology was introduced, where n-channel and p-channel MOSFETs are fabricated side by side on the same substrate.

The principal division of semiconductor electronics are analog and digital circuits. CMOS devices found application in both fields but it rapidly captured the digital market. That is due to the fact that CMOS gates dissipate power only during switching and require only few devices. CMOS technology became attractive also thanks to low fabrication costs and to the possibility of placing both analog and digital circuits on the same chip, improving the performances and reducing the cost of packaging.

The modern digital logic is completely based on that type of technology, where different type of transistors are combined in order to perform a certain logic operation. So we can say that nowadays digital logic circuits depends entirely on the performance of a single type of device: the MOSFET.

Till now the scaling of the CMOS devices has provided the needed performance

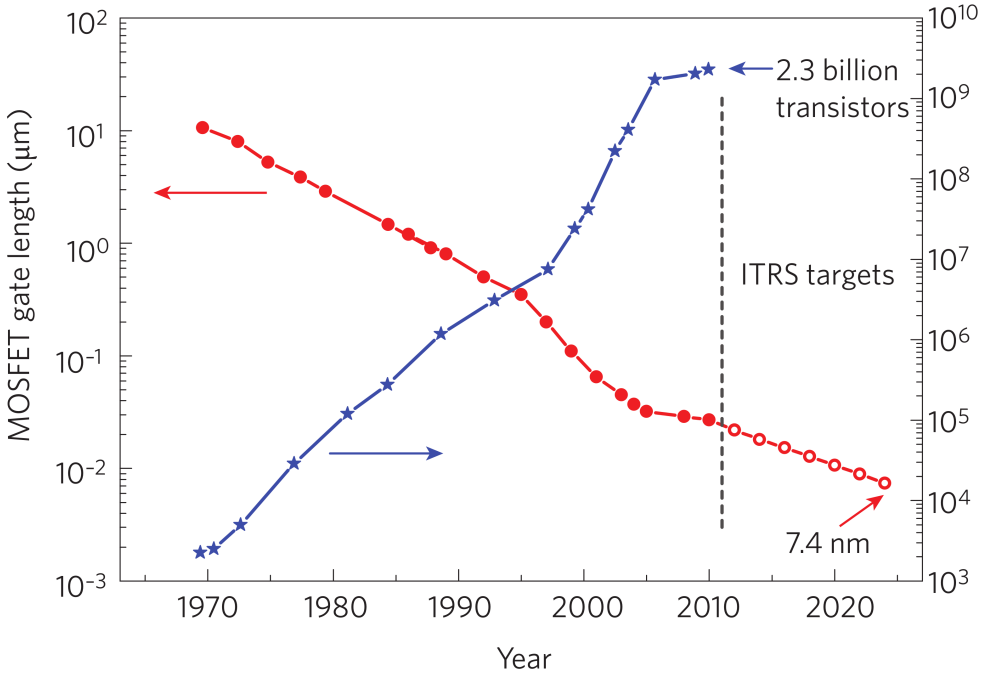


Figure 1.1.1: MOSFET channel length scaling in integrated circuits (filled red circles) and International Technology Roadmap for Semiconductor (ITRS) targets (empty red circles) together with the number of transistors per processor chips (blue stars). Taken from reference [31].

improvements required by the market from one generation of integrated circuits to the next one. Figure 1.1.1 show the progress made by the companies in shrinking the dimension of the devices from 1970 to nowadays. However it seems now that the MOSFET scaling is approaching its limit and that in the future it will be necessary to introduce new materials and device concepts in order to have performance improvements. In this view graphene could be the material of the future CMOS technology.

The aim of this chapter is to introduce the basic knowledge about physics and modeling of standard silicon MOSFETs and to compare these devices with graphene field effect transistors (GFETs).

1.2 Physics and characteristics of MOSFETs

The basic structure of an n-channel MOSFET is shown in figure 1.2.1a. It can be seen as a four-terminal device that is formed by a p-type semiconductor substrate, which is named bulk (B), into which there are two n^+ regions, that formed the source (S) and drain (D). A metal contact on an insulator is named gate (G) and it acts as a control electrode [18, 35, 3, 36]. Figure 1.2.1b shows the device

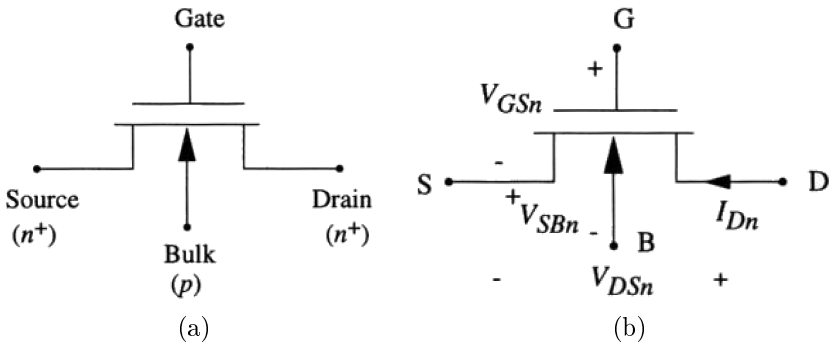


Figure 1.2.1: n-channel MOSFETs (a) symbols, (b) current and voltages. Taken from reference [18].

voltages, where V_{GSn} represents the gate-source voltage that is used to control the value of the drain current I_{Dn} which flows through the device from drain to source.

The main part of the device, as it can be seen in figure 1.2.2a, consists of a metal-oxide-semiconductor (MOS), which is formed by the gate (G) on top of an insulating layer (typically silicon dioxide). Underneath the gate there is a p-type silicon epitaxial layer (S) on top of a p^+ substrate. The physics of the MOS structure, when is coupled with the source and the drain, is the origin of the I-V characteristic of the device.

There are several basic device parameters that have to be considered for the circuits design. One of them is the channel length L of the MOSFET, which is defined by the distance between the two n^+ regions shown in figure 1.2.2a, since this critical dimension influences the electrical characteristics of the device. There is also the channel width W, indicated in figure 1.2.2b, which is the region that supports the current flow between source and drain, that together with the channel length defines the aspect ratio, defined as W/L. Other important parameters are the insulator thickness x_{ox}, the substrate doping N_a and the drawn channel length L' (figure 1.2.2b), that is effective length of the gate electrode and is larger than the channel length L.

Another extremely important parameter of a MOSFET is the threshold voltage V_T and it is used as a reference for the I-V characteristics of the device. Typically the transistors in CMOS design are based on enhancement-mode (E-mode) where V_{GSn} is used to enhance the conduction between the source and the drain. This parameter is particularly relevant for high performance circuits and in an n-channel E-mode transistor it assumes a positive value ($V_{Tn} > 0$), which range usually from 0.5 V to 0.9 V.

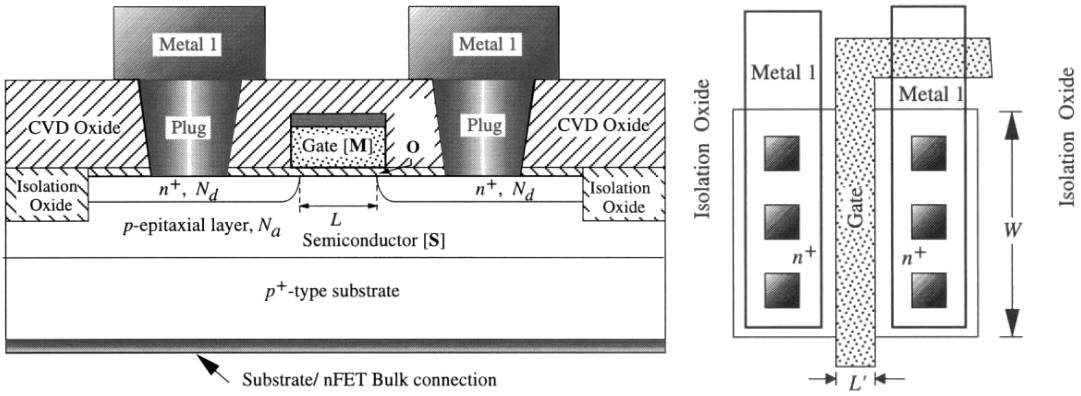


Figure 1.2.2: nFETs (a) cross-sectional view, (b) top view. Taken from reference [18].

In an ideal n-channel MOSFET the transistor is placed in a cut-off state when the gate source voltage has the value $V_{GSn} < V_{Tn}$, and ideally the current that flows into the channel is zero. Instead if the voltage of the gate source is increased to $V_{GSn} > V_{Tn}$ the transistor is in the active mode and it can conduct a current I_{Dn}. In practice the value of V_{GSn} relative to V_{Tn} determines the value of the current I_{Dn} inside the channel and so if the transistor is in a ON or OFF state.

The physics of a p-channel MOSFET will not be described since if is the complement of the one that will be used for an n-channel device. The p-channel MOSFET is obtained by changing n-type regions of source and drain into p-type, p-type substrate into n-type, reversing the electrons with holes, the polarities of the voltages and the direction of the current.

1.2.1 The threshold voltage

The central MOS structure,which has the characteristics of a simple capacitor, is the part of the MOSFET that can allow the conduction between source and drain. This unique property is due to the fact that an electric field can penetrate into a semiconductor for a small distance and this can change the charge distribution at the the surface.

Typically the oxide capacitance per unit area is expressed as $C_{ox} = \varepsilon_{ox}/x_{ox}$ in units of F/cm^2, where ε_{ox} is the oxide permittivity. In current technologies the oxide thickness is less then 10 nm, which gives values of the oxide capacitance of the order of 10^{-7} F/cm^2 or even bigger. In this type of device it is better to have thin oxides in order to increase capacitance since in this way the conduction in the MOSFET can be better controlled.

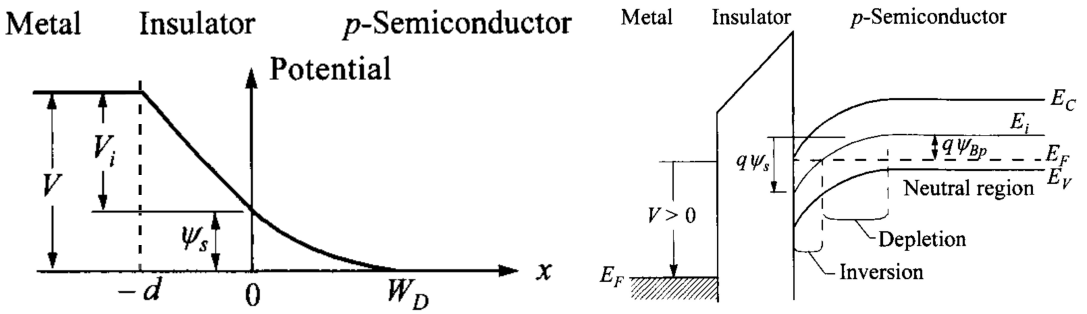


Figure 1.2.3: (a) potential distribution in a MOS structure, (b) band diagram of the MOS under strong inversion. Taken from reference [35].

With the aim of trying to find the origin and the characteristics of the threshold voltage, let us start at the beginning to take into account the basic MOS structure. The gate voltage V_G is used to control the charge carrier density at the semiconductor surface and in particular negative charge is induced in the semiconductor region under the oxide when $V_G > 0$. This is due to the field-effect: using the external voltage an electric field can be created and it can penetrate into the semiconductor and control the charge densities. Using Kirchoff's voltage law on the circuit the gate voltage can be expressed like

$$V_G = V_{ox} + \phi_S \tag{1.2.1}$$

where V_{ox} is the voltage across the oxide and ϕ_S is the surface potential whose behavior is shown in figure 1.2.3a. From the expression 1.2.1 can be understood that if V_G is increased, also ϕ_S is increased and this gives stronger electric field in the semiconductor.

The surface charge density Q_S can be defined as the total charge at the surface of the semiconductor into the p-type bulk and is expressed with units of C/cm^2. As it can be seen in figure 1.2.4a, for small values of V_G a depletion region is created and it is formed by a negative space charge in order to support the electric field. This type of charge is due to ionized acceptor atoms and it takes the name of bulk charge, that can be written as

$$Q_B = -\sqrt{2q\varepsilon_{Si}N_a\phi_S}. \tag{1.2.2}$$

In this formula, ε_{Si} is the permittivity of the silicon, N_a is the acceptor doping density in the substrate and q is the elementary charge. In this case Q_S is completely formed by the bulk charge, which means $Q_S = Q_B$, and since it is

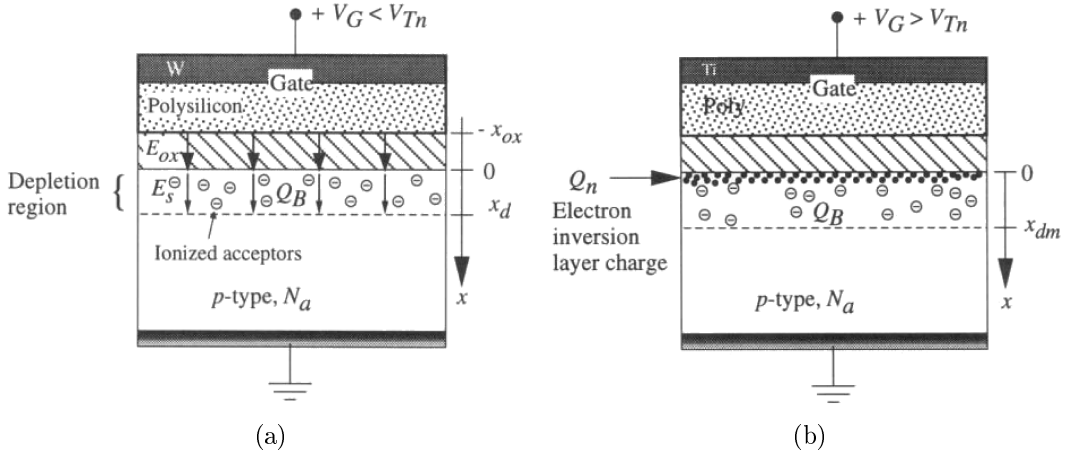


Figure 1.2.4: (a) depletion and (b) inversion in a MOS structure. Taken from reference [18].

composed by ionized atoms it is an immobile charge.

When the gate voltage is increased to reach the threshold voltage ($V_G = V_{Tn}$), at the surface of the semiconductor a thin layer of electrons, called inversion layer, is formed with a surface charge density Q_n. In the case $V_G > V_{Tn}$ the inversion charge Q_n will contribute to the surface charge density, which is given now by

$$Q_S = Q_B + Q_n. \tag{1.2.3}$$

This operational mode is called inversion and is characterized by the charges present in figure 1.2.4b.

The depletion and inversion mode could also be described by the point of view of the energy band across the MOS. When a positive voltage is applied to the metal, its Fermi level moves downward and in this way it creates a field in the oxide layer which accelerates the negative charges in the direction of the metal electrode. A similar field is induced in the p-semiconductor and causes the band bending at the surface. If the voltage is high enough, the band bending is so strong that it can create a wide depletion region and an inversion region as shown in figure 1.2.3b. In the case of depletion the valence band close to the surface is more far away from the Fermi level with respect to the valence band in the bulk. The holes concentration close to the surface is lower than the one in the bulk and it can be seen as if the holes are repelled from the interface with the oxide by the positive voltage. For inversion the conduction band at the surface is closer to the Fermi level compared to the valence band and the holes are depleted from the surface. This condition is energetically favorable for electrons to populate the

conduction band and the surface behaves like an n-type material.

Differently from bulk charge, inversion charge is made by mobile electrons that can move in the direction parallel to the surface. If V_G is increased further the charge in the depletion region Q_B remains approximately constant while Q_n continues to increase, providing a larger current from source to drain.

In this mode the surface potential ϕ_S that is needed to form the inversion layer can be written as

$$\phi_S \approx 2\left|\phi_F\right| = 2\left(\frac{kT}{q}\right) ln\left(\frac{N_a}{n_i}\right) \tag{1.2.4}$$

where ϕ_F is the bulk Fermi potential and $n_i \approx 1.45 \times 10^{10} cm^3$ is the intrinsic density of silicon. From the equation 1.2.4 it can be seen that the value of the surface potential depends on the substrate doping N_a, which is typically equal to 10^{15} cm^3 and gives values of $\phi_S \approx 0.58$ V.

In order to find an expression for the threshold voltage it has to be taken into account the case when the inversion layer has just started to form, which means $V_G = V_{Tn}$ and $Q_n \approx 0$, so that Q_B is the total charge at the surface. Using the capacitive relation $Q = CV$ together with equations 1.2.2 and 1.2.4, the oxide voltage is given by

$$V_{ox} = \frac{|Q_B|}{C_{ox}} = \frac{\sqrt{2q\varepsilon_{Si}N_a\left(2\left|\phi_F\right|\right)}}{C_{ox}}. \tag{1.2.5}$$

Due to the Kirchhoff equation the ideal threshold voltage is given by

$$V_{Tn}^{ideal} = 2\left|\phi_F\right| + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_{Si}N_a\left(2\left|\phi_F\right|\right)} \tag{1.2.6}$$

which is ideal expression because it takes into account that the MOS structure is a perfect insulator and ignores the fact that the materials that form the substrate and the gate are usually different.

To find an expression for the threshold voltage that can be used for realistic MOS structure two effects has to be taken into account: charge traps in the oxide that can modify the electric field and the difference in terms of electrical characteristics of the different materials of the gate and the substrate. To do this the flat-band voltage has to be introduced and it can be written as

$$V_{FB} = (\phi_G - \phi_S) - \frac{1}{C_{ox}}\left(Q_f + Q_{ox}\right) \tag{1.2.7}$$

where $(\phi_G - \phi_S)$ is the difference of the work function between the gate and the substrate, Q_f is the surface charge density at the interface between insulator and

semiconductor and Q_{ox} is the charge trapped in the oxide. Q_{ox} mostly originates from the alkali impurity ions, such as Na^+ and K^+, that are trapped in the oxide. These kind of impurities are able to move under an applied electric field and in this way they can generate unstable threshold voltages. To reduce the effect of trapped charge the oxidation process is performed in chlorinated atmosphere, so the alkali contaminants form neutral NaCl and KCl salts do not affect the V_{Tn} stability. Q_f instead is due to the change in composition passing from silicon to silicon dioxide and in cannot be removed, it can only be minimized by the use of a thermal annealing.

Now incorporating the flat-band voltage into equation 1.2.6, V_{Tn} can be expressed as

$$V_{Tn} = V_{FB} + 2\left|\phi_F\right| + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_{Si}N_a\left(2\left|\phi_F\right|\right)}. \tag{1.2.8}$$

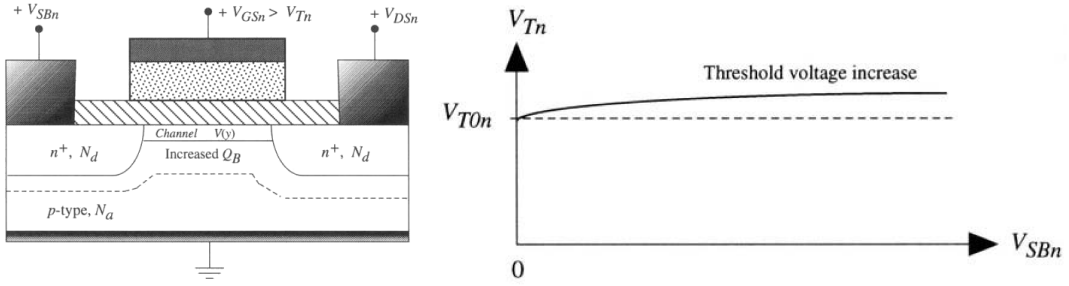
However, with standard processing condition, the flat-band voltage assumes typically negative values and usually it brings to negative threshold voltages. For standard CMOS circuits V_{Tn} should have positive values since positive values of power supply are used. This problem can be solved implanting acceptor ions into the substrate, with a dose D_1 that gives the number of inons/cm^2. Using this process, called threshold adjustment ion implant, an additional bulk charge is introduced at the surface and this induce a positive shift of the threshold voltage that can reach typical working values of 0.7 V. With that adjustment in the inversion regime the inversion charge density can be expressed like

$$Q_n = -C_{ox}\left(V_G - V_{Tn}\right) \tag{1.2.9}$$

which modifies the equation 1.2.8 as follows

$$V_{Tn} = V_{FB} + 2\left|\phi_F\right| + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_{Si}N_a\left(2\left|\phi_F\right|\right)} \pm \frac{qD_1}{C_{ox}}. \tag{1.2.10}$$

Now that an expression for the threshold voltage in a MOS structure was found, it should be updated in the case of a MOSFET. In fact, even if the value of V_{Tn} is similar to the one of the MOS structure, equation 1.2.10 should be modified in order to take in account the application of a voltage between source and drain. Considering the MOSFET device in figure 1.2.5a, with the p-type substrate grounded, the application of source-bulk voltage V_{SBn}, induces a shift of the threshold voltage to higher values. This is due to the fact that V_{SBn} introduces a reverse-bias at the interface between the p-substrate and the n-channel, which increases the bulk depletion charge Q_B. This is called the body-



(a) (b)

Figure 1.2.5: (a) MOSFET biased in inversion mode, (b) body bias effect. Taken from reference [18].

bias effect and as a result the threshold voltage can be written as

$$V_{Tn} = V_{FB} + 2\left|\phi_F\right| + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_{Si}N_a\left(2\left|\phi_F\right| + V_{SBn}\right)} \pm \frac{qD_1}{C_{ox}}. \tag{1.2.11}$$

If zero-body bias threshold voltage V_{T0n} is defined, applying V_{SBn} the threshold voltage is increased by

$$\triangle V_{Tn} = (V_{Tn} - V_{T0n}) = \gamma\left(\sqrt{2\left|\phi_F\right| + V_{SBn}} - \sqrt{2\left|\phi_F\right|}\right) \tag{1.2.12}$$

where γ is the body-bias factor in units of $V^{1/2}$ and it can be written as

$$\gamma = \frac{1}{C_{ox}}\sqrt{2q\varepsilon_{Si}N_a}. \tag{1.2.13}$$

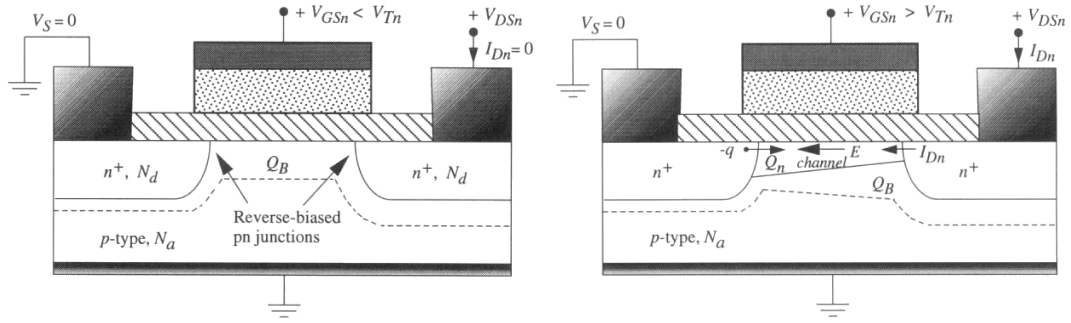
In this way, using equations 1.2.11 and 1.2.12, the threshold voltage is given by

$$V_{Tn} = V_{T0n} + \gamma\left(\sqrt{2\left|\phi_F\right| + V_{SBn}} - \sqrt{2\left|\phi_F\right|}\right) \tag{1.2.14}$$

which is plotted in figure 1.2.5b in which the square-root dependance of V_{Tn} can be observed.

1.2.2 Current-voltage characteristics

The I-V characteristics of MOSFETs can be derived using some basic assumptions: the MOS capacitor structure is an ideal insulator, the are no interface traps nor mobile oxide charges, only drift current will be taken into account, the doping in the channel is uniform and everything will be treated using the gradual-channel approximation. The last condition is the most important and it assumes that the longitudinal field E_y (along the channel) is much less then transverse



(a) (b)

Figure 1.2.6: MOSFET charges in (a) OFF state and (b) ON state. Taken from reference [18].

field E_x (perpendicular to the channel) and is valid for long channel length with dimensions around 20 μm.

Consider the case in which the MOSFET is in OFF state, when $V_{GSn} < V_{Tn}$ as shown in figure 1.2.6a. In this mode the gate voltage is not high enough to induce the formation of the inversion layer, so the only charge present is the bulk charge Q_B below the oxide. Between the drain and the source there are two p-junctions, which have respectively a reverse-bias across it and a zero-bias applied, that blocks the flow of current, resulting in $I_{Dn} \approx 0$. While the situation in the ON state is completely different. In fact with $V_{GSn} \geq V_{Tn}$, the electron inversion layer is formed and allows the formation of the conduction channel from the drain to the source as it can be seen in figure 1.2.6b. In this way when drain to source voltage V_{DSn} is applied the channel electric field is formed. The field forces the electrons to move from the source to the drain resulting in a drift current I_{Dn} in the opposite direction, that can be calculated in function of V_{GSn} and V_{DSn}. In the channel, in this regime, there is an electron inversion charge given by

$$Q_n = -C_{ox} \left[V_{GSn} - V_{Tn} - V(y) \right] \tag{1.2.15}$$

where $V(y)$ is the electrical potential function generated by V_{DSn} that creates an electric field in the channel $E = dV/dy$. The negative sign in the equation 1.2.15 is needed since the channel consists of negative electrons, in this way $Q_n < 0$. The corresponding boundary conditions of the channel potential are

$$\begin{aligned} V(0) &= V_S = 0 \\ V(L) &= V_{DSn}. \end{aligned} \tag{1.2.16}$$

In order to find the I-V equations for the MOSFET it can be used the channel

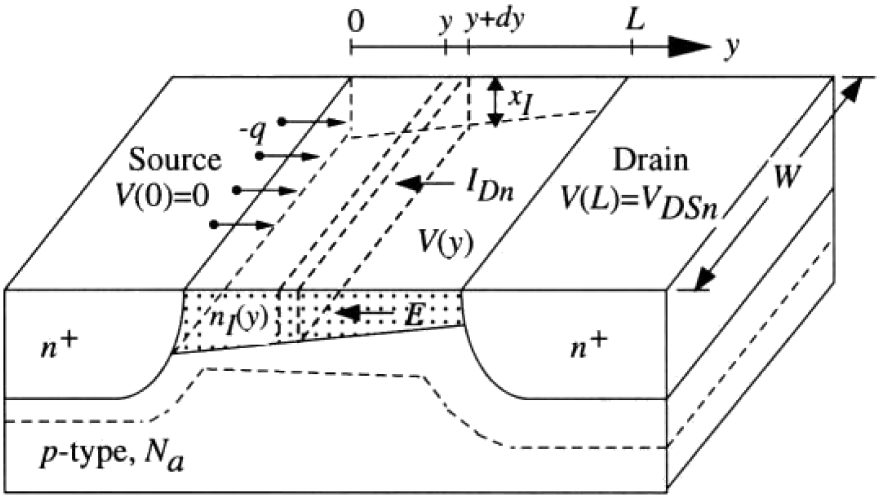


Figure 1.2.7: Channel geometry of the MOSFET. Taken from reference [18].

geometry shown in figure 1.2.7. Note that the channel act as a resistor and by considering a differential segment dy with a rectangular shape, through which the current I_{Dn} flows, the resistance is given by

$$dR = \frac{dy}{\sigma A_l} \tag{1.2.17}$$

where σ is the conductivity of the section and A_l is the cross-sectional area. This area could be expressed as $A_l = Wx_l$, where W is the channel width and x_l the thickness of the channel inversion layer. The conductivity is given by $\sigma = q\mu_n n_l$, where μ_n is the electron surface mobility and n_l is the electron density of the channel. Combining these two informations the denominator of equation 1.2.17 can be written as

$$\sigma A_l = q\mu_n n_l W x_l = -\mu_n W Q_n. \tag{1.2.18}$$

At this point the voltage that drops on the segment of the channel dy is

$$dV = -\frac{I_{Dn} dy}{\mu_n W Q_n} \tag{1.2.19}$$

where the negative sign is due to the fact that the current flows in the -y direction. Substituting Q_n and integrating from $y = 0$ to $y = L$ the general expression for I_{Dn} is expressed by

$$I_{Dn} = k_n \left(\frac{W}{L}\right) \int_{V=0}^{V_{DSn}} \left[V_{GSn} - V_{Tn} - V(y)\right] dV \tag{1.2.20}$$

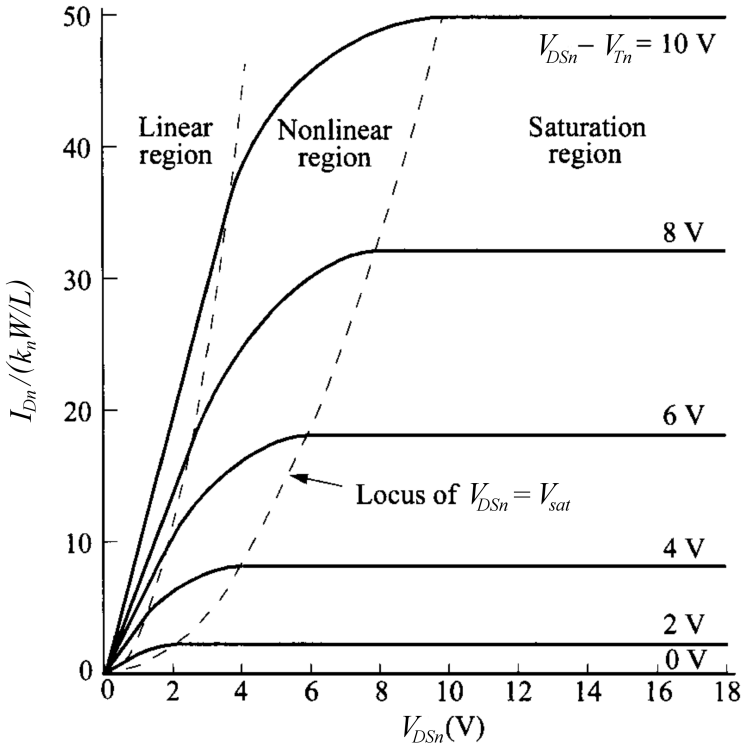


Figure 1.2.8: I-V characteristics of MOSFET, where the dash lines separate the different regions. Taken from reference [35].

where $k_n = \mu_n C_{ox}$ in units of A/V^2 is the process parameter. Since the aspect ratio (W/L) is an important parameter of the geometry of the MOSFET that influence the current of the device, it is worthed to introduce the device parameter that can be written as

$$\beta_n = k_n \left(\frac{W}{L} \right). \tag{1.2.21}$$

Starting from equation 1.2.20 and assuming V_{Tn} constant in the channel it is possible to find the the basic I-V characteristics of an ideal MOSFET using the square-law model. This approach predicts that at the beginning, for a given value of V_{GSn}, the drain current I_{Dn} increase linearly with the drain voltage V_{DSn} (linear region), then it starts to level off (nonlinear region) and at the end it approaches a saturated value (saturation region). In figure 1.2.8 it is possible to see the three different regions and in particular the dash line on the right it corresponds to the values of V_{DSn} where the drain current reaches its maximum value I_{Dsat}.

Now the different operations of the device will be discussed in order to better understand what happens in every single region. Start to consider the case in which a positive voltage is applied to the gate, high enough to obtain an inversion at the interface of the semiconductor. With a drain-source voltage applied, the

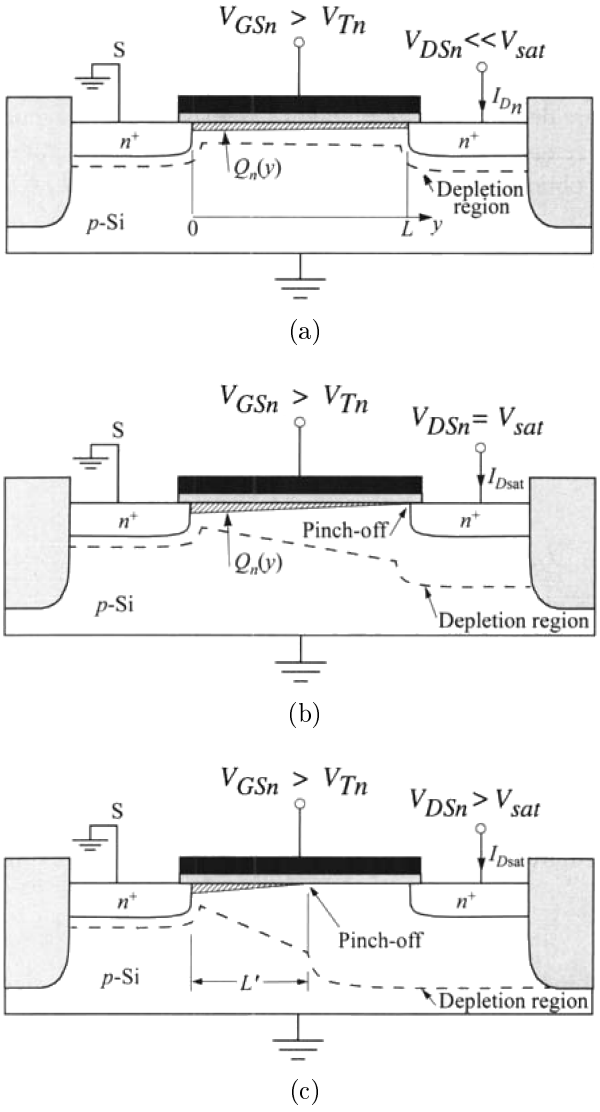


Figure 1.2.9: MOSFET operating in (a) linear region, (b) on set of saturation when $V_{DSn} = V_{sat}$ and (c) in saturation region. Taken from reference [35].

conductive channel will be formed and in this way the current can flow from source to drain, as shown in figure 1.2.9a.

If a small V_{DSn} is taken into account equation 1.2.20 can be reduced to

$$I_{Dn} = \frac{\beta_n}{2}\left[2\left(V_{GSn} - V_{Tn}\right)V_{DSn} - V_{DSn}^2\right] \tag{1.2.22}$$

and it describes the so called non-saturated current flow. In particular for values of $V_{DSn} \ll (V_{GSn} - V_{Tn})$ the channel act as a linear resistor equal to

$$R_{on} = \frac{1}{\beta_n(V_{GSn} - V_{Tn})}. \tag{1.2.23}$$

This corresponds to the linear region where the current I_{Dn} is proportional to the drain source voltage V_{DSn}.

As the drain-source voltage increases the current goes into the non linear region since the charge close to the drain is reduced by the channel potential. It will reach a point at which the inversion charge at the drain $Q_n(L)$ is approximately zero as can be seen in figure 1.2.9b. This point is called of pinch-off point. The equation 1.2.20 predicts increasing of the current that reaches a peak and for higher V_{DSn} it drops. The decrease in the current is not physical and it corresponds to the case in which the charge of the inversion layer at the drain becomes $Q_n(L) = 0$.

The maximum value of the current I_{Dsat} occurs when

$$\frac{\partial I_{Dn}}{\partial V_{DSn}} = \frac{\beta_n}{2}\left[(V_{GSn} - V_{Tn}) - V_{DSn}\right] = 0. \tag{1.2.24}$$

In this condition the value of V_{DSn} defines the saturation voltage as the increment

$$V_{sat} = V_{GSn} - V_{Tn} \tag{1.2.25}$$

as shown in figure 1.2.8. Equation 1.2.22 is valid only for $V_{DSn} \leq V_{sat}$. The current I_{Dsat} at the saturation voltage can be obtain simply noting that in this condition $V(L) = V_{sat}$ and $Q_n(L) = 0$, which gives

$$I_{Dsat} = I_{Dn}\,|_{V_{DSn}=V_{sat}} = \frac{\beta_n}{2}\left(V_{GSn} - V_{Tn}\right)^2. \tag{1.2.26}$$

Looking the the figure 1.2.8, that shows a family of curves for different values of V_{sat}, the border between the non linear region and the saturation region is approximately parabolic since it can be written

$$I_D\,|_{border} = \frac{\beta_n}{2} V_{sat}^2. \tag{1.2.27}$$

For voltages higher then V_{sat} the simplest approximation is to extend the maximum value of current and in this way keep it constant. In fact in the saturation region the pinch-off point starts to move toward the source keeping the voltage of the pinch-off point at V_{sat}. In practice the number of carriers, and so the current, that arriving at the pinch-off point remain the same. The only thing that change is the channel length that in this case it reduces to L_{sat} as shown in figure 1.2.9c. The expression for I_{Dn} in the saturation region could be modified trying to take into account the phenomena of the so called channel length modulation, which becomes relevant when L_{sat} is a very small part of the

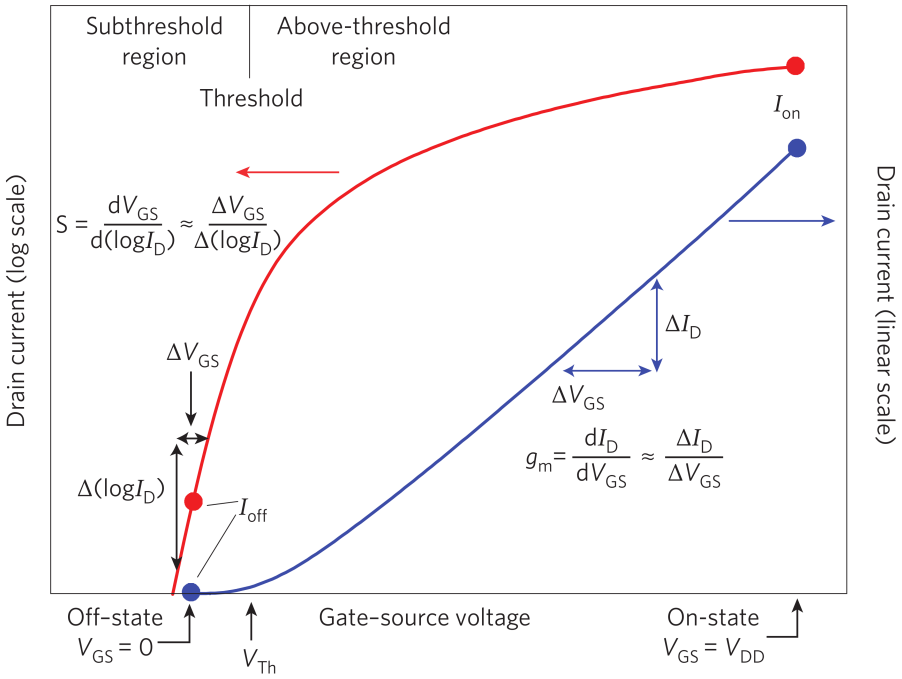


Figure 1.2.10: Drain current in function of gate-source voltage on a logarithmic scale (on the left) and on a linear scale (on the right). Taken from reference[31].

original channel. The value of the current could be adjust by starting from the saturated value and adding a factor, writing in this way

$$I_{Dn} = \frac{\beta_n}{2} (V_{GSn} - V_{Tn})^2 [1 + \lambda (V_{DSn} - V_{sat})] \tag{1.2.28}$$

where λ is the channel-length modulation parameter. The change of the effective length of the channel will slightly increase the drain current with the increment of the drain-source voltage.

Depending on the gate and source-drain voltages the MOSFET can be biased in the sub-threshold region when $V_{GSn} < V_{Tn}$, or into the saturation region when $V_{GSn} > V_{Tn}$. In a linear scale the drain current I_{Dn} seems to approach zero immediately below the threshold voltage. If everything is plotted on logarithmic scale it can be seen that in reality the drain current remains at non negligible values for $V_{GSn} < V_{Tn}$, as shown in figure 1.2.10. This is due to the fact that the inversion charge density Q_n does not drop instantaneously to zero. The sub-threshold region shows how fast the current drops and so it describes how a MOSFET switches off. This behavior is particularly important for low-voltage and low-power applications, such as in digital logic and in memory circuits.

It is useful to define the sub-threshold swing S, which is the inverse of the sub-threshold slope, that quantifies how fastly the transistor is turned off by the

gate voltage V_{GSn}. This parameter is defined as the gate voltage that is needed in order to induce a change in the drain current of one order of magnitude, so it can be calculated as

$$S = \frac{dV_{GSn}}{d\left(logI_{Dn}\right)} = 2.3\frac{kT}{q}\left(1 + \frac{C_d}{C_{ox}}\right) \tag{1.2.29}$$

where C_d is the depletion capacitance which will be introduced in the next section. The minimum value of S can be found by considering $C_{ox} \to \infty$ which leads to 60 mV/dec at room temperature, but in the state of the art MOSFETs is typically 80-100 mV/dec due to parasitic components. In order to have sharp sub-threshold voltage, which means small values of S, it would be excellent to have low channel doping, thin oxide layers low interface traps density and low temperature operation.

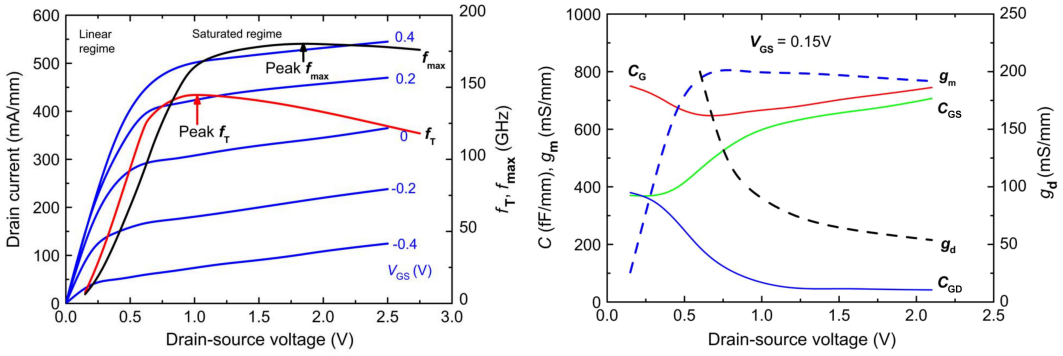
Considering the saturation region, another important figure of merit that can be defined is the transconductance g_m that corresponds to how well the device current responds to a voltage change, as is shown in figure 1.2.10. It can be expressed a

$$g_m = \frac{\partial I_D}{\partial V_{GSn}} \mid_{V_{DSn}=const} = \beta_n \left(V_{GSn} - V_{Tn}\right). \tag{1.2.30}$$

In practice the transconductance g_m represents the sensitivity of the device. It would be better to have high values of g_m so that a small change in V_{GSn} results in a large change of I_{Dn}. Typical values of g_m are between 10 and 30 mS/μm for a 65 nm technology node.

In figure 1.2.10 another important parameter can be extracted: the ON/OFF ratio. It represents the ratio between the ON current and the OFF current of the MOSFET and it gives an idea of the performance of the digital switching of the device. Higher the value the better the switching capability is and in the current state of the art of silicon CMOS digital circuits the ratio $I_{\text{on}}/I_{\text{off}}$ is in between 10^4 and 10^7 [31].

In radio-frequency applications figures of merit are the cut-off frequency f_T and the maximum oscillation frequency f_{MAX}. In this case the transistor is driven in the ON-state and a small radio frequency signal that has to be amplified is superimposed on the gate-source voltage V_{GS} [31, 32]. Small variation of V_{GS} can change the concentration of the carriers in the channel and consequently the drain current of the device. For such operation the device does not need to be turned-off. The cut-off frequency is the frequency at which the magnitude of the small-signal current gain drops to unity (0 dB) while the maximum oscillation frequency corresponds to the case in which the power gain equals unity. The maximum



(a) (b)

Figure 1.2.11: (a) Drain current and cut-off frequency, (b) transconductance, gate capacitance and output conductance in function of the drain-source voltage V_{DS} of a radio-frequency GaAs transistor. Taken from reference [31].

oscillation frequency f_{MAX} represents the upper limit of the transistors action which loses the capability to amplify a signal. In radio-frequency applications both f_T and f_{MAX} should to be maximized. They are given by

$$f_T = \frac{g_m}{2\pi} \frac{1}{(C_{GS} + C_{GD})\left[1 + g_d\left(R_S + R_D\right) + C_{GD} g_m \left(R_S + R_D\right)\right]} \tag{1.2.31}$$

$$f_{MAX} = \frac{g_m}{4\pi C_{GS}} \frac{1}{\sqrt{g_d\left(R_i + R_S + R_G\right) + g_m R_G \frac{C_{GD}}{C_{GS}}}} \tag{1.2.32}$$

where C_{GS} and C_{GD} are the gate-source and gate-drain capacitances and R_i, R_S, R_D and R_G are the internal, source, drain and gate resistances. In order to achieve high values of f_T and f_{MAX} the transconductance g_m has to be maximized and the output conductance g_d expressed as

$$g_d = \frac{\partial I_D}{\partial V_{DS}} \, |_{V_{GS}=const} \tag{1.2.33}$$

minimized. All other capacitances and resistances of the radio-frequency device must also be minimized. The values of all these components depend on the DC value of the drain-source voltage V_{DS} and the gate-source voltage V_{GS}. As can be seen in figures 1.2.11a and 1.2.11b the dependance on V_{DS} is particularly crucial for the FET operation. The peak of the cut-off frequency f_T occurs when the device is driven into saturation regime where g_m is close to its maximum value and g_d, C_{GS} and C_{GD} are as small as possible. For this reason the saturation region of the drain current I_D is important to reach the highest operating speed of the device. This is even more evident looking the peak of f_{MAX} which is moved

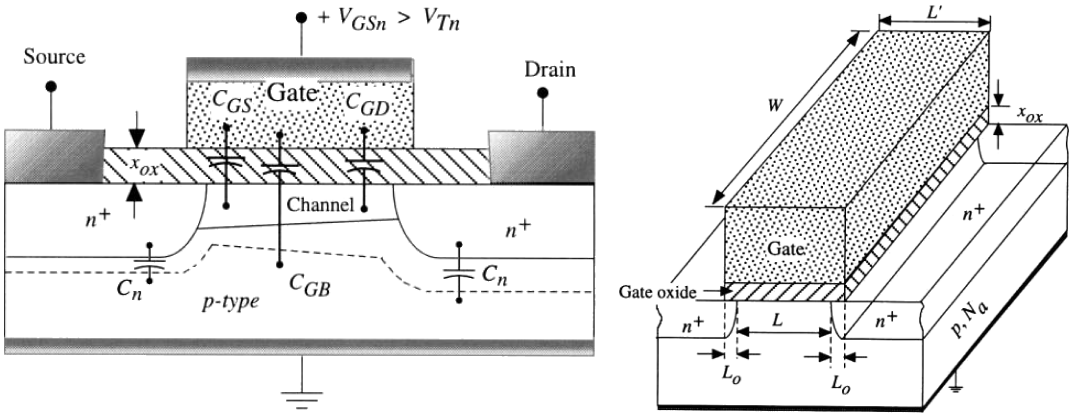


Figure 1.2.12: (a) Parasitic capacitances and (b) prospective view of a MOSFET. Taken from reference [18].

deeper in the saturation where g_d plays an essential role for achieving high power gain and consequently for the maximum oscillation frequency.

1.2.3 MOSFETs modeling

An important aspect of the MOSFET technologies are the parasitic components that exist mainly due to physical structure and operation mode. They must be considered when a device is modeled since these elements are the limiting factors of the circuit performances. In real devices when the current flows from the channel to the terminal contact, there is a small drop in the value of the voltage at the drain and source regions due to finite silicon resistivity and metal contact resistance. As the channel length decreases source and drain parasitic resistances cannot be neglected and can generate current degradation. Two parameters that can affect the performances are the access resistances due to ungated part of the channel and the contact resistance in the region where the current flows into the metal line.

Another key role in MOSFET devices is played by parasitic capacitances since they influence the switching delay of a logic gate. In fact the capacitance determines how fast the gate of a FET can be charged to a certain potential. Looking at figure 1.2.12a the MOSFET's capacitances could be divided into two groups. The first regards the MOS structure, related with the gate oxide capacitance, and is constituted by C_{GB}, C_{GS} and C_{GD}. Instead the second set takes into account the depletion capacitance of the drain and the source C_n.

To define the parasitic capacitances, the gate region of a MOSFET shown

in figure 1.2.12b has to be considered. The value of the oxide capacitance of a transistor is very important since it influences the drain current through the process parameter k_n and is expressed as $C_{ox} = \varepsilon_{ox}/x_{ox}$ with units of F/cm^2. For this type of structure the gate capacitance C_G has to be take into account which is the input capacitance seen looking to the gate and is given by

$$C_G = C_{ox} W L^{'} \tag{1.2.34}$$

where $L^{'} = L + 2L_0$ with L_0 is the gate overlap distance. The gate capacitance can be divided into different contributions in this way

$$C_G = C_g + 2C_{ol}. \tag{1.2.35}$$

In particular $C_g = C_{ox} W L$ is the gate capacitance in which the "effective" channel length L is considered and $C_{ol} = C_{ox} L_0 W$ is the overlap capacitance due to overlap of the gate on both the drain and the source side of the transistor. For a simple estimation it is sufficient to use the expression 1.2.34 of C_G with a reasonable value of C_{ox} to have a quite good estimation.

Now for a better characterization gate-channel capacitances have to be examined and they are represented by C_{GB}, C_{GS} and C_{GD} expressed in farads. They originate from the coupling of the gate electrode with the channel and all the three contributions are non linear since the channel characteristics depend on the applied voltage. A way to define these parasitics is to calculate how the application of a given voltage changes the gate charge Q_G. C_{GS} and C_{GD} can be written as

$$\begin{aligned} C_{GS} &= -\left(\frac{\partial Q_G}{\partial V_S}\right) \\ C_{GD} &= -\left(\frac{\partial Q_G}{\partial V_D}\right) \end{aligned} \tag{1.2.36}$$

where V_S and V_D are the source and drain potential measured with respect to the gate voltage. C_{GB} instead could be seen as composed by the series of the gate oxide capacitance and the depletion capacitance C_d

$$C_{GB} = WL \left(\frac{1}{C_{ox}} + \frac{1}{C_d}\right)^{-1}. \tag{1.2.37}$$

The behavior of the three capacitors is shown in figure 1.2.13. In the saturation and non saturation region C_{GB} could be neglected since the inversion layer acts as a shield and the charge is supplied by the source and the drain. What has to be underlined is that all the MOS capacitances increase with the channel width W and this is not perfect from the physical point of view, but they are sufficient

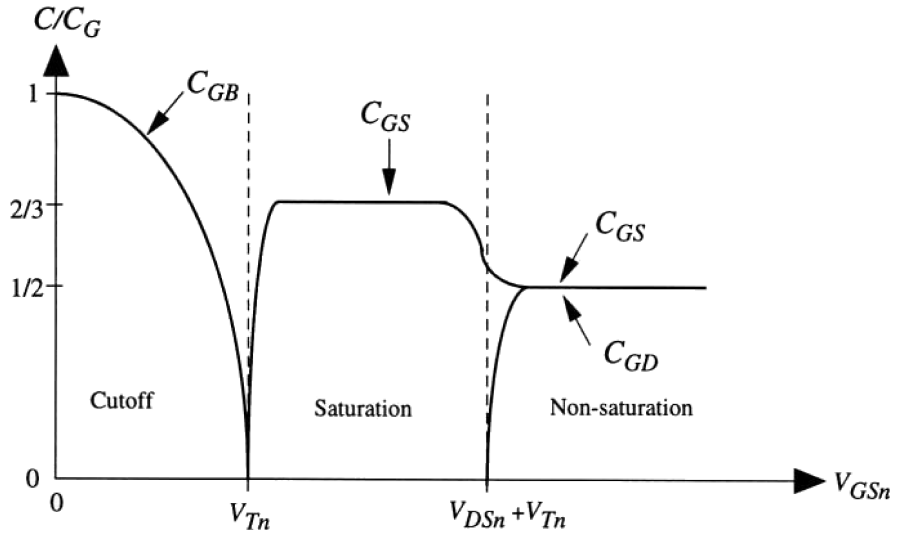


Figure 1.2.13: Gate-channel capacitances normalized to the value of C_G as a function of the operational regions of the MOSFET. Taken from reference [18].

for a first estimation for the design of the transistors.

Another important contribution to the limitation of the switching speed of the device is the depletion capacitance C_d, in units of F/cm^2, which also has a non linear behavior. C_d originates from the ionized dopants in the proximity of a p-n junction as is illustrated in figure 1.2.14a. If a step doping profile is considered with doping densities N_a and N_d for the p-region and the n-region respectively, the zero-bias capacitance can be defined as

$$C_{d0} = \frac{\varepsilon_{Si}}{x_{d0}} \tag{1.2.38}$$

where

$$x_{d0} = \sqrt{\frac{2\varepsilon_{Si}\phi_0}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)} \tag{1.2.39}$$

is the zero-bias depletion width, with ϕ_0 defined as the built-in voltage. The value of x_{d0} became larger increasing the applied reverse-bias voltage V_R. This means that the depletion capacitance can be expressed in function of V_R as

$$C_d\left(V_R\right) = \frac{C_{d0}}{\sqrt{1 + \frac{V_R}{\phi_0}}} \tag{1.2.40}$$

which gives the behavior shown in figure 1.2.14b.

In a MOSFET structure a great contribution of the parasitic capacitances comes from the depletion capacitance of the source and the drain regions C_n. In

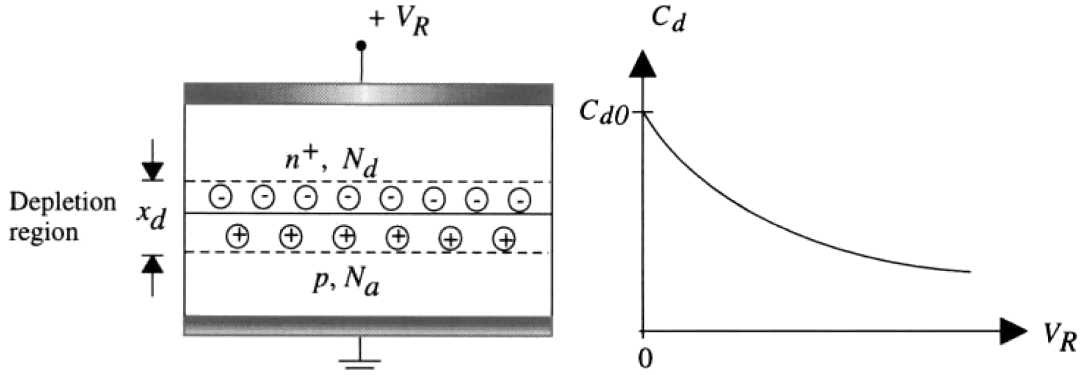


Figure 1.2.14: (a) depletion charges in a p-n junction, (b) behavior of the depletion capacitance. Taken from reference [18].

order to describe them it is necessary to consider the typical n-channel MOSFET region represented in figure 1.2.15. Source and drain regions have three dimensional geometry, so to simplify the calculation they can be split into bottom and sidewall part and the total capacitance will be obtained adding the two components. Starting from the bottom capacitance it can be easily written as

$$C_{bot} = C_{d0} W X \tag{1.2.41}$$

where W is the width and X the lateral extension of the source or drain region. The sidewall capacitance instead is calculated taking into account that the sidewall has a depth x_j, which brings the expression

$$C_{side} = C_{d0sw} x_j 2 \left(W + X\right). \tag{1.2.42}$$

So the total zero-bias depletion capacitance of the source or drain region is given by

$$C_n = C_{bot} + C_{side} = C_{d0} W X + C_{d0sw} 2 \left(W + X\right). \tag{1.2.43}$$

The value of C_n is used for a first approximation of the performance of the MOSFET device. In fact the depletion capacitance is non linear and, since C_n decreases when the reverse-bias voltage is increased, the zero-bias depletion capacitance overestimates the contribution in a digital circuit analysis. All these contributions will simply give an estimation of the total capacitance at every node in digital CMOS circuits and will give an idea about the circuit performance.

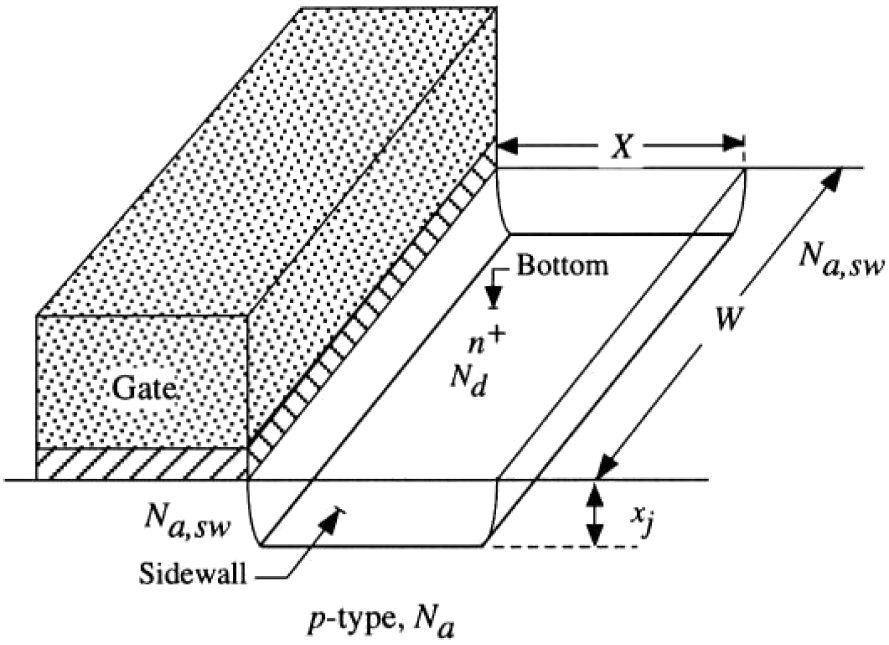


Figure 1.2.15: MOSFET junction geometry for the calculation of the source and drain depletion capacitances. Taken from reference [18].

1.2.4 Scaling theory

The MOSFET model discussed so far is valid only for devices with channel length of the order of 20 μm. In fact for smaller transistors the physics change and affect the circuit operation. In this situation the scaling theory helps to understand how the device characteristics are affected by the reduction of the device dimensions. The scaling theory can be applied both to surface (lateral scaling) and vertical (vertical scaling) features of a MOSFET so that, ideally, the long-channel behavior of the internal electric field is preserved. With this approach it is useful to introduce the scaling factor S $>$ 1, that is used to shrunk the dimensions of the long-channel transistors. In the case of the lateral scaling the channel width W, the channel length L and the channel area $A = WL$ for a scaled device becomes

$$\begin{aligned} W' &= \tfrac{W}{S}, \\ L' &= \tfrac{L}{S} \\ A' &= \tfrac{A}{S^2} \end{aligned} \tag{1.2.44}$$

as shown in figure 1.2.16. As we saw in section 1.2.2 the current in the MOSFET is proportional to the aspect ratio (W/L) which is invariant for a transistor with reduced lateral dimensions. This type of scaling results only in the reduction of the area of the channel.

The MOSFET is an electric field-effect device, so performing only lateral

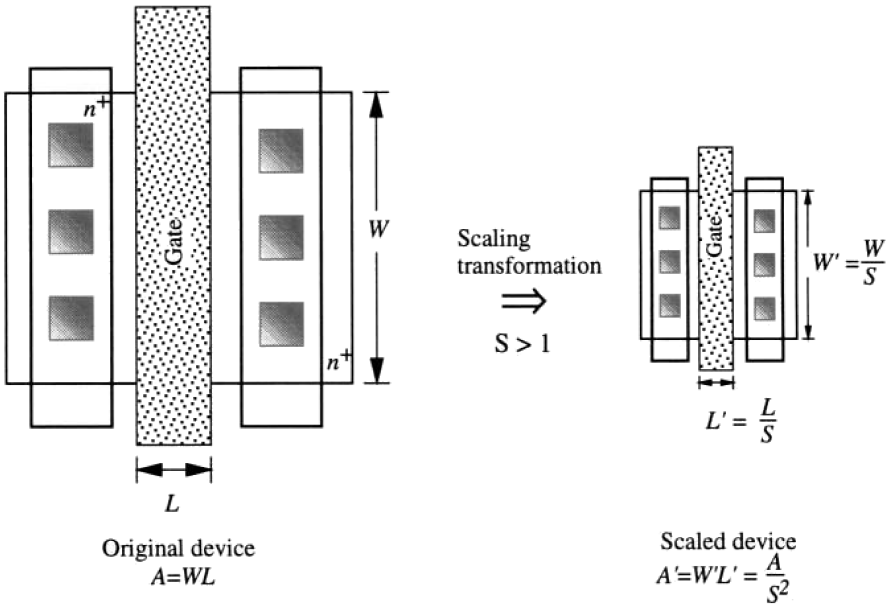


Figure 1.2.16: MOSFET lateral scaling. Taken from reference [18].

scaling the electric field distribution in the channel will be modified and this will change the current-voltage characteristics. In order to restore the original field distribution also vertical dimensions must be reduced as it can be seen in figure 1.2.17a. The most important parameters that have to be scaled are the oxide thickness x_{ox} and the junction depth x_j, which are reduced by the scaling factor S as is illustrated in figure 1.2.17b. These two dimensions have an impact on the device I-V characteristics through the values of the process parameter k and the device parameter β. In particular larger values of k and β are desirable and this can be done simply by increasing the oxide capacitance C_{ox} i.e., by reducing the oxide thickness x_{ox}. In fact for the scaled devices it can be written

$$\begin{aligned} x'_{ox} &= \tfrac{x_{ox}}{S} \\ C'_{ox} &= \tfrac{\varepsilon_{ox}}{x'_{ox}} = SC_{ox} \end{aligned} \tag{1.2.45}$$

which will increase the process and the device parameters to

$$\begin{aligned} k' &= Sk \\ \beta' &= S\beta \end{aligned} \tag{1.2.46}$$

since the aspect ratio is invariant. It has to be considered that the field-effect in a MOSFET is related to the doping density N_a and N_d. This means that these two parameters must be modified as well to new values N'_a and N'_d in order to preserve the same operational characteristics. It has to be underlined that chan-

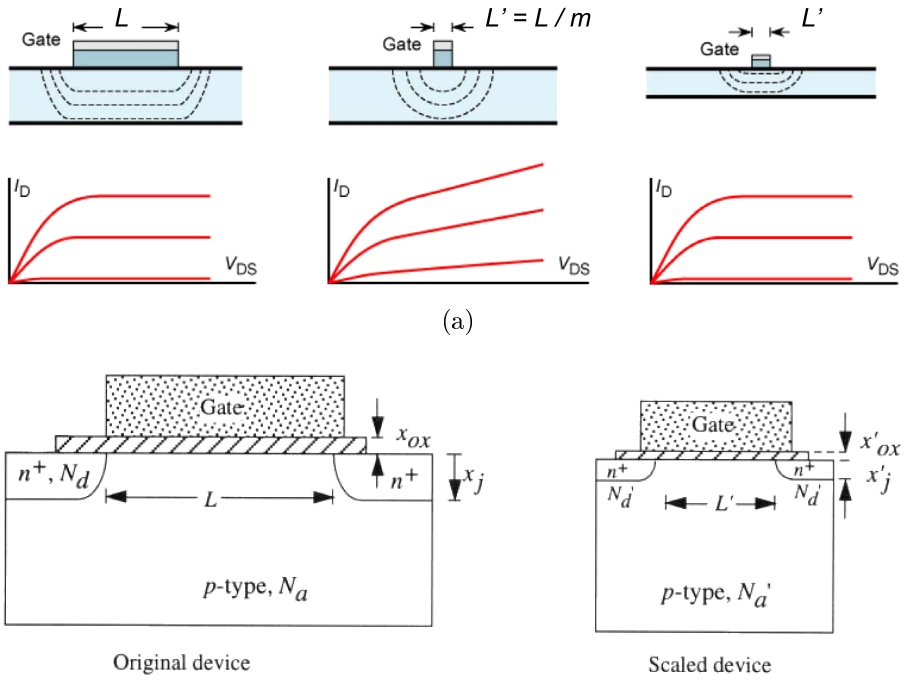


Figure 1.2.17: (a) Electric field distribution in scaled devices, (b) MOSFET vertical scaling. Taken from reference [18].

nel doping cannot be increased indefinitely, otherwise it will lead to p-n junction breakdown. The gate oxide thickness, for short CMOS channel lengths, has approached the low nanometer scale. In this range the gate tunneling current could be too high in order to have an operating device. To solve this problem high permittivity dielectrics must be used as gate insulators, as aluminum oxide, hafnium oxide and yttrium oxide. In fact they can have a thicker physical thickness for the same capacitance, reducing in this way the gate tunneling current.

The reduction of all these parameters will influence the square-law equations of a MOSFET. In particular two type of scaling should be considered: full-voltage scaling and constant-voltage scaling. In both cases the analysis starts by taking into account the equations 1.2.22 and 1.2.26 for the non-saturated I_D and saturated I_{Dsat} currents respectively for the original device. In the full-voltage scaling the scaling factor S is also used to reduce the applied voltages V_{DS} and V_{GS} to

$$\begin{aligned} V'_{DS} &= \frac{V_{DS}}{S} \\ V'_{GS} &= \frac{V_{GS}}{S}. \end{aligned} \tag{1.2.47}$$

Assuming to adjust the ion implantation so that even the threshold voltage can

be scaled to $V'_T = V_T/S$, the currents I'_D and I'_{Dsat} of the scaled device become

$$\begin{aligned} I'_D &= \frac{I_D}{S} \\ I'_{Dsat} &= \frac{I_{Dsat}}{S}. \end{aligned} \tag{1.2.48}$$

The interesting aspect of this approach is the power dissipation of the transistor. In fact the power of the scaled device turns into

$$P' = I'_D V'_{DS} = \frac{I_D}{S}\frac{V_{DS}}{S} = \frac{P}{S^2}. \tag{1.2.49}$$

So in this way the area and the power dissipation of the scaled device are reduced by the same factor, keeping constant the power dissipation per unit area.

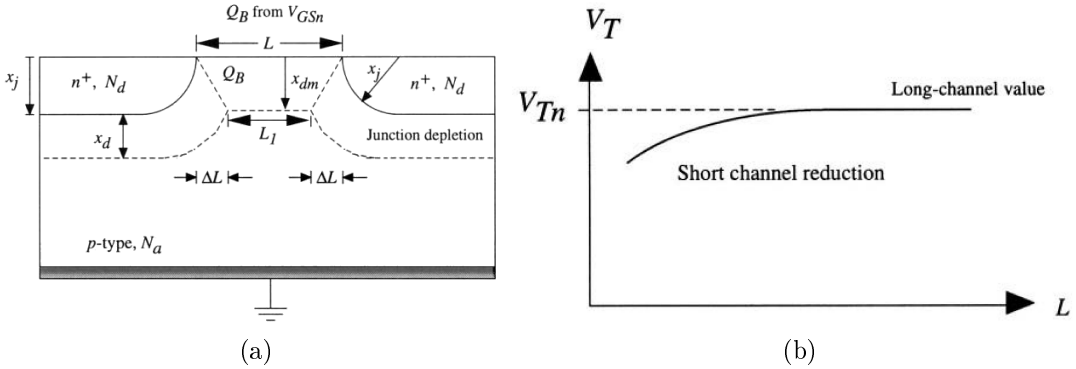
In general is not possible to change arbitrarily the power supply. So the constant-voltage scaling takes into account how the I-V characteristics are affected when the dimension of the transistors are scaled while the applied voltages are kept constant. Proceeding as before, I'_D, I'_{Dsat} and P' can be written as

$$\begin{aligned} I'_D &= SI_D \\ I'_{Dsat} &= SI_{Dsat} \\ P' &= SP \end{aligned} \tag{1.2.50}$$

showing that with this approach there is a problem related to the power dissipation, which is increased by a factor of S.

1.2.5 Short-channel effects and narrow-width effects

Even if the scaling theory is applied correctly, as the dimension of the channel are reduced at about 2 μm there are deviations from the long-channel behavior. Short-channel effects arise from the fact that it is not possible anymore to use the gradual-channel approximation. In fact for small transistors the potential distribution becomes two-dimensional, since it depends both on the transverse field E_x and longitudinal field E_y which are controlled respectively by the gate voltage and the drain bias. This behavior can cause mainly four type of short-channel effects: the threshold voltage V_T is non constant with L, the drain current I_D cannot saturate with drain voltage V_D, I_D is not anymore proportional to $1/L$ and the transistor characteristics degrade with time. The most important results for small transistors is that the threshold voltage is reduced below the value found in section 1.2.1. In case of long-channel devices the gate voltage V_{GS} supports the bulk charge Q_B on the entire area of the channel (WL) underneath the gate. This



(a) (b)

Figure 1.2.18: (a) Bulk charge distribution in short-channel MOSFET, (b) short-channel effect on threshold voltage. Taken from reference [18].

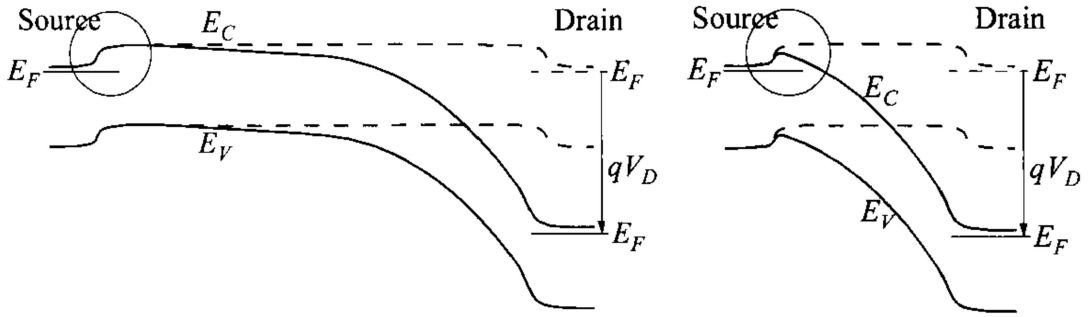
is not valid for a short-channel since there are some depletion charges induced by the drain-bulk and source-bulk pn junctions which cannot be ignored. For this reason, if only the bulk charge is considered, the threshold voltage will be overestimated.

In order to calculate the reduction of the threshold voltage it is useful to take into account the bulk charge distribution for the small-geometry transistors shown in figure 1.2.18a. The gate voltage that supports the bulk charge Q_B has a trapezoidal shape looking at the cross-section. In particular the upper part of the trapezoid has a length L, while the bottom one has a length $L_l = L - 2\,(\Delta L)$ where $\triangle L$ is considered to be the same on both sides. Assuming that the depletion thickness x_d of the pn junction corresponds to the maximum MOS depletion depth, from the simple geometry of figure 1.2.18a the change in the threshold voltage can be calculated. So including the short-channel effect the threshold voltage for an n-type MOSFET can be written as

$$(V_{Tn})_{SCE} = V_{Tn} + (\triangle V_{Tn})_{SCE} \tag{1.2.51}$$

where V_{Tn} is the long-channel value. From the calculation it turns out that $(\triangle V_{Tn})_{SCE}$, which depends on the channel length L, has a negative value and this explain the reduction of the threshold voltage from the long-channel value. In figure 1.2.18b is illustrated the general trend of the threshold voltage $(V_{Tn})_{SCE}$ in function of the channel length L.

Short-channel effects start to be important when the depletion regions of the source an the drain are a relevant fraction of the channel length. If in extreme case the widths of the depletion regions approach the channel length, the device will be influenced more seriously. Usually this phenomenon is called of punch-through



(a) (b)

Figure 1.2.19: Energy band diagram at the surface of the semiconductor for (a) long-channel MOSFET and (b) short-channel MOSFET, showing the DIBL effect. The dashed line represent the case for $V_D = 0$, while for the solid line $V_D > 0$. Taken from reference [35].

and it consists in a large leakage current between the source and the drain. The origin of the punch-through arises from the lowering of the barrier close to the source and is called drain-induced barrier lowering (DIBL) used in figures 1.2.19a and 1.2.19b. In practice the drain bias can influence the barrier at the source if the channel is too short, so that the carrier concentration in the channel at the source is no longer fixed. This can be shown looking at the energy band diagram at the semiconductor surface. In the case of a long-channel MOSFET, the drain voltage can change the effective channel length but the barrier at the source remains the same as it can be seen in figure 1.2.19a. This is not valid for a short-channel transistor since the lowering of the source barrier gives origin to an injection of extra carriers as shown in figure 1.2.19b. In punch-through condition this effect is present both in the above threshold and sub-threshold regimes.

As the channel becomes shorter and shorter the internal electric field increases. This has an effect on the value of mobility since $\mu = v/E$, where v is the drift velocity of the carriers. For low electric field the device is in the linear region so the velocity is proportional to the electric field and the mobility can be assumed as constant. In devices with channels length smaller then 1 μm the electric field can exceed 10^3 V/cm leading to the non linear region and it is not possible anymore to use a constant value for the mobility of electrons and holes. As the strength of the field is increased the drift velocity approaches its maximum level v_s which is called saturation velocity as it can be seen in figure 1.2.20. This value is the maximum velocity that a charge carrier can reach inside the lattice. This limitation is due to the fact that, even if the electric field tries to increase the energy, collisions tend to remove energy from the carrier. In the case of silicon

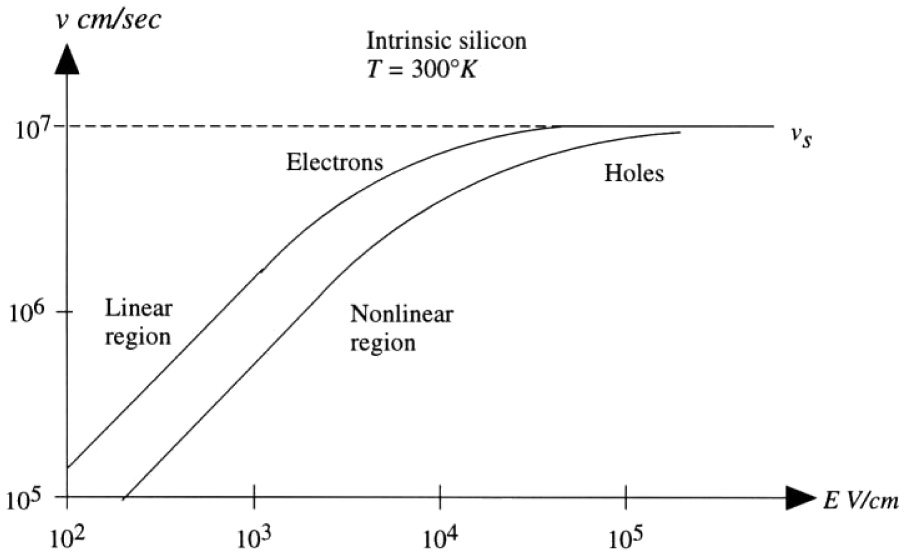


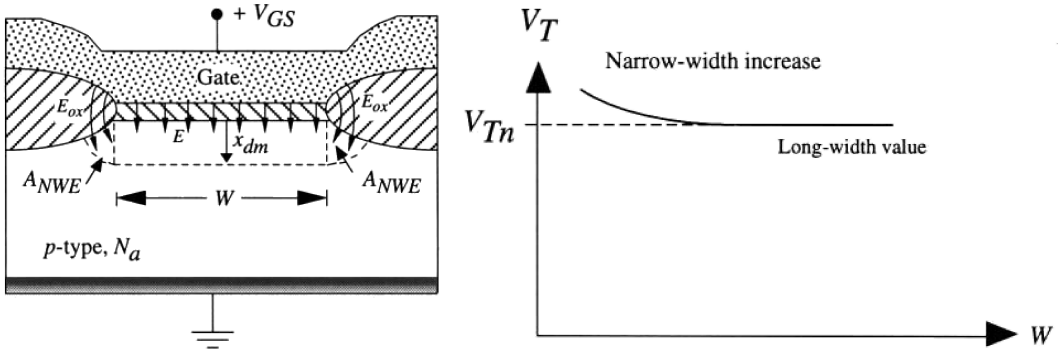
Figure 1.2.20: Velocity-Electric field relations in bulk silicon. Taken from reference [18].

at room temperature $v_s = 10^7$ cm/s. This phenomena will also have an effect on the current of the MOSFET: as the electric field will become higher, the current will increase together with the drift velocity. When the velocity saturates also the current reaches a constant value. It has to be noticed that this kind of mechanism is completely different from the case of constant mobility. In fact the constant value of the current is not due to the pinch-off point, but it is due to the saturation velocity of the carriers. So the saturation current can be expressed as

$$I_{Dsat} = WC_{ox}v_s\left(V_{GS} - V_T\right)$$

when $V_{DS} \geq V_{sat}$.

The reduction of the channel length is not the only thing that can influence the value of the threshold voltage. In fact decreasing the channel width W leads to narrow width effects. In particular the threshold voltage, due to the shrinking of W, exhibits larger value compared to the one found using the gradual-channel approximation. This behavior can be understood by looking at figure 1.2.21a. What happen is that the fringing electric fields deplete the silicon on an area which is greater than the one defined by the channel width W. Since this component of the bulk charge is not considered in a large-width devices, the threshold voltage V_T is underestimated in the case of narrow-width transistors. Considering the geometry of the cross-section of the MOSFET shown in figure 1.2.21a the reduction of V_T can be calculated. The narrow-width effect correction arise from the presence of the charges due to the extra area A_{NWE} estimated by assuming



(a) (b)

Figure 1.2.21: (a) Bulk charge distribution in narrow-width MOSFET, (b) narrow-width effect on threshold voltage. Taken from reference [18].

that the depletion edges have a circular shape. In this way the threshold voltage can be written as

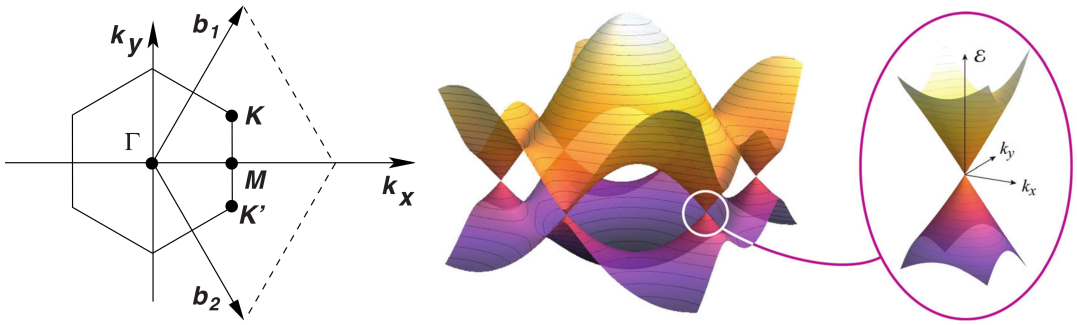
$$(V_{Tn})_{NWE} = V_{Tn} + (\triangle V_{Tn})_{NWE} \tag{1.2.52}$$

with $(\triangle V_{Tn})_{NWE} > 0$ and its behavior due to the shrinking of the channel width W can be seen in figure 1.2.21b.

During all these years it was possible to scale silicon MOSFETs following the Moore's law. Today this technology allows the fabrication of transistors with dimension of the gate length of 22 nm. The international roadmap for semiconductor (ITRS) predicted the need for devices with gate length of 10 nm by 2020. However it is complicated to further scale Si MOSFETs keeping the needed performances in the presence of short-channel effects. For this reason it will be necessary in the future to introduce devices based on different physics or on materials other than silicon in order to continue to improve the MOSFETs performances.

1.3 Graphene field-effect transistors

In the search of new materials that could represent an option for the post-silicon electronics some material properties should be considered. The ideal semiconductor should have a wide band-gap, high value of mobility, be producible on large scale, compatible with Si CMOS technology, good interface with dielectrics with a long term stability and low contact resistance [32]. Graphene represents a good trade off to these requirements. A single layer of graphene is a two-dimensional material that is composed of carbon atoms arranged in a hexagonal lattice. It was isolated for the first time in 2004 [28] with the technique of the micro-mechanical



(a) (b)

Figure 1.3.1: First Brillouin zone of graphene lattice [5] and band structure of graphene with the linear dispersion around the K points. Taken from reference [6].

exfoliation, but today it can be grown on metals with the chemical vapor deposition (CVD) process in order to produce it on wafer-scale. Graphene is not a semiconductor like silicon, but is a semi-metal with a zero band-gap. Differently from Si MOSFETs, which are unipolar, the charge carriers in graphene channel can be changed from electron to holes simply applying a gate potential. The minimum carrier density point is called the Dirac point. Looking at the band structure in figure 1.3.1, the valence and conduction bands meet at the K points of the Brillouin zone and, differently from semiconductors where the bands have a parabolic-shape, they have linear energy-momentum dispersion close to the Dirac points. In this case electrons, propagating through graphene lattice, mimic relativistic particle. For that reason particle should be described by Dirac-like equations rather then Schrodinger equations [11]. Because of the absence of a band-gap, transistors made with graphene cannot be turned-off and their ON/OFF ratios are poor ($<$100), which in theory makes such devices unsuitable for most logic applications. The impossibility of switching devices off also results in large static power consumption, absence of which is one of the great advantages of the Si CMOS. It was demonstrated that it is possible to modify the band structure of graphene in order to open a band-gap. This can be done by confining graphene in one dimension by forming nanoribbons, by biasing bilayer graphene or by applying some strain to the monolayer [31].

Despite the fact that graphene does not have a band-gap, it offers many advantages in FET channels. The biggest one is its high carrier mobility at room temperature. In fact on exfoliated graphene usually mobilities of 10000-15000 cm^2/Vs are measured, reaching values of 10^6 cm^2/Vs in case of suspended graphene devices. For large area-graphene grown with CVD process and trans-

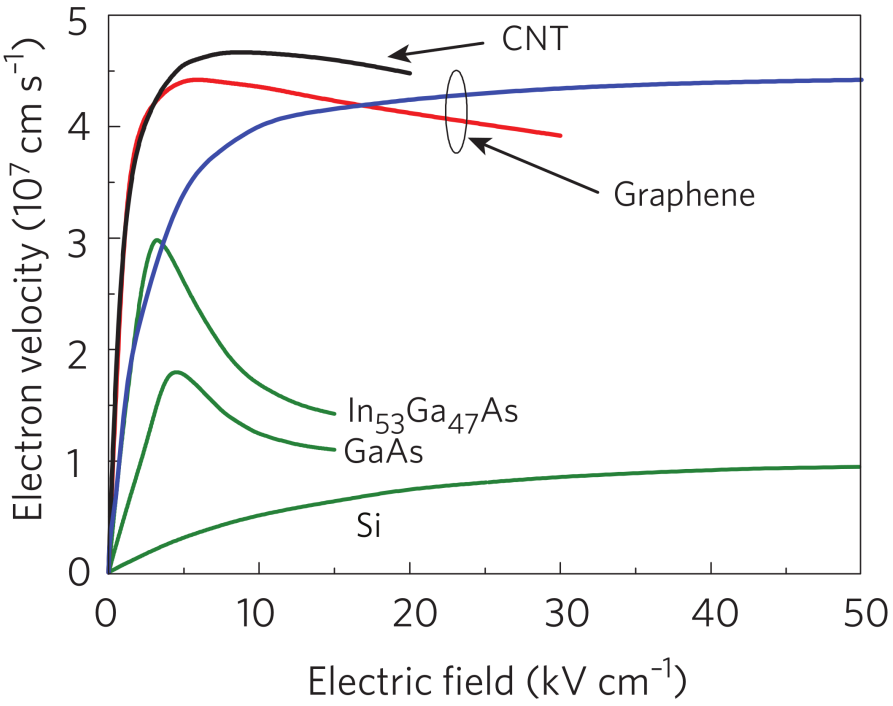


Figure 1.3.2: Electron drift velocity-electric field for standard semiconductor, carbon nanotubes (simulation) and large-area graphene (simulation). Taken from reference [31].

ferred on SiO_2 substrates, mobility up to 4000 cm^2/Vs has been measured [31]. Even if these values are lower compared to the exfoliated graphene they still are about 4 times larger than the mobility of bulk silicon. Such high values of mobility are useful for the fabrication of transistors for very high frequency applications. In conventional semiconductors the hole mobility is much lower compared to the electron mobility. In fact for silicon the ratio μ_h/μ_e is approximately 0.3 and even lower for III-V semiconductors, while for graphene that difference is not so large and the hole mobility can be even higher than that of electrons[32].

The deposition of oxides on the top of graphene for the realization of top-gated transistors seems to be a problematic issue since it was leading to the degradation of the mobility of the devices. During these years several steps were made in that direction until it was possible to obtain mobilities of 24000 cm^2/Vs in top-gated graphene FETs [31]. This demonstrates that it is possible to fabricate good devices with dielectrics with a long term stability simply by making the proper choice of the materials and optimizing the deposition process.

In the state-of-the-art FETs, where devices have short-channels, high electric fields reduce the importance of the mobility. In fact in this situation the saturation velocity of the carriers becomes an important figure of merit, as it was

explained in section 1.2.5. Figure 1.3.2 shows the graph of the electron drift velocity versus the electric field for common semiconductors and simulated plots for large-area graphene and carbon nanotubes (CNT) [31]. From the simulation in the case of graphene it can be seen that for high electric fields the velocity does not drop as fast as in the case of III-V semiconductors and moreover recent experiments shown values of saturation velocities of $3 - 6 \cdot 10^7$ cm/s (close to Fermi velocity of $\sim 10^8$ cm/s) for large-area graphene [32]. Considering that the velocity saturation for silicon MOSFETs is 10^7 cm/s it can be concluded that graphene has some advantages over the conventional semiconductors.

Another advantage of graphene consists in the fact that graphene is a 2D-material and it could allow to fabricate devices with extremely thin channels that can result in better gate control of the channel. Even though graphene is so thin it can provide carrier sheet densities of 10^{12} cm^{-2}, similar the the conventional transistors, allowing in this way the appropriate MOSFET operation [32]. Moreover an atomically thin channel can be scaled to extremely short channel lengths without experiencing short-channel effects, allowing graphene to be more scalable compared to standard FETs [31]. The two dimensional nature of graphene can in addition enable monolithic 3D integration.

For the operation of transistors another key issue is the contact resistance between the source and drain metal and the underneath graphene layer. Typically it is given in the normalized form in units of resistance times the contact width, $\Omega \cdot \mu$m, and its value should be as low as possible. Contact resistances of first graphene field-effect transistors (GFETs) were quite high compared to Si MOSFETs, but in the meanwhile significant progress has been made reducing its value to 100-200 $\Omega \cdot \mu$m, about 5 times the contact resistance in Si [10, 16, 22, 27, 38, 45].

Traditional semiconductors usually are rigid materials and cannot be used for emerging technologies such as flexible and printable electronics [32]. In that fields organic semiconductors are used, but those materials have low mobility. Large area graphene is bendable and has a significantly higher mobility compared to organic semiconductors such as pentacene. These characteristics allowed graphene to be deposited on flexible substrates and used in inks for printable electronics.

All these properties can allow graphene to be the material of the future together with the fact that it is fully compatible with existing CMOS technologies and gives the possibility for hybrids with silicon for high performance and flexible devices. However, due to the absence of a band-gap and small values of the ON/OFF ratio, GFETs so far have been used mainly in niche application such as frequency doublers [41], radio-frequency mixers [40] and phase modulators [15].

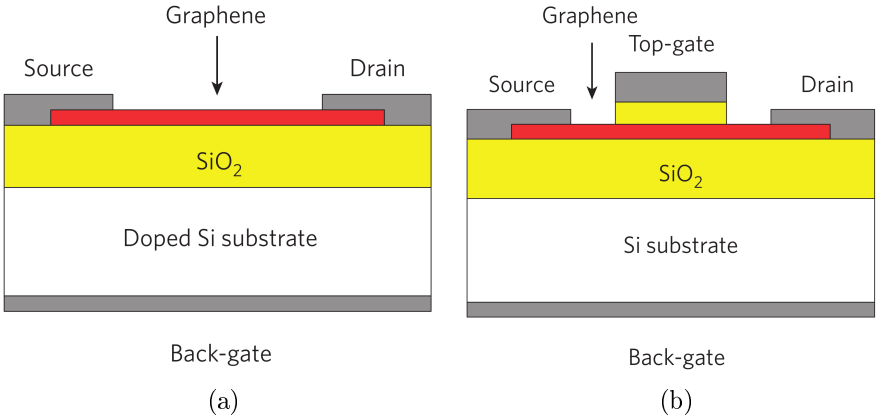
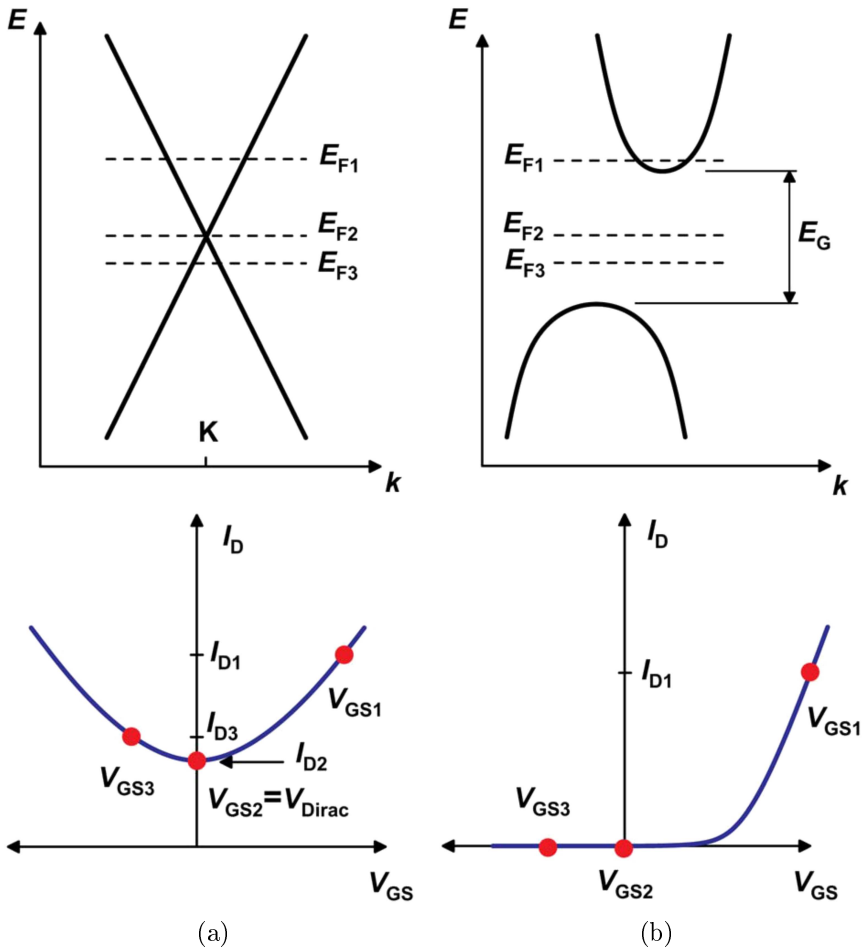


Figure 1.3.3: Different type of graphene field-effect transistors with (a) back-gate structure and with (b) the combination of top-gate and back-gate. Taken from reference [31].

1.3.1 Transfer characteristics

The first graphene field-effect transistor (GFET) was reported in 2004 [28]. Differently from Si MOSFETs, that are usually top-gated, this type of graphene device had a layer of 300 nm of SiO_2 underneath the channel, that was used as back-gate dielectric, and a back-gate made by doped silicon substrate, as illustrated in figure 1.3.3a. The main reason to use that configuration was that it was necessary to have a good optical visibility for the graphene flakes and this can be obtained only for specific thicknesses of the oxide [4]. Back-gated transistors where useful to prove the FET operation when graphene is used as the channel in a MOSFET structure, but they suffered from huge parasitic capacitances. Moreover with such a thick layer of oxide it is hard to have a good control over the channel and such GFETs cannot be integrated with other components. An important step for the improvement of the performances of GFETs was made in 2007 when the first graphene top-gated device was made [19]. During these years top-gated transistors, like the one shown in figure 1.3.3b, were fabricated using both exfoliated graphene and graphene grown on metals (such as copper and nickel) for the channel and growing SiO_2, Al_2O_3 and HfO_2 as top-gate dielectric. All these and the use of top-gate in combination with the back-gate allowed to GFETs to compete with Si MOSFETs.

GFETs have a unique transfer characteristic due to the fact that the conduction and valence bands are not separated by a band-gap like in conventional semiconductors. Like in Si MOSFETs the carrier density in the channel is influenced by the potential difference between the channel and the gates, but in



(a) (b)

Figure 1.3.4: Band diagram and transfer characteristics of (a) gap-less graphene and (b) n-channel Si MOSFET. Taken from refence [32].

graphene this also can change the type of carriers and due to the missing gap GFETs cannot be switched off, as it can be seen in figure 1.3.4a [32].

When a positive gate voltage V_{GS1} is applied, the Fermi level tends to move to the position E_{F1} inside the conduction band promoting an electron accumulation in the channel that corresponds in the transfer characteristics to the value I_{D1} of the drain current. As the bias applied to the gate is reduced, the Fermi level is shifted downward and decreases the concentration of the electrons causing the reduction of the drain current.

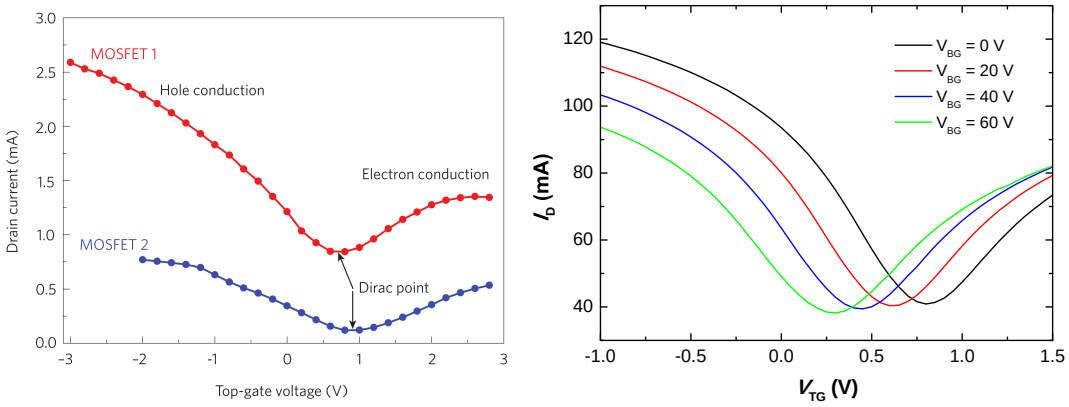
In case that the gate voltage is equal to $V_{GS2} = V_{\text{Dirac}}$ the Fermi level E_{F2} is located exactly where the conduction and valence bands meet. This corresponds to the case of minimum carrier density and drain current I_{D2}. This point is called Dirac or charge neutrality point. Here in theory the density of the electronic states should vanish but there is still a residue of conductivity due to the presence of electron and hole puddles, caused by a small overlap between the conduction

and valence bands, and for that reason graphene undergoes ambipolar transition without the depletion of the carriers [25]. In the case of undoped graphene, as in figure 1.3.4a, $V_{\text{Dirac}} = 0$ and this also represents the flat-band voltage of the device when the drain current is at its minimum. As soon as the gate voltage assume negative values the type of carriers change from electrons to holes. So, for the case of figure 1.3.4a, for the gate voltage V_{GS3} the Fermi level moves inside the valence band and the carrier concentration increase again and rise the drain current to the value I_{D3}. So reducing the gate voltage V_{GS} the graphene channel can change from n-type to p-type MOSFET and this phenomenon is called ambipolar conduction. This behavior gives to graphene the peculiar transfer characteristic of two branches separated by the Dirac point.

For conventional semiconductors as silicon the situation is completely different as it can be seen in figure 1.3.4b. Start with considering an n-type channel where a positive bias V_{GS1} is applied, the Fermi level moves inside the conduction band and there is an electron accumulation in the channel associated to the drain current I_{D1}, like in the case of graphene. But if the gate voltage is reduced to lower values such as V_{GS2} or V_{GS3} the Fermi level goes rapidly into the gap and switches off the transistor.

Ideally the Dirac point of graphene should be located at $V_{GS} = 0$, but this is not always true. In fact the position of the charge neutrality point depends on several factors. This shift could be due to the difference of the work functions between the gate and the graphene, the density and type of charges at the top and bottom interface of the channel and due to doping of graphene introduced by adsorbed ambient impurities. Figure 1.3.5a shows the typical behavior of two GFETs that have the Dirac points at positive voltages, which stems for p-doping. As it can be seen the transfer characteristics are not fully symmetrical, in particular the electron mobility is lower then the hole mobility, and the ON/OFF ratios are quite low. The asymmetry is caused by a geometry dependance due to the combination of the formation of p-n junctions between the channel and the source and drain regions together with the contact resistances that depend on the gate voltage [29]. Typically the values of ON/OFF ratio of transistors fabricated with large-area graphene are around 2-20, while for devices realized with exfoliated graphene, due to a higher crystallographic quality, it can reach values of 100 at room temperature [44].

Due to graphene transistor properties it is useful to use MOSFETs structure like the one in figure 1.3.3b in order to take advantage of the combination of two different gates. In fact due to its ambipolar conduction it is possible to

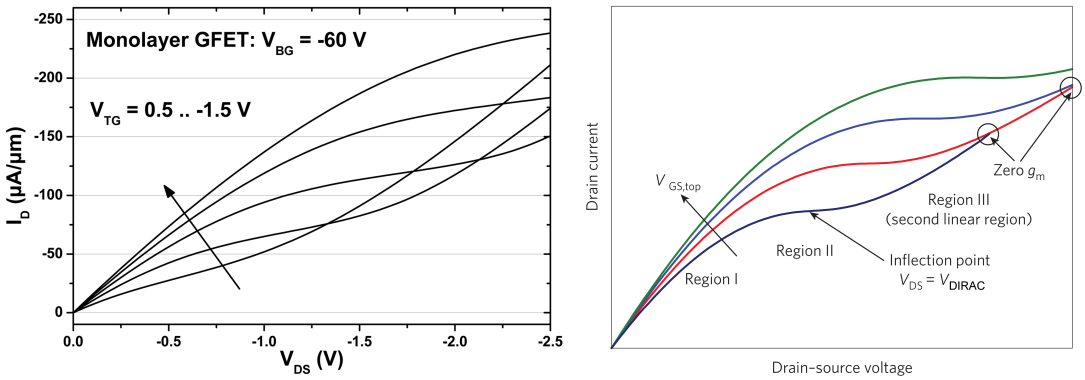


(a) (b)

Figure 1.3.5: (a) Typical transfer characteristic for two different MOSFETs with large-area graphene used as channel, taken from refence [31], (b) transfer characteristic of a GFET at different values of the back-gate voltage.

use the back gate to dope electrostatically the graphene channel. As shown in figure 1.3.5b, applying different constant voltages to the back-gate shifts the Dirac point in the transfer characteristics. In this way it is possible first of all to choose between n-type or p-type channel and moreover to have the control on the threshold voltage of the top-gated channel [26].

An important characteristic for the performances of a MOSFET is the transconductance as explained in section 1.2.2, that describe how well the device current responds to a voltage change. It was found experimentally in top-gated GFETs a transconductance $g_m \approx 150\ \mu S/\mu m$ [26]. Extrapolating the effects of the series resistance and considering $v_{sat} = 5.5 \cdot 10^7$ cm/s in a single layer graphene, the transistor intrinsic transconductance becomes equal to 833 $\mu S/\mu m$. This value should be compared to the transconductance of a Si MOSFET with a 65 nm channel length with a gate capacitance of 1.77 $\mu F/cm^2$ that has $g_m \sim 1.5\ \mu S/\mu m$, which results to be lower then in the GFET. The value of the transconductance of that particular graphene device could be even higher if it would have the same gate capacitance as in the Si MOS structure. For that transistor a dielectric layer HfO_2 that has a gate capacitance of 0.762 $\mu F/cm^2$ was used. So using a higher gate capacitance, similar to the one used in standard silicon technology, the GFET would have $g_m > 2.9\ \mu S/\mu m$ which is definitely higher then the values of the transconductances of the state-of-the-art silicon MOSFETs.



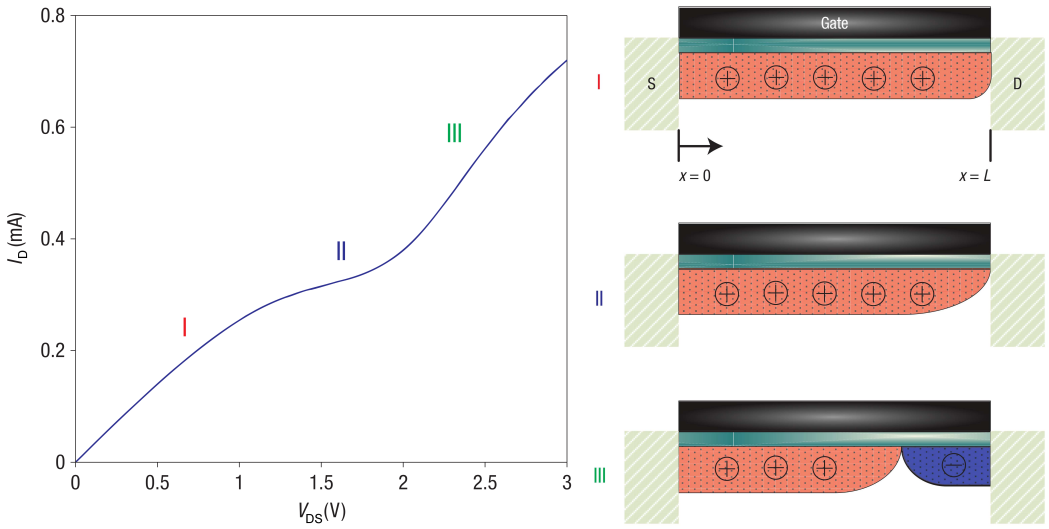
(a) (b)

Figure 1.3.6: Output characteristics in GFETs (a) with linear-shape or weak saturation [34], (b) with saturation and second linear region. Taken from reference [32].

1.3.2 Current-voltage characteristics

Output characteristics of an ambipolar GFET differ significantly from the unipolar Si MOSFETs. The charges that contribute to the drain current in a Si MOSFET are always electrons (n-channel) or always holes (p-channel) inside all the regions of the operation of the device. In case of graphene instead, since it is a zero-gap material, the conduction is due to electrons when the Fermi level is above the Dirac point or due to holes if the Fermi level is below the charge neutrality point. Moreover when the channel makes the transition from n-type to p-type the conduction is due to both types of charges. This of course has an influence on the drain current, as can be seen in figure 1.3.6a, that shows a linear-shape without any saturation or only with a weak saturation which are both disadvantages for the device operation [31]. Such characteristic is due to the lack of the band-gap that does not allow a correct pinch-off of the channel. However some graphene devices are able to show a saturation behavior which then is followed by a second linear region as shown in figure 1.3.6b. In this case at large values of the drain-source voltage V_{DS} there is a crossing between output characteristic at different gate voltages that can lead to a zero or even negative transconductance g_m, which is highly undesirable condition [31].

The saturation region assumes a relevant importance for GFETs applications if the intrinsic voltage gain $A_v = g_m/g_d$ is considered. In order to have values of intrinsic gain comparable to Si MOSFETs the output conductance must be small and this can be achieved only if the GFET operates in the saturation regime. The same discussion can be made for f_T and f_{MAX} of radio-frequency



(a) (b)

Figure 1.3.7: (a) I-V characteristic of a GFET with three different regions, (b) carriers concentration under the top-gate region. Taken from reference [26].

transistors. For this reason it is important to fully understand the nature of the saturation behavior in GFETs. This phenomenon was explained for the first time in 2008 by Meric [26].

For small values of the drain-source voltage the output characteristic is in the linear regime (region I of figure 1.3.7a) and the current along the channel is carried by holes as illustrated in figure 1.3.7b. As V_{DS} starts to increase the carrier density in the channel decreases leading to high electric field. When the potential at the end of the drain equals to the Dirac voltage, the field reaches its maximum value and the part of the channel close to the drain starts to change the type of carriers. In this situation $V_{DS} = V_{\text{DIRAC}}$ and the vanishing carrier density produces a sort of pinch-off region close to the drain and the drain current enters the saturation regime shown in figures 1.3.7a and 1.3.7b. In the region II the drain current is relatively insensitive the the drain-source voltage since the electric field is so high to cause velocity saturation in the channel and allow I_D to assume constant values. The I-V characteristics show a strong saturation regime when the channel is in the unipolar region and moreover the saturation indicates the transition to the ambipolar conduction. As the drain-source voltage is further increased ($V_{DS} > V_{\text{DIRAC}}$) the minimum density point stays inside the channel and produces the pinch-off point that starts to move in the direction of the source end. This condition corresponds to the the region III in figures 1.3.7a and 1.3.7b and it represents the second linear region in the output characteristics of the transistor because the carrier density begins to increase again while the electric

field goes down. In region III, as it can be seen in figure 1.3.7b, the carriers on the left side of the minimum density point are holes and on the drain side electrons. The voltage that drops across the left side of the channel stays at the fixed bias V_{DIRAC} while on the right side the voltage increases to $V_{DS} - V_{\mathrm{DIRAC}}$. In this ambipolar regime the pinch-off point is the place where the holes, coming from the source, and the electrons, injected from the drain, recombine but, since graphene has no band-gap, the recombination does not release energy.

The saturation effect is highly affected by the type of oxide that is used for the gate. In the case of a back-gated device, made by 300 nm of SiO_2, the saturation regime would be weak due to inferior electrostatic coupling with the gate electrode compared to the top-gated transistor. In GFETs it is possible to observe an output characteristic similar to the punch-through of the Si MOSFETs. It can happen at high values of V_{DS} when the transistor starts to be weakly coupled to gate electrode in particular for short-channel devices. For extremely high values of the drain-source voltage, the drain current can lead to irreversible thermal damage of the graphene channel. It should be underlined that an important role in the saturation behavior in a graphene transistor is played by the contact resistance [29]. In fact, in case of poor contacts, a great part of the voltage drops across the contacts and it becomes difficult to observe the saturation region in the device. For that reason in order to reach high electric fields for the saturation characteristics as the quality of graphene is improving, increasing the mobility, it should be necessary to improve the quality of the contacts.

The drain current in function of the drain-source voltage can be modeled using as a reference the cross-view of the GFETs in figure 1.3.8a. Start to consider the carrier concentration in the channel for different values of V_{DS} in the output characteristic as it was shown in figure 1.3.7b. So using a field-effect model the concentration can be expressed as

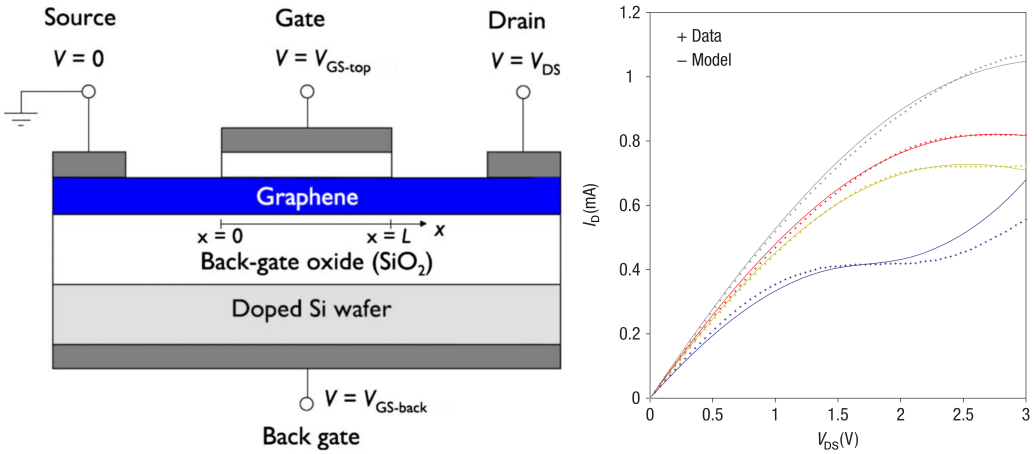
$$n\left(x\right) = \sqrt{n_0^2 + \left(C_{top}\left(V_{GS-top} - V\left(x\right) - V_0\right)/e\right)^2} \tag{1.3.1}$$

where

$$V_0 \cong V^0_{GS-top} + \left(C_{back}/C_{top}\right)\left(V^0_{GS-back} - V_{GS-back}\right) \tag{1.3.2}$$

works as the threshold voltage and $V\left(x\right)$ is the potential inside the graphene channel. $V^0{}_{GS-\mathrm{top}}$ and $V^0{}_{GS-\mathrm{back}}$ represent the top-gate and back-gate voltage of the Dirac point respectively and their role is similar to the one of the flat-band voltage in Si MOSFETs [37].

In order to make an accurate model to derive the drain current, above a



(a) (b)

Figure 1.3.8: (a) Cross-section of a GFET taken from reference [37], (b) comparison of modeled and measured I-V curves taken from reference [26].

certain critical electric field E_{crit} the drift velocity has to reach the saturation value v_{sat}. So there can be a saturation behavior if $V_{GS-\text{top}}$ has a value such that $V_{\text{DIRAC}} > E_{\text{crit}}L$. This model assumes that the velocity saturation in graphene is due to the optical-phonon scattering. The coupling with the phonons is so strong that when the electrons reach the energy threshold for the phonon emission, they are immediately scattered. This assumption implicates that the Fermi sea is shifted by the phonon energy $\hbar\Omega$ and the saturation velocity can be expressed as

$$v_{sat} = v_F \frac{\hbar\Omega}{E_F}. \tag{1.3.3}$$

Differently from semiconductors where the drift velocity is limited from the thermal velocity, in case of graphene the limiting factor is the Fermi velocity v_F and the derivation for v_{sat} is not valid close to the Dirac point. Using a velocity saturation model the drift velocity can be approximated to

$$v_{drift}(x) = \frac{\mu E}{1 + \mu E / v_{sat}} \tag{1.3.4}$$

where it is assumed that the electron and hole mobility and velocity saturation are the same.

With all these considerations the drain current in the graphene channel can be expressed as

$$I_D = \frac{W}{L} \int_0^L en(x)\, v_{drift}(x)\, dx \tag{1.3.5}$$

where L is the channel length and W the channel width. Figure 1.3.8b shows

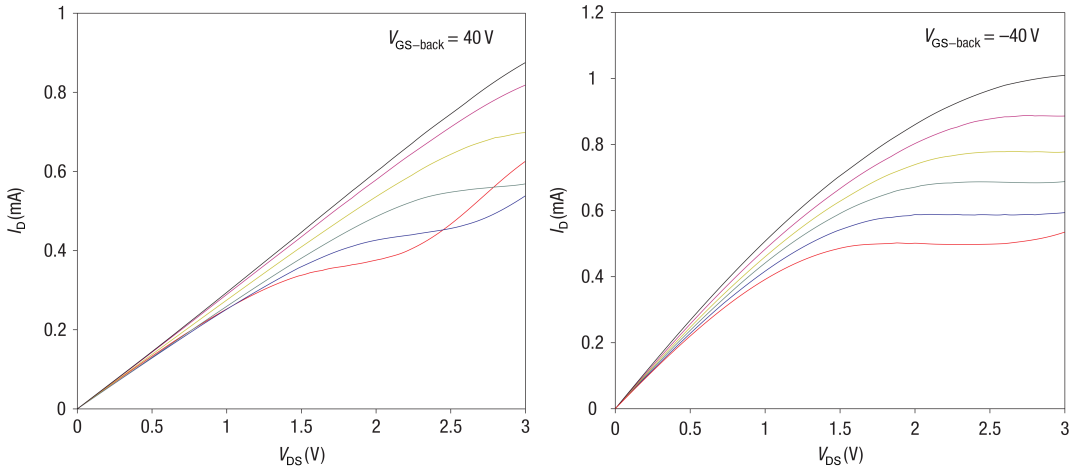
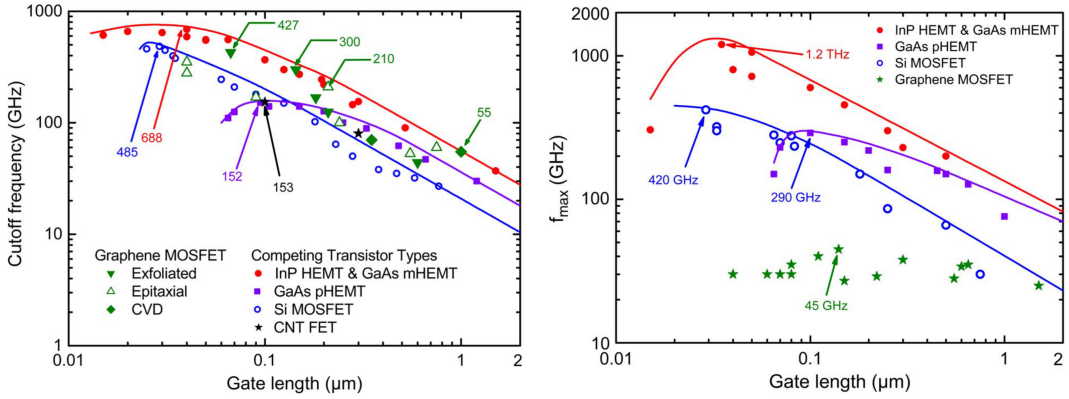


Figure 1.3.9: Drain current in function of drain-source voltage, at different top-gate voltages, for two different values of the back-gate voltage. Taken from reference [26].

the comparison of the output characteristics derived with this model and the experimental data.

The use of back-gate and top-gate in the GFETs can be important. Since the control over the channel by the back-gate is weaker then the top-gate (due to the thicker layer of the oxide) the former could be used to move the Dirac point of the graphene channel as it was explained before. This can make a huge difference because it can allow to have high electric field in the channel and so to reach the saturation velocity. As can be seen in figure 1.3.9a the voltage applied to the back-gate does not allow to have high fields and it can only be observed a linear-shape of the drain current with no saturation or weak saturation. Instead the back-gate voltage applied to the device in figure 1.3.9b allows to have the conditions for the saturation of the drift velocity and the drain current shows the saturation behavior for different values of the top-gate voltage.

Even though GFETs have shown weak current saturation so far, it is impressive to see the values that these devices can achieve in terms of cut-off frequency f_T. In fact only in saturation regime it is possible to reach small values of the output conductance that can maximize f_T [31]. Figure 1.3.10a shows the cut-off frequency of the best performance GFETs realized with exfoliated, epitaxial and CVD graphene compared with the state of the art of the radio-frequency transistors [32]. The highest value of f_T reported so far for graphene MOSFETs is 427 GHz for a device with a gate length of 67 nm [8]. This is still slower than the best conventional radio-frequency transistors but is showing definitely better



(a) (b)

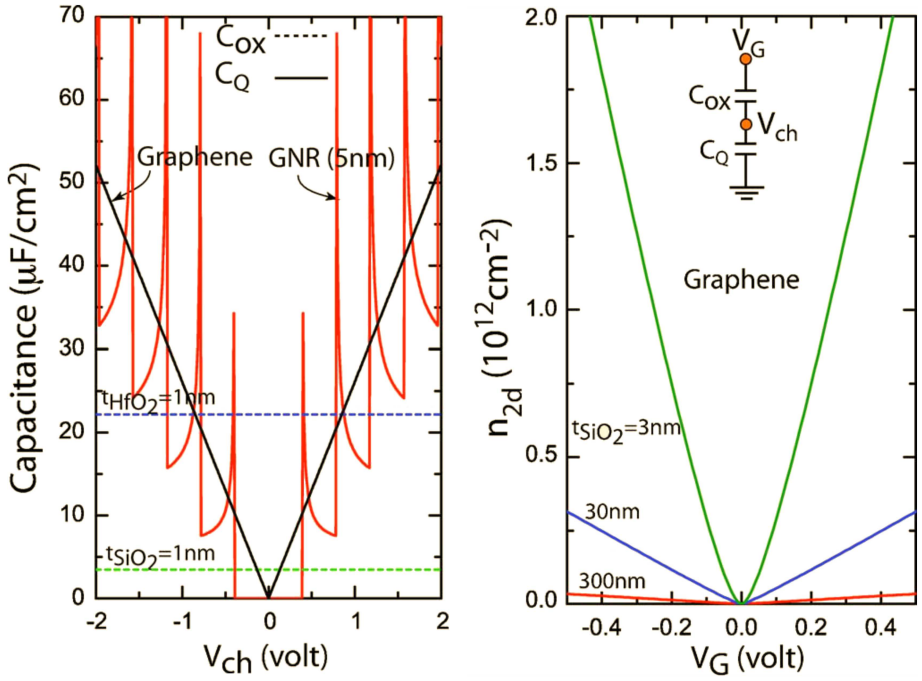
Figure 1.3.10: (a) Cut-off frequency f_T and (b) maximum oscillation frequency f_{MAX} of GFETs in function of the gate length L together with competing radiofrequency FET devices. Taken from refence [32].

performances than Si MOSFETs with the same gate length and is getting really close to the characteristics of GaAs device.

In contrast GFETs show poor performances in terms of maximum oscillation frequency f_{MAX} as it can be seen in figure 1.3.10b. The highest reported f_{MAX} for graphene transistors is 45 GHz which is one order of magnitude smaller than the f_{MAX} of conventional state-of-the-art semiconductor MOSFETs and it even does not show any gate length dependence [32]. The main reason for the different behavior of f_T and f_{MAX} is due to the output conductance g_d which has a second order effect on the cut-off frequency compared to the strong effect that it has on the maximum oscillation frequency. Current saturation and small values of g_d are much more important for power gain than for current gain. For this reason large values of the output conductance in graphene devices significantly reduce the intrinsic gain of GFETs with respect to Si MOSFETs.

1.3.3 Quantum capacitance

As it was shown graphene behaves differently from conventional semiconductor materials used as channels of MOSFETs. Another thing that has to be considered for the performances of GFETs is the gate capacitance together with the density of states (DOS). In fact the gate capacitance C_G in case of transistors with finite density of states cannot be approximated simply with the oxide capacitance $C_{ox} = \varepsilon_{ox}/t_{ox}$, where ε_{ox} and t_{ox} are the dielectric constant and the thickness of the gate oxide [37]. In case of graphene devices the quantum capacitance C_q connected in series with C_{ox} has to be considered. The overall gate capacitance



(a) (b)

Figure 1.3.11: (a) 2D carrier density in a single layer graphene as the function of gate voltage for different oxide thicknesses, (b) quantum capacitance of graphene compared with other parallel-plate capacitances. Taken from reference [9].

in this way becomes

$$C_G = \frac{C_q C_{ox}}{C_q + C_{ox}} \tag{1.3.6}$$

and its value is smaller than the oxide capacitance in particular close to the Dirac point.

Defining V_{ch} as the potential across C_q, as it can be seen in figure 1.3.11b, the quantum capacitance is derived from the channel charges with respect to the channel potential and can be defined as

$$C_q = \frac{\partial Q}{\partial V_{ch}} \tag{1.3.7}$$

and it represents the extra voltage that is required to modulate the charge density in the channel [9]. Differently from 2D electron systems, where the quantum capacitance assumes constant values and it depends only on the effective mass of the carriers in the parabolic limit, in graphene the quantum capacitance depends on the position of the Fermi level since it has a linear band structure [29]. Writing

the total charge in the graphene sheet as

$$Q = q\left(p - n\right) \tag{1.3.8}$$

where q is the electron charge, the quantum capacitance becomes

$$C_q = \frac{2q^2kT}{\pi\left(\hbar v_F\right)^2} ln\left[2\left(1 + cosh\frac{qV_{ch}}{kT}\right)\right]. \tag{1.3.9}$$

Under the condition $qV_{ch} \gg kT$ it reduces to

$$C_q \approx q^2\frac{2qV_{ch}}{\pi\left(\hbar v_F\right)^2} = q^2\rho_{gr}\left(qV_{ch}\right) \tag{1.3.10}$$

where ρ_{ch} is the linear density of states of graphene and its behavior in function of the channel potential is illustrated in figure 1.3.11a compared with other oxide gate capacitances.

Figure 1.3.11b shows instead the dependance of the channel charges in graphene with the gate voltage for different thicknesses of SiO_2 and can be expressed as

$$n_{2D} = n_G - n_q\left(\sqrt{1 + 2\frac{n_G}{n_q}} - 1\right) \tag{1.3.11}$$

where n_G is the carrier density that neglects the quantum capacitance and n_q is the one that arises only from the quantum capacitance [9]. For very thin layer of the gate oxide the modulation of the carriers is very strong since in that condition $C_q \approx C_{ox}$, while the field effect becomes weaker, as the thickness increases, since $C_{ox} \ll C_q$. So reducing the effective oxide thickness of the gate the quantum capacitance of graphene can reach comparable value with the oxide capacitance and have a negative impact on the performances of the device. Moreover the dependance of the gate capacitance could introduce non linearity that could have a negative impact on the linear metrics of amplifiers and mixers [29].

The effect of the quantum capacitance over the gate capacitance is shown in figure 1.3.12 and as it can be seen in this range of voltages C_q can change the gate capacitance by a factor of two. This means that the quantum capacitance must be treated in a correct way, in its voltage dependance, in order to make an accurate modeling of the GFETs. In fact it also influences the drain current of the device through the voltage V_0 of equation 1.3.2, where the overall gate capacitance is represented by the top gate capacitance C_{top}.

In conclusion it can be said that GFETs are a quite new type of devices compared to Si MOSFETs that have been investigated for so many years. Even if

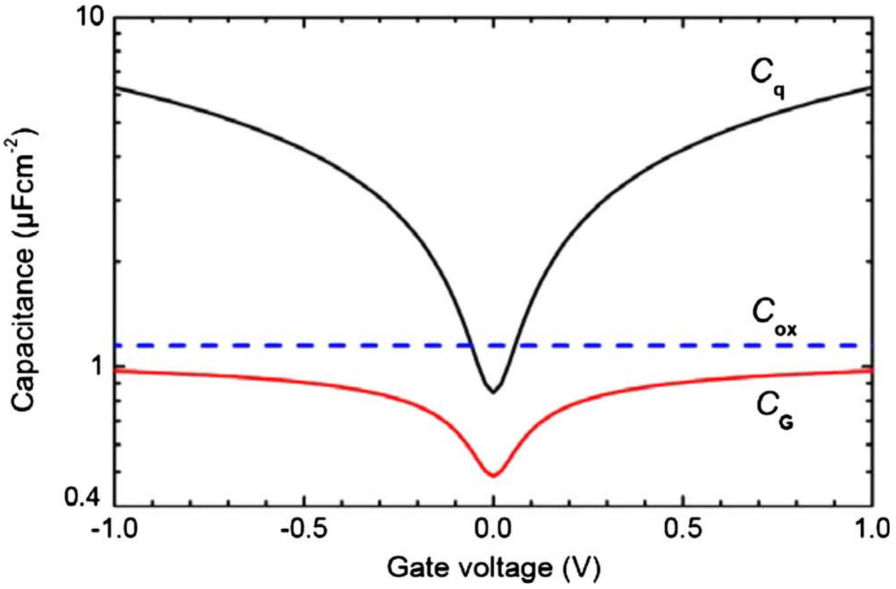


Figure 1.3.12: Effect of the quantum capacitance of graphene on the gate capacitance. Taken from reference [37].

they have a short history, graphene transistors reached impressive performances and further progress can be done. However the missing band-gap still represents a problem when graphene has to compete with conventional semiconductors and for that reason it has not been used so far in logic applications but only in niche analog application. However the unique properties of graphene could help it to become the leading material in other fields such as devices for beyond CMOS logic, devices with ambipolar conduction, applications on flexible substrates and printable electronics [32].

Chapter 2

Graphene digital electronics

2.1 Introduction

The extremely high mobility of charge carriers in graphene at room temperature makes this material suitable for applications in high-speed electronics. However graphene has been used only in niche applications in analogue and radio-frequency electronics so far. In fact the absence of the band-gap results in the impossibility to switch-off GFETs which limits use of graphene in logic applications.

The first graphene devices were not capable of signal amplification which is the ultimate requirement for the fabrication of realistic electronic circuits. For that reason GFETs must exhibit over-unity intrinsic voltage gain $A = g_m/g_d$, where g_m is the transconductance and g_d is the output conductance, in order to be able to realize analogue voltage amplifiers and digital logic gates which represent the main building block of the analogue and digital electronics respectively.

In the past the ambipolar conduction of GFETs was used in digital application to obtain logic gates with a fewer number of transistors compared to Si MOSFETs [33]. Also graphene complementary digital inverters with stable and separated output logic levels have been fabricated [39] but they exhibited very small values of the voltage gain. For this reason this thesis is on the fabrication of graphene logic gates which can be used in realistic high-frequency digital circuit due to the high mobility of graphene.

This chapter illustrates all the steps and the outstanding performances of graphene transistors that effectively allows the fabrication of real digital electronics with graphene, starting from graphene voltage amplifiers to graphene ring oscillators.

2.2 Graphene amplifiers

In order to realize real digital circuits using graphene, graphene voltage amplifiers with a sufficiently high value of voltage gain A_v must be fabricated first. $A_v > 1$ is very difficult to obtain in GFETs because this two dimensional material is limited by the absence of the band-gap which prevents the depletion of charge carriers. This is responsible for the limited control of the gate voltage over the drain current which reduces the transconductance g_m of the GFETs compared to conventional FETs. Moreover the absence of the depletion leads to a weak drain current saturation in graphene devices that increases the value of the output conductance g_d and consequently reduces the voltage gain. Signal amplification is important in electronic circuits since the amplitude of analogue or digital signals must be sufficiently large to drive a load, to overcome the noise or to have well separated logic levels in digital circuits [3].

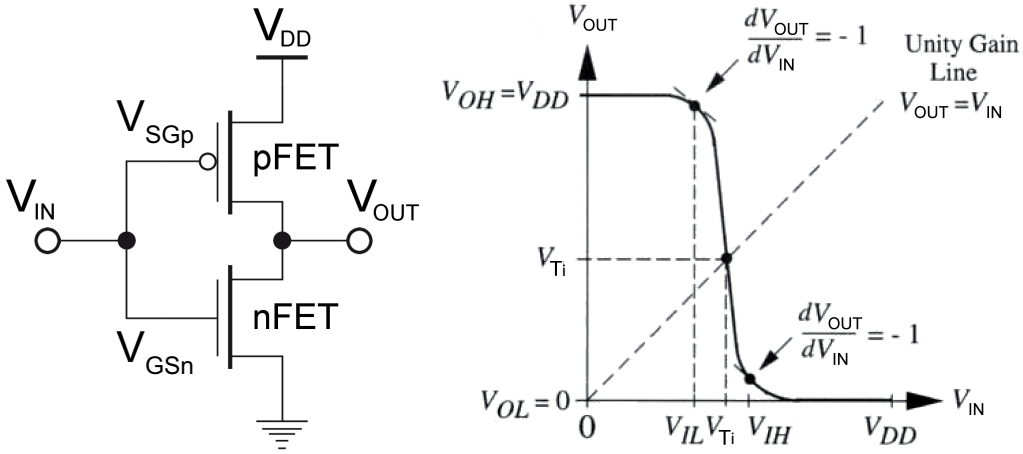
2.2.1 Complementary inverters

Voltage amplifiers can be realized as integrated complementary inverters which are also the main building block of the silicon CMOS digital electronics. A CMOS inverter performs the basic NOT function in digital electronics. An inverter is fabricated by using two MOSFETs of opposite type, as can be seen in figure 2.2.1a, that share the same gate to which the input voltage V_{IN} is applied. The output voltage V_{OUT} is taken at the common drain electrode of the FETs. In this way only one of the two transistors is conductive when there is a stable high or low voltage at the input [18]. Apart from the voltage gain, other important parameters are: the speed, the supply voltage, the power dissipation, the linearity, the noise and also the maximum voltage swing of the device. CMOS digital circuits usually have large voltage swing and dissipate power only when the input signal is switched. The performances of the inverters are also influenced by the input and output impedances since they determine how the circuit interacts with the previous or following stage.

The complementary operation can be understood simply considering the relation between the input signal V_{IN} and the gate-source voltages of the MOSFETs [18]. Looking at figure 2.2.1a the bias voltages of the transistors can be expressed as

$$\begin{aligned} V_{GSn} &= V_{IN} \\ V_{SGp} &= V_{DD} - V_{IN} \end{aligned} \tag{2.2.1}$$

where the input voltage assumes values in the range from 0 to V_{DD} and V_{DD}



(a) (b)

Figure 2.2.1: (a) Circuit diagram of a Si CMOS complementary inverter and (b) the its voltage transfer curve. Taken from reference [18].

is the supply voltage. When the applied input voltage is high $V_{IN} = V_{DD}$, the gate-source voltages of the FETs are equal to

$$\begin{aligned} V_{GSn} &= V_{DD} \\ V_{SGp} &= 0 \end{aligned} \tag{2.2.2}$$

resulting in the cut-off of the p-MOSFET while the n-MOSFET is conductive. In this way the n-type transistor provides the current path to the ground resulting in

$$V_{OUT} = V_{OL} = 0 \tag{2.2.3}$$

where V_{OL} is the output low voltage and is the smallest value that the output signal can assume. If a low voltage is applied to the input $V_{IN} = 0$, it ends up with

$$\begin{aligned} V_{GSn} &= 0 \\ V_{SGp} &= V_{DD} \end{aligned} \tag{2.2.4}$$

and this time the p-MOSFET is conductive while the n-MOSFET is switched-off. Now there is a conductive path to the power supply which results in

$$V_{OUT} = V_{OH} = V_{DD} \tag{2.2.5}$$

where V_{OH} is the output high voltage and is the largest value of the output signal.

To each of the two cases logic levels can be associated as

$$\begin{aligned} V_{OUT} = 0 &\rightarrow \text{Boolean } 0 \\ V = V_{DD} &\rightarrow \text{Boolean } 1 \end{aligned} \tag{2.2.6}$$

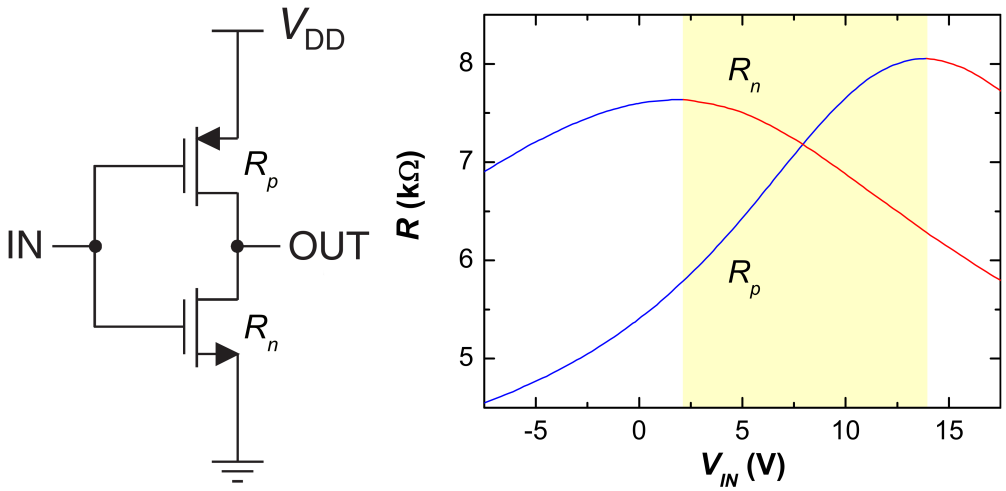
in a positive logic convention. From this, as illustrated in figure 2.2.1b, it can be seen that the inverter exhibits a full-rail output voltage swing (i.e., equal to the entire supply voltage range) which helps to clearly distinguish between logic 0 and logic 1.

Figure 2.2.1b shows the transfer curve of a typical complementary inverter where it is possible to see the inversion operation since at low values of the input voltage the output voltage assumes high values and vice-versa. The steepness of the transition gives an indication regarding the value of the voltage gain of the amplifier since it can be expressed as

$$A_v = \frac{dV_{OUT}}{dV_{IN}}, \tag{2.2.7}$$

i.e., the higher the slope the higher the gain and also the larger the range of input voltages which can be interpreted as logic levels 1 and 0. From the transfer curve other several critical voltages can be extracted: the input voltages V_{IH} and V_{IL} and the inverter threshold voltage V_{Ti}. The input high voltageV_{IH} is the smallest value that the input signal can assume in order to interpret the output of the inverter as a logic level 1. Instead the input low voltage V_{IL} is the highest value of V_{IN} that corresponds to the logic level 0. Both V_{IH} and V_{IL} can be defined as points in which the slope of the transfer curve is equal to -1. V_{Ti} is the inverter threshold voltage which is located at the midpoint between the two Boolean levels. This is also an input voltage at the highest gain point which is located where the intersection of transfer curve and the unity gain line $V_{OUT} = V_{IN}$.

Another important characteristic of this type of device is the switching time. This feature is particularly relevant in CMOS digital circuits since it represents the time that the output signal needs to stabilize after the change of the input voltage. The switching time is influenced by the capacitor of the output node C_{OUT} that needs to be charged and discharged during switching. C_{OUT} is due to all the internal contributions of the FETs and the external load. The switching time is relevant for an inverter since is the limiting factor in terms of high-frequency operation of CMOS digital circuits.



(a) (b)

Figure 2.2.2: (a) Circuit diagram of a graphene inverter, (b) resistance curves as a function of the input voltage. The range of the complementary operation is shaded in yellow [39].

2.2.2 Graphene low-frequency voltage amplifier

Complementary integrated graphene digital inverters have been fabricated for the first time in 2009 [39] using two GFETs of opposite type on the same graphene flake. In contrast to silicon, graphene has an ambipolar conduction and therefore only for certain values of the input voltage the complementary operation can be achieved. As fabricated both the GFETs show the same behavior in terms of the position of the Dirac points and in this way it is impossible to have two channels with carriers of the opposite type. For that reason an electrical annealing was performed in order to change the position of the charge neutrality point of the FET R_n of the inverter in figure 2.2.2a. This process removes ambient contamination due to the Joule heating of the chosen channel and shifts the Dirac point to lower values of V_{IN} compared to the other transistor as it can be seen in figure 2.2.2b. The complementary operation is obtained using input voltages in between the two Dirac points (yellow area in figure 2.2.2b) where an n-channel (R_n) and a p-channel (R_p) are formed at the same time. However, The electrical annealing process is quite detrimental for the devices and the yield of working transistor after this process is quite low.

Differently to Si CMOS in which there is no static power dissipation, the graphene inverter is always conductive due to the absence of the band-gap. As a

consequence the output stage dissipates power which can be expressed as

$$P = \frac{V_{DD}^2}{R_n + R_p} \tag{2.2.8}$$

where R_n and R_p are the resistances of the two GFETs. One way to reduce the static power dissipation could be achieved by increasing the resistance of the GFETs but this would also affect the transient response time leading to slower devices compered to the standard Si CMOS technology. In case of graphene inverters there is a trade-off between the static power dissipation and the maximum speed that the device can reach.

Until 2011 all analog graphene circuits were not capable of amplifying small alternating current (AC) voltage signals since these devices were exhibiting low voltage gain mainly due to the fact that back-gated transistors (made with a thick layer of 300 nm of SiO_2) were used. This limits the extent of modulation of the carriers and reduces the transconductance g_m. Low g_m suppresses the signal amplification which limits the use of graphene not only in analogue electronics, but also in digital electronics, because this prevents the coupling between digital logic gates due to a mismatch of the output and input voltage swing. For that reason graphene voltage amplifiers should be fabricated using top-gated structures that use thin high-k gate insulators since $A_v \propto \varepsilon_{ox}/t_{ox}$, where ε_{ox} and t_{ox} are the dielectric constant and the thickness of the oxide respectively.

It has been shown that it is possible to obtain a very thin layer of aluminum oxide simply by evaporating an aluminum layer directly on graphene [20]. As the device is exposed to the air the aluminum oxidizes and form a very thin ($\sim$ 4 nm) layer of oxide at the interface with graphene, as shown in figure 2.2.3 [24]. The aluminum oxide has also the advantage to lower the gate voltages of the transistors with respect to those used in standard back-gated graphene devices. This is due to an enhancement of the coupling capability with respect to the conventional 300 nm SiO_2 back-gate due to a thinner layer of oxide and a larger relative permittivity. This AlO_x/Al gate stack was used to fabricate digital graphene logic gates that were exhibiting an over-unity voltage gain but only at cryogenic temperatures [20].

Another way to increase the voltage gain is to improve the current modulation of graphene devices by to fabricating graphene nanoribbons (GNRs) in which quantum confinement opens a band-gap. In this way the current ON/OFF ratio

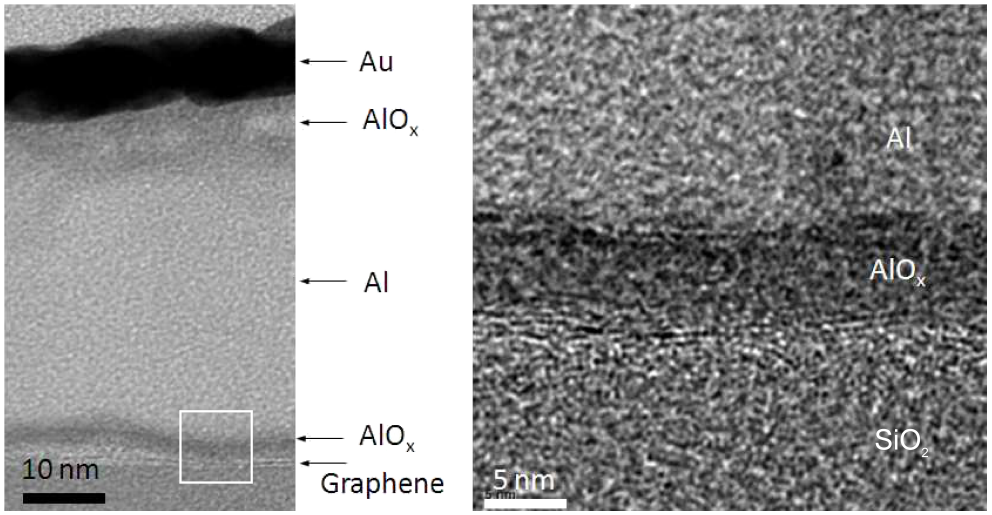


Figure 2.2.3: Cross-sectional TEM image of the AlO_x/Al gate stack. Taken from reference [24].

is increased and since g_m can be expressed as

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \approx \frac{I_{D,ON} - I_{D,OFF}}{V_{GS,ON} - V_{GS,OFF}} = \frac{I_{D,ON}}{V_{GS,ON} - V_{GS,OFF}} \tag{2.2.9}$$

this should enlarge the value of the transconductance. However this advantage is significantly reduced by the fact that in the state-of-the-art GNRs the ON current is strongly attenuated by carrier scattering on disordered ribbon edges.

Over-unity voltage gain equal to 4.5 dB was achieved at room temperature but only using a complex six-finger-gate FET configuration [14]. However such a value of A_v is not large enough for realistic applications of graphene amplifiers since voltage gain larger than 10 dB is needed. Without this graphene circuits should always rely on Si MOSFETs for the signal amplification and therefore a hybrid technology would be needed, which leads to increased production costs.

In order to realize graphene voltage amplifiers exhibiting over-unity voltage gain, AlO_x/Al gate stack was used in graphene inverters fabricated on single graphene flakes [13]. Graphene flakes were exfoliated from a highly ordered pyrolytic graphite with the standard scotch-tape method and deposited on SiO_2/Si substrates. Exfoliated monolayers were used due to their high crystallographic quality in order to estimate the upper performance limits of graphene amplifiers. All the electrodes of the devices were patterned by electron beam lithography (EBL) followed by the deposition of the contacts in an electron beam evaporator. In the first step, 100 nm of Al were evaporated directly on graphene for the realization of the common gate. As the devices are exposed to the air a

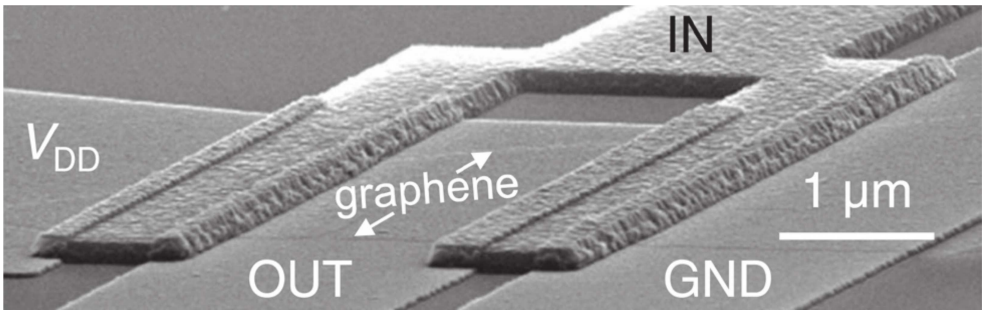


Figure 2.2.4: A scanning electron microscope image of the graphene amplifier [13].

very thin layer of AlO_x naturally forms at the interface between graphene and Al. Typical dimension of the graphene channels were length L = 1 μm and width W = 3 μm. In the second step source and drain contacts, made of 5 nm of Ti and 35 nm of Au, were patterned. These contacts were overlapped with the gate electrode as it can be seen in figure 2.2.4. The aim of the overlap was to fabricate a graphene channel that is fully covered by a gate as in conventional Si MOSFETs. This increases g_m and therefore A_v because the access resistances (ungated parts of the channel) are eliminated. The AlO_x layer forms all around the Al gate electrode and this prevents short circuits with the source and drain contacts. Moreover the complete coverage of the channel with the gate electrode helps to dissipate the heat more efficiently allowing to reach higher drain currents I_D and so larger voltage gains A_v. The absence of ungated parts also stabilizes the electrical properties of the devices during the operation since it reduces the influence of atmospheric contaminants at high currents. Moreover the graphene channel is partially screened by the gates from water charge traps present on the substrate which suppress the hysteresis.

It was found that it was extremely important to fabricate the gate first in order to have reasonably large yield of working devices. In fact if source and drain electrodes are realized before gates the yield of working devices was extremely low ($\sim 5\%$). This could be due to the large thermal expansion coefficient of Al, one of the largest among metals, and is three times larger than that of Ti.

Figure 2.2.5a shows the circuit diagram of the fabricated graphene amplifiers where Z = 1 MΩ||13 pF is the input impedance of the oscilloscope that was used to measure the input and output signals. 50 Ω is output resistance of the input source meter and R_L is used to simulate the input resistance of the next stage. The Dirac point shift that allows to obtain the complementary operation is different from the one previously explained. As fabricated both FETs (F1 and F2) in the inverter have the same charge neutrality point, as it can be seen in figure 2.2.5b.

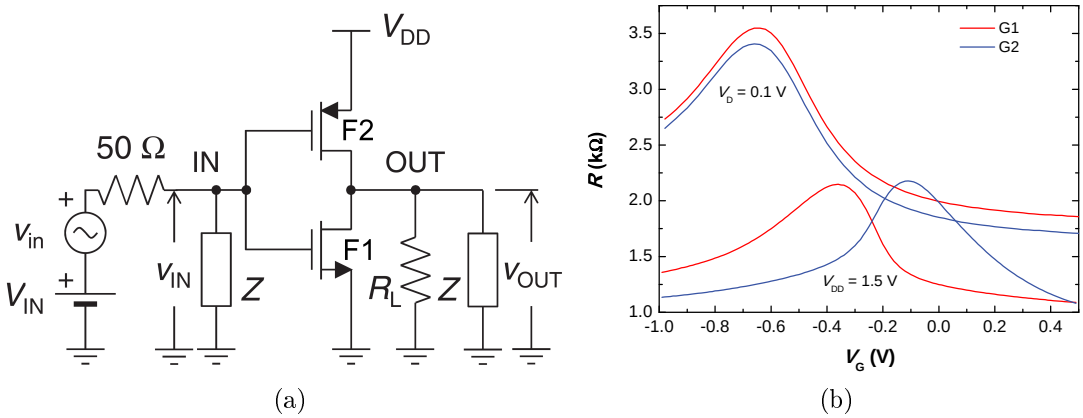


Figure 2.2.5: (a) Circuit diagram a the graphene voltage amplifier, (b) Dirac points shift for complementary operation [13].

In order to shift the Dirac points it is enough to apply a supply voltage $V_{DD} > 0$. This increases the potential of the channel F2 with respect to the potential of the channel F1 and shifts the Dirac point of F2 to higher input voltages with respect to F1 as is evident in figure 2.2.5b. In particular the complementary operation is obtained when the input voltage V_{IN} assumes values approximately between -0.4 V and -0.1 V. The fabricated complementary inverter is considered to be fully integrated since both transistors are made on the same graphene flake. It has to be underlined that the load resistance R_L and the output resistance of the input voltage source are not necessary for the signal amplification.

The transfer curve, the output voltage V_{OUT} as a function of the input voltage V_{IN}, and the corresponding voltage gain of graphene amplifier are shown in figure 2.2.6, for a supply voltage $V_{DD} = 1.5$ V and the load resistance $R_L \to \infty$. The blue curve represents the static (DC) transfer curve and the Q point is the point at which the amplifier has the highest voltage gain, which in this case is $|A_v|_{max} = 2.1$. In particular, the Q point is placed at $V_{IN,Q} = -0.22$ V and $V_{OUT,Q} = 0.71$ V and differently from conventional complementary inverters it is not located where the input and output signals are equal. The low-frequency voltage gain in figure 2.2.6 is simply obtained as a derivative of the transfer curve, equation (2.2.7), and the Q point is placed exactly were the slope of the transfer has a maximum. The red curve is the dynamic (AC) transfer curve obtained by applying an input voltage $v_{IN} = V_{IN} + V_{in}sin(2\pi ft)$, where $V_{IN} = -0.22$ V and $V_{in} = 0.35$ V at a frequency $f = 1$ kHz. The small hysteresis that can be observed is not due to hysteric behavior of the graphene transistors, since the channel is fully covered, but due to parasitic reactive components present in the circuit. When the graphene amplifier is biased at the operating point Q and a

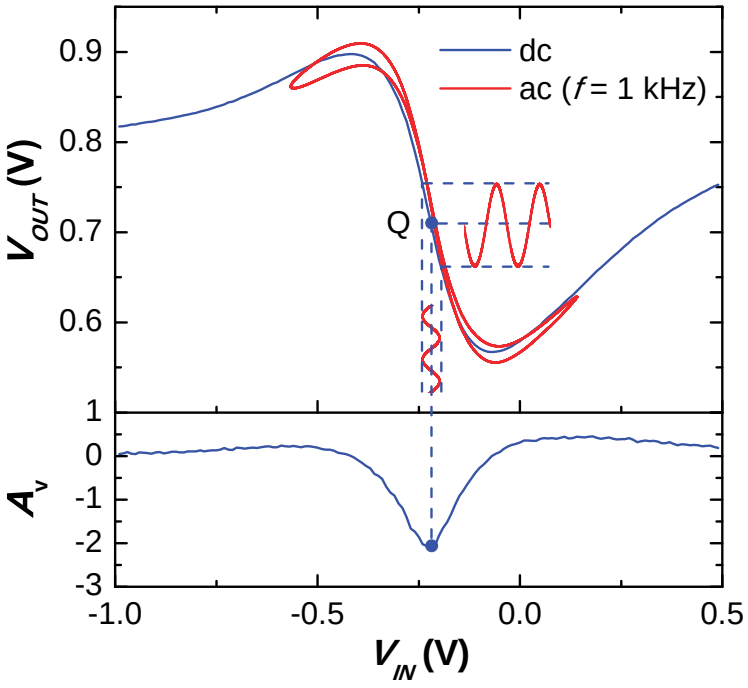


Figure 2.2.6: DC (blue) and AC (red) transfer curves of the amplifier together with the low-frequency voltage gain at room temperature [13].

small AC signal is super-imposed to the input signal $V_{IN,Q}$ the amplitude of the output AC signal is amplified by the factor $|A_v|_{max}$.

In order to improve the low-frequency voltage gain it is necessary to derive its expression in order to find which parameters affect its value. Considering the circuit of the fabricated graphene amplifier and assuming that both the GFETs have approximately the same characteristics i.e., $g_{m1} = g_{m2} = g_m$ and $r_{d1} = r_{d2} = r_d$, the voltage gain A_v can be expressed as

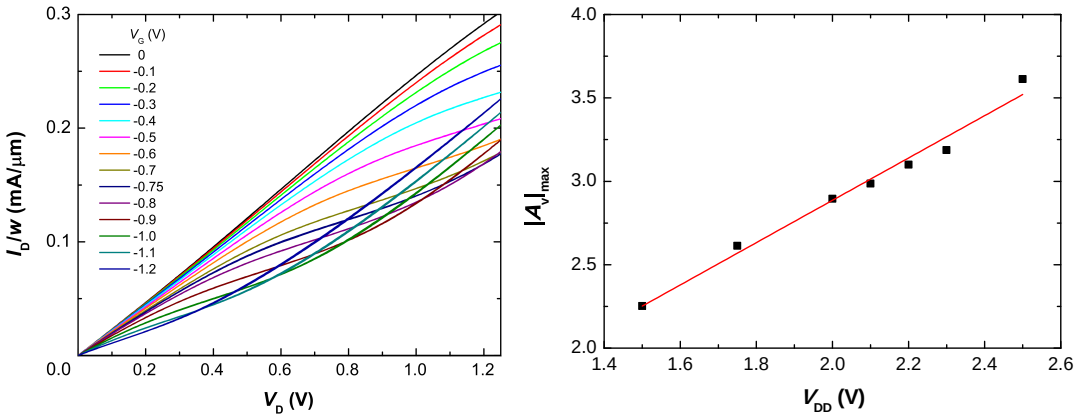
$$A_v = -2g_m(\frac{r_d}{2} \parallel R_L). \tag{2.2.10}$$

A_v reaches maximum value for $R_L \to \infty$ when

$$A_v = -g_m r_d. \tag{2.2.11}$$

Now it is necessary to find expressions for g_m and r_d in function of the parameters of the DC circuit. To do this a simple model in which the GFETs are treated as voltage-controlled resistors will be used. With this consideration the drain current I_D is equal to

$$I_D = \frac{W}{L}(\sigma_0 + \frac{\varepsilon_{ox}}{t_{ox}}\mu \mid V_{GS} - V_0 \mid)V_{DS} \tag{2.2.12}$$



(a) (b)

Figure 2.2.7: (a) Current-voltage characteristics of one of the FETs of the complementary inverter, (b) maximum value of the voltage gain of the graphene amplifier in function of the supply voltage V_{DD} at room temperature. Taken from reference [13].

where σ_0 is the conductance of graphene at the Dirac point, ε_{ox} and t_{ox} are the dielectric constant and the thickness of the oxide respectively, μ is the mobility of the carriers in graphene and V_0 is the Dirac voltage. Knowing this g_m and r_d can be expressed as

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{\varepsilon_{ox}}{t_{ox}} \mu V_{DS} \frac{W}{L} \tag{2.2.13}$$

$$g_d = \frac{1}{r_d} = \frac{\partial I_D}{\partial V_{DS}} = (\sigma_0 + \frac{\varepsilon_{ox}}{t_{ox}} \mu (V_{GS} - V_T)) \frac{W}{L} \tag{2.2.14}$$

and consequently the low-frequency voltage gain is

$$A_v = \frac{V_{DS}}{V_{GS} - V_0 + \sigma_0 t_{ox}/(\varepsilon_0 \mu)}. \tag{2.2.15}$$

The maximum value of the voltage gain is obtained when the amplifier is biased at the operating point Q at which

$$\begin{aligned} V_{DS} &\approx \tfrac{V_{DD}}{2} \\ V_{GS} - V_0 &\approx \tfrac{V_{02} - V_{01}}{2} = \tfrac{\triangle V_{IN}}{2} \end{aligned} \tag{2.2.16}$$

that gives

$$|A_v|_{max} = \frac{V_{DD}}{\triangle V_{IN} + 2\sigma_0 t_{ox}/(\varepsilon_{ox}\mu)}. \tag{2.2.17}$$

It has to be underlined that $\triangle V_{IN}$ depends on the value of V_{DD} and that there is a proportionality between the maximum voltage gain and V_{DD}. From equation (2.2.17) it can be observed that in order to increase the voltage gain high-k

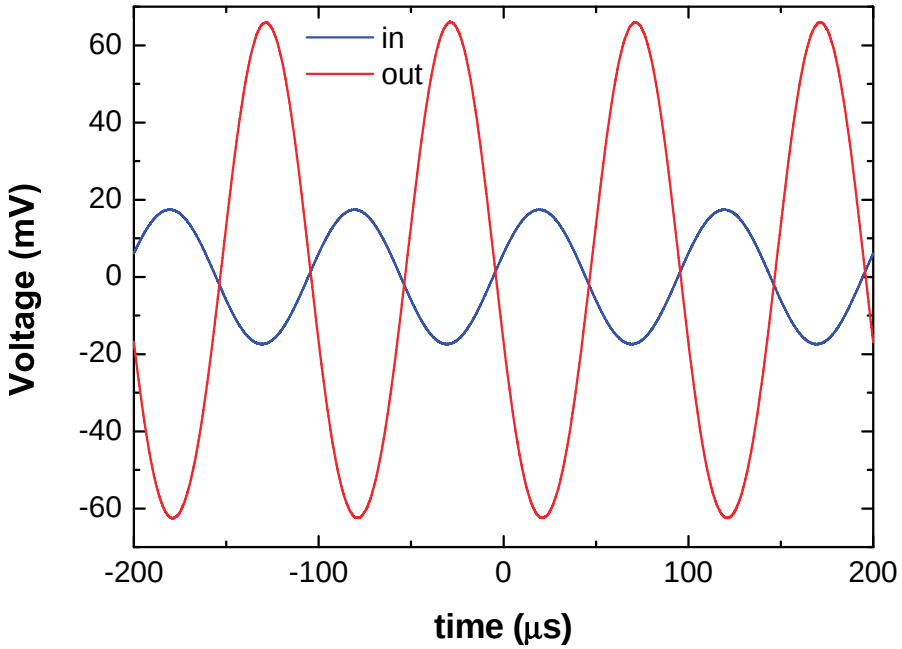
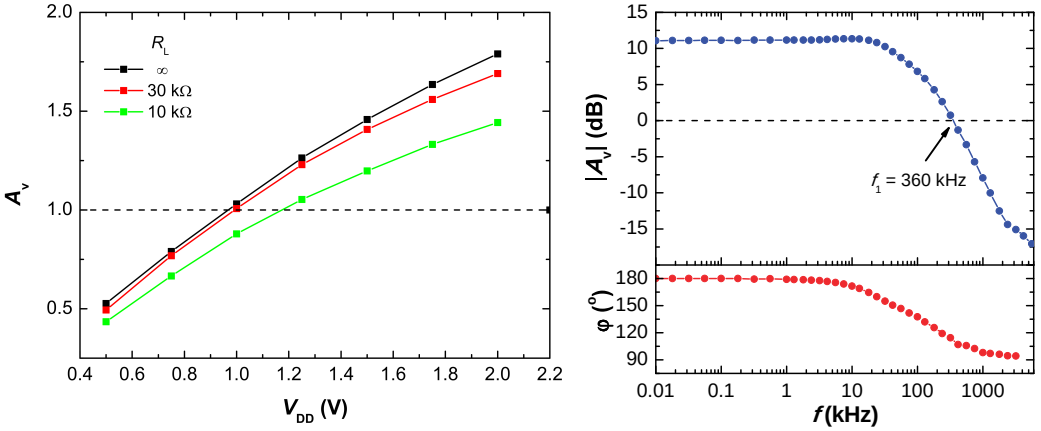


Figure 2.2.8: Amplification of the AC component of the input voltage at a frequency of 10 kHz at room temperature [13].

thin gate insulator should be used, as explained before, and high supply voltages should be applied to the complementary inverter. This has been experimentally demonstrated with the graphene amplifier as it can be seen in figure 2.2.7b. Increasing the value of V_{DD} it is possible to increase linearly the maximum voltage gain which reaches $|A_v|_{max} = 3.7$ or 11.4 dB at $V_{DD} = 2.5$ V. At this supply voltage it was possible to obtain a value of the transconductance $g_m/W \approx 0.5\,\mathrm{mS}/\mu\mathrm{m}$ that combined with a relatively large value of the output resistance $r_d W = 10.5\,\mathrm{k\Omega}\cdot\mu\mathrm{m}$ (even without saturation of the drain current) allows voltage gains above 10 dB at room temperature.

Figure 2.2.8 shows the AC components of the input and output signals of the voltage amplifier measured at $V_{DD} = 2.5$ V, frequency $f = 10$ kHz and $R_L \to \infty$. In this particular case the maximum voltage gain equal to 3.7 was reached with the DC component of the input signal $V_{IN} = 0.15$ V. The amplitude of the oscillations of the output signal was larger than 60 mV. With such a supply voltage a maximum drain current density $I_D/W = 0.26\,\mathrm{mA}/\mu\mathrm{m}$ was reached at the Q point which is five times smaller than the breakdown current density of exfoliated graphene [23]. In theory it would be possible to increase V_{DD} without destroying the graphene channels but this was not attempted for other reasons. In fact with V_{DD} applied to F2, the potential that drops across the gate oxide is equal to $V_{DD} - V_{IN}$. For supply voltages larger than 2.5 V the potential across the oxide is higher than 2.35 V which is close to the breakdown voltage of the



(a) (b)

Figure 2.2.9: (a) Voltage gain of the complementary amplifier measured in dynamic mode at $f = 1$ kHz for different values of the load resistance, (b) frequency response of the magnitude of the voltage gain of the graphene amplifier and the phase shift between input and output signals [13].

gate insulator. The measured leakage current of all fabricated devices did not influence the performances of the inverters since $I_G < 1$ nA.

The output resistance of the graphene complementary amplifier $r_{d1} \parallel r_{d2} = r_d/2 \sim 1.8\,\text{k}\Omega$ should be able to preserve the high gain when the output of the inverter is loaded with a resistance larger than 18 kΩ. The attenuation of the signal caused by the voltage drop across the output resistance of the amplifier is present at all finite input resistances of the next amplifying stage. However, as long as $R_L \gg r_d/2$ the attenuation is negligible. This is demonstrated in figure 2.2.9a which shows A_v of the graphene voltage amplifier for three different finite values of R_L.

Figure 2.2.9b shows the frequency response and the phase shift of one of the fabricated amplifiers for frequencies below 5 MHz. As it can be seen the maximum voltage gain of 11.4 dB is preserved up to $f = 20$ kHz and it decreases by 3 dB at $f_{-3dB} = 70$ kHz which defines the bandwidth of the amplifier. For large values of the frequency the magnitude of the gain is reduced by 18 dB/dec, which is very close to the 20 dB/dec that indicates the presence of a dominant pole at f_{-3dB}. The fabricated inverter has an unity-gain frequency $f_1 = 360$ kHz at which it can be used as a buffer stage, while for higher frequencies the signal is attenuated. The phase shift also confirms that the device is able to perform the inversion of the signal at low frequencies while at higher frequencies the shift is reduced to 90°. The plots in figure 2.2.9b indicate the presence of a dominant pole at $f_{-3dB} = 70$ kHz. This pole does not originate from the amplifier but

from the parasitic capacitances of the cables used to connect the device to the measurement equipment. A total conductor-to-ground capacitance $C_c \approx 0.5$ nF was found which gives a pole frequency $f_p \approx 100$ kHz that almost coincides with the measured cut-off frequency.

It was also found that almost all realized devices exhibited over-unity voltage gain, but only a smaller number, typically 20%, exhibited high voltage gain. For high gain it is important to have two identical GFETs in an inverter with symmetric transfer curves I_D vs V_{GS} even at large supply voltages V_{DD}. In comparison silicon amplifiers exhibit higher values of voltage gain, typically larger than 20 dB. The demonstrated technology is not scalable since it is based on graphene flakes obtained by mechanical exfoliation and it suffers from large static power dissipation due to the impossibility to turn-off the graphene transistors. However these results reported the highest voltage gain obtained at room temperature for any kind of graphene amplifiers realized until that time. The obtained results are important since they pave the way for the use of GFETs as the main building blocks of analogue electronics.

2.2.3 Graphene audio voltage amplifiers

A possible application of graphene voltage amplifiers could be amplification of audio signals since the fabricated devices are capable of high signal amplification at room temperature in the audio frequency range ($<$ 20 kHz). In the audio applications it is also necessary that amplifiers work in a linear mode without introducing any noticeable harmonic distortions.

Conventional audio amplifiers are made in two stages. The first stage, that is called preamplifier, has the purpose to amplify the input signal and has the same functionality as previously discussed graphene voltage amplifiers. The second stage is a power amplifier which exhibits a unity voltage gain and is used to match the signal amplified by the previous stage to a low impedance load such as a loudspeaker. Audio voltage amplifiers have to provide a high fidelity reproduction and for this reason they should have a very low noise figure. Graphene is a perfect candidate for that type of applications since it exhibits very low $1/f$ noise which is the main noise component that affects the audio frequency range. Moreover audio voltage amplifiers realized with graphene should also benefit from its high mechanical and chemical stability and its high thermal conductivity [2].

Figure 2.2.10 shows the power spectrum of the output signal of the graphene voltage amplifier at frequencies below 20 kHz and biased at $V_{DD} = 1.5$ V and $R_L \to \infty$, when a sinusoidal signal at a frequency $f_0 = 1$ kHz with an amplitude

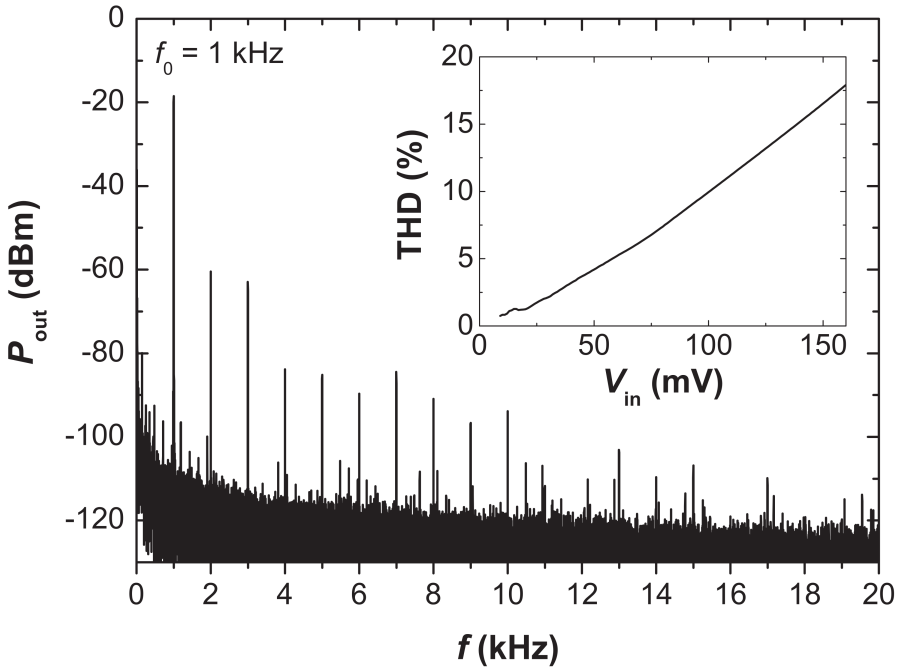


Figure 2.2.10: Power spectrum and the total harmonic distortion (insert) of the output signal [13].

V_{in} = 18 mV is applied at the input. The non linearity of the complementary amplifier results in the generation of higher order harmonics at frequencies $f = nf_0$ where $2 \leq n \leq 20$. For the fabricated device the second harmonic is 42 dB below the first one and as the main contributions to the harmonic distortion are due to the second and third harmonics. The total harmonic distortion (THD) is a measure of the non linearity of the voltage amplifiers and is calculated as the square root of the total output power of all the higher order harmonics divided by the power of the first harmonic [13]. The insert of figure 2.2.10 shows the THD as a function of the amplitude of the sinusoidal input signal. In the case shown in the main panel (V_{in} = 18 mV) $THD = 0.999\%$, which demonstrates the high fidelity of the reproduction of the amplifier since the signal is considered audible for THD less than 3%. As the amplitude of the input sinusoidal signal is increased, the total harmonic distortion also increases due to increased device non linearity. $THD = 17.5\%$ is reached at the amplitude $V_{in} = 160$ mV when for the output signal reaches the minimum and maximum of the transfer curve where the gain drops to zero. In conclusion the fabricated graphene voltage amplifier could be used for high fidelity signal amplification in audio applications since the total harmonic distortion is smaller than 1% at low input signal amplitudes.

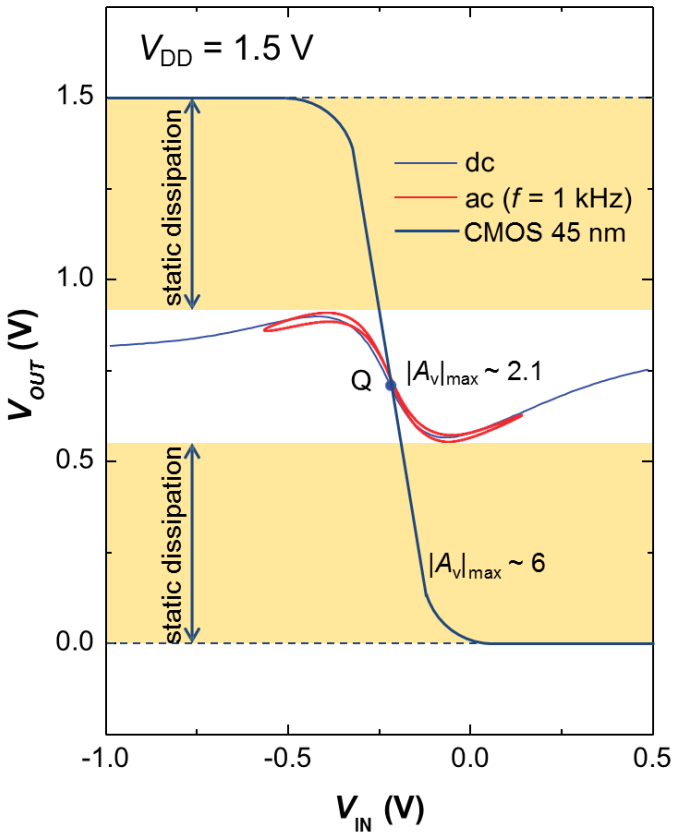


Figure 2.2.11: Transfer curves of conventional silicon CMOS with gate length of 45 nm and of the fabricated graphene voltage amplifier.

2.2.4 Digital logic gate

Successful fabrication of graphene voltage amplifiers exhibiting over-unity the voltage gain opens a possibility to realize digital logic gates that are the main building blocks of digital electronics. The complementary inverter could be used as a digital logic gate simply applying the square-wave signal at the input. However the fabrication of a graphene amplifier represents only the first step in the realization of realistic graphene digital circuits. High-gain in logic gates operating at room temperature has not been demonstrated so far. High values of the voltage gain in digital applications are required in order to distinguish between the two logic levels and to match the input and output signals. Without the matching it is not possible to cascade digital logic gates and realize realistic graphene digital circuits.

Figure 2.2.11 shows a comparison between the transfer curves of a graphene inverter and a conventional silicon CMOS with a gate length of 45 nm at the same supply voltage $V_{DD} = 1.5$ V. Si inverters have a voltage gain which is almost 3 times larger with respect to that of graphene inverters but the gate length is 3

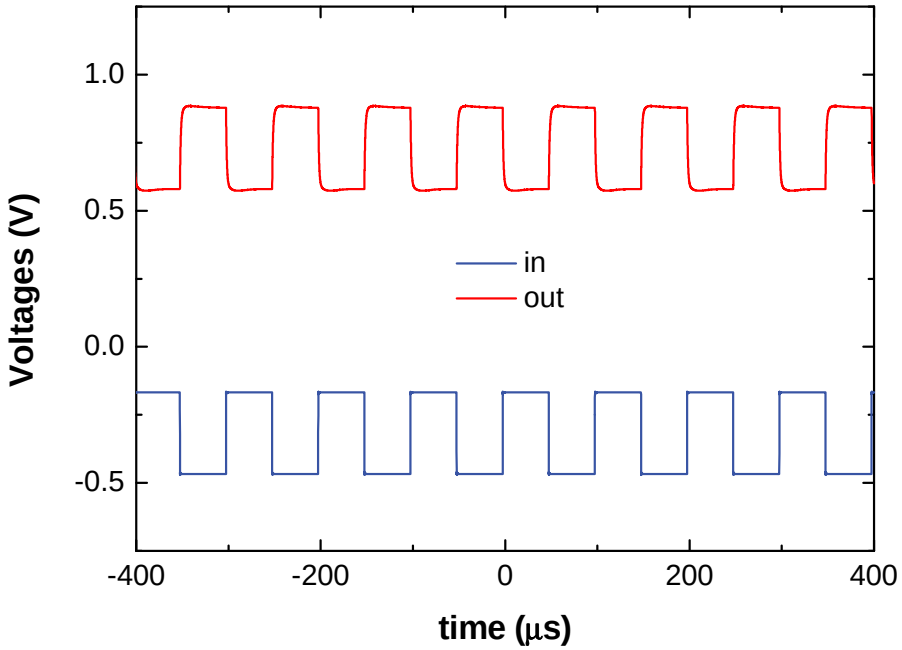


Figure 2.2.12: Digital waveforms of the input and output signals measured on the fabricated graphene voltage amplifier.

μm in the latter compared to 45 nm in the former. What is more evident is the difference of the output voltage swings between the two devices. In the case of silicon inverter the voltage swing reaches almost 100% of the supply voltage V_{DD} compared to $\sim$ 20% of the graphene inverter in which the GFETs cannot be turned-off in both logic states. This could affect the ability to distinguish between the two logic levels. As it can be seen in figure 2.2.11 the graphene device suffers from large static power dissipation represented by the yellow areas. Those are mainly due to the fact that in the two logic states both GFETs are in the ON state (in contrast to silicon inverter in which one FET is always switched-off). Therefore graphene logic gates cannot compete with conventional semiconductors in terms of power consumption.

A square-wave signal at $f = 5$ kHz with an offset equal to V_{Th} (which is the threshold voltage of the inverter) was applied to the fabricated graphene complementary inverter was biased at $V_{DD} = 1.5$ V. As it can be seen in figure 2.2.12 it was not possible to obtain the matching between the input and output signals even though it was possible to distinguish the two logic levels. It was found that the mismatch between the signals is equal to the voltage at the Dirac point of the unbiased GFETs. All GFETs were realized with mechanically exfoliated graphene having charge neutrality points far away from zero due to the absorption of ambient impurities on the flake prior to the fabrication. In order to match the signals it is necessary to fabricate GFETs with Dirac points close to zero.

2.3 Cascading wafer-scale graphene inverters

As it was shown in the previous section there are same issues that has to be addressed in order to improve graphene logic gates. It should be noted that all previously demonstrated devices were realized using mechanically exfoliated graphene which is not a scalable technology since typical lateral dimensions of this material are $\sim$ 10 μm. In order to solve the problem of the scalability of complementary graphene inverters it would be necessary to use large-scale graphene films grown by chemical vapor deposition (CVD). This material was provided by the group of Prof. Eric Pop from the Stanford University (previously at the University of Illinois at Urbana-Champaign). Graphene monolayers where grown by CVD on Cu foils with a CH_4 precursor at 1000 °C [43]. After growth Raman spectra were acquired to confirm the absence of the D peak which indicates the presence of defects in the graphene film. Subsequently graphene was removed from the Cu foil and transferred to conventional Si substrates with 300 nm of SiO_2 on top in order to realize graphene logic gates.

2.3.1 Input and output matching

GFETs which were fabricated on large-scale graphene films and integrated into digital complementary inverters are shown in figure 2.3.1. Several graphene stripes with a width $W = 20$ μm and a length $L_{st} = 100$ μm were patterned using EBL and then etched using the reactive ion etcher with an oxygen plasma. After etching, AlO_x/Al gate stacks were fabricated by direct evaporation of Al on graphene stripes followed by the fabrication of source and drain contacts, made of Ti/Au (5/35 nm or 2/35 nm). These contacts partially overlap with the gate electrodes in order to have a full coverage of the channel. This was obtained by a self-aligned top-gate fabrication process that does not leave ungated parts. A thinner adhesion layer of Ti helps to reduce the contact resistance of the FETs and increases the voltage gain. From the optical image of the fabricated structure in figure 2.3.1 it is possible to distinguish six gate electrodes coming from the right that corresponds to three different complementary inverters. The orange electrodes are the source and drain contacts. In order to compact the structure and therefore minimize the inhomogeneity between the GFETs the source contacts are shared between two neighboring inverters (the third electrode from the bottom and the top) [30].

GFETs were realized with channel length $L = 2$ μm and exhibited the same characteristics as fabricated. In order to obtain the complementary operation the

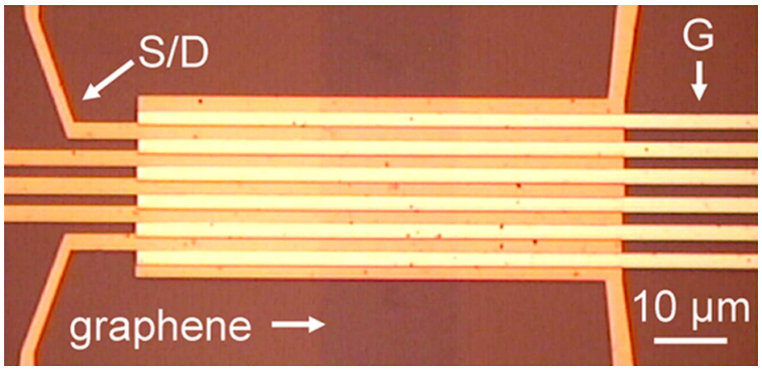


Figure 2.3.1: Optical microscope image of a series of graphene integrated complementary inverters [30].

input voltage V_{IN} must lie between the Dirac points of the two transistors of the inverter. This was achieved by applying a supply voltage $V_{DD} > 0$. It has been found that the highest gain point Q lies halfway between the Dirac points of the FETs. For that reason when negligible values of the supply voltage are applied to the amplifier the two Dirac points are not shifted and the input and output voltages at the highest gain point are

$$\begin{aligned} V_{IN} &= V_0 \\ V_{OUT} &= 0 \end{aligned} \tag{2.3.1}$$

where V_0 is the Dirac voltage of the unbiased transistor. As V_{DD} is increased the corresponding voltages are

$$\begin{aligned} V_{IN} &\approx V_0 + \tfrac{V_{DD}}{1+\alpha} \\ V_{OUT} &= \tfrac{V_{DD}}{1+\alpha} \end{aligned} \tag{2.3.2}$$

where α is the ratio of the resistances of the two FETs at $V_{IN} = 0$. In this way

$$V_{IN} - V_{OUT} = V_0 \tag{2.3.3}$$

represents the mismatch between the input and output voltages at the highest gain operating point Q. For that reason, in order to be able to match the input and output signals in air, inverters with GFETs that exhibit Dirac points close to zero (typically $V_0 < 0.2$ V) were used in measurements.

Figure 2.3.2 shows the DC characteristics of the graphene inverter under ambient conditions for different values of the supply voltage. As it can be seen, it was possible to reach voltage gain of $|A_v|_{max} \approx 5$ at $V_{DD} = 2.5$ V , which is higher than the results obtained with graphene devices fabricated in section 2.2.2. Amplifiers with GFETs with smaller gate length, 1 μm and 500 nm, were also

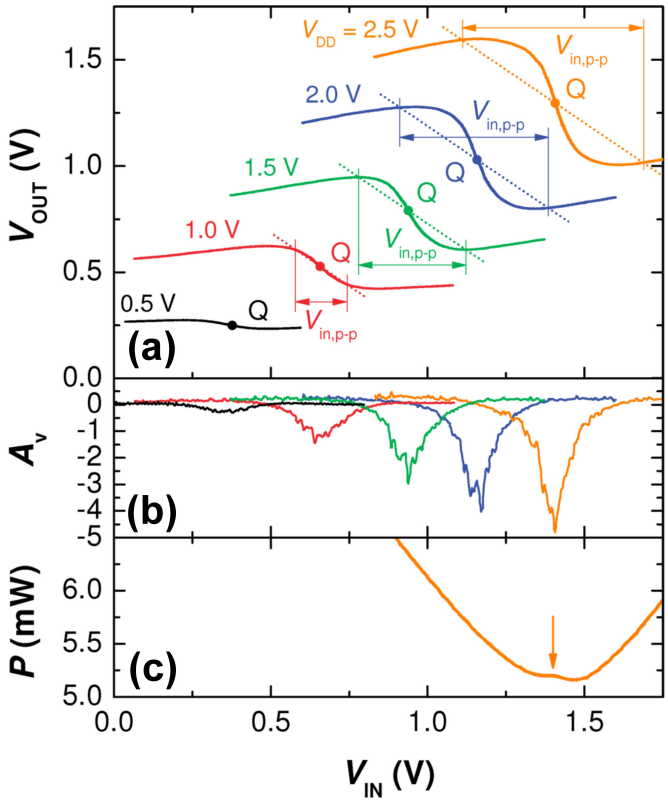
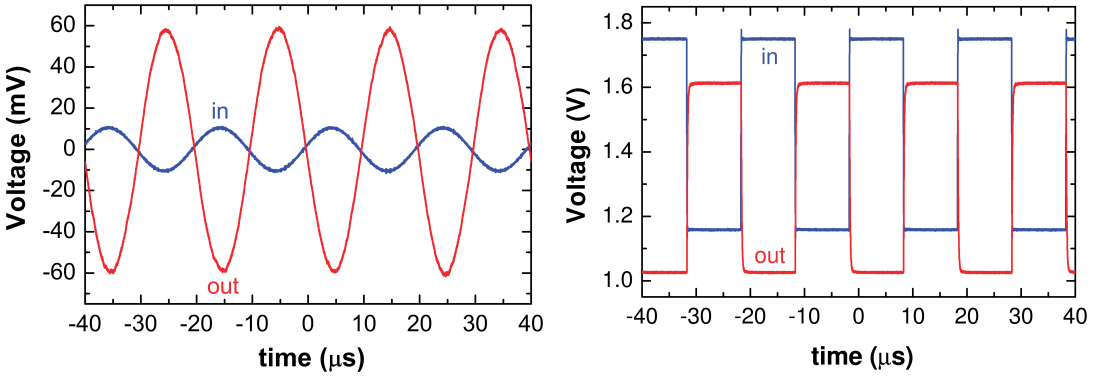


Figure 2.3.2: DC characteristics of large-area graphene inverters under ambient conditions at different supply voltages [30]. (a) transfer curves, (b) voltage gain and (c) power dissipation.

fabricated but the voltage gain was found to be independent from L as expected. With such a large-area graphene it was possible to obtain current ON/OFF ratios of up to to 3.4 (in the previously investigated amplifiers fabricated from mechanically exfoliated graphene ON/OFF < 2). Such large ON/OFF ratios and symmetry of the fabricated graphene transistors resulted in obtained high values of the voltage gain. It can be observed in figure 2.3.2 that at the highest gain point Q $V_{IN} \approx V_{OUT}$, which is a requirement for signal matching.

The voltage gain should be larger than one in the vicinity of the Q point in order to have the same voltage swing at the input and output. This can be understood from figure 2.3.2 where $V_{in,p-p}$ represents the voltage swing at which $V_{in,p-p} = V_{out,p-p}$ and is given by the intersection of the transfer curve with the unity-gain line passing through the Q point. The higher the supply voltage, the higher the voltage gain and larger the voltage swing.

Figure 2.3.2(c) shows the power dissipation of the graphene complementary inverter, biased at $V_{DD} = 2.5$ V, as a function of the input voltage. A local maximum is located at the highest gain point Q around which the power dissipation



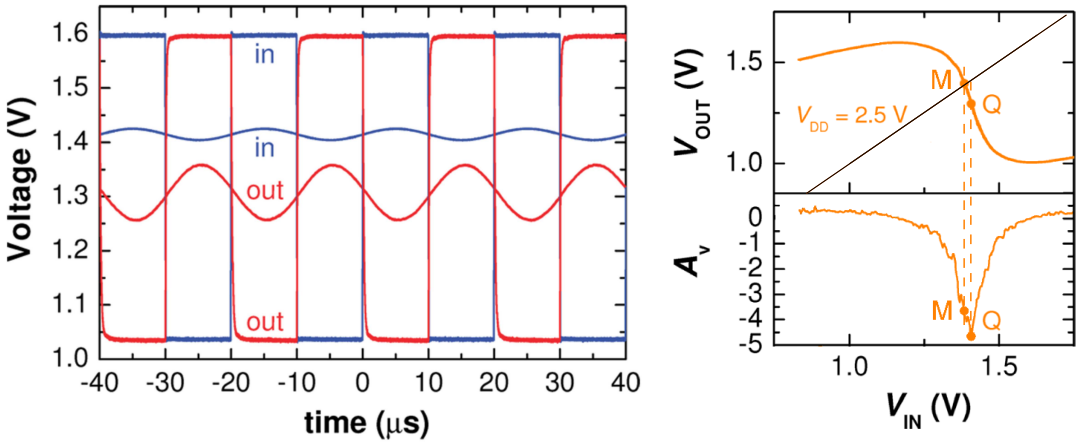
(a) (b)

Figure 2.3.3: (a) AC signal amplification with voltage gain $A_v = -5.3$ and (b) digital waveforms at the highest gain point Q, both measured under ambient conditions [30].

decreases. The power dissipated at the minimum is 98% of the power at the local maximum. Away of the Q point the power increases again since for these values of the V_{IN} the inverter exits the complementary operation between the two Dirac points. The dissipation could be reduced simply by lowering the supply voltage, but this will result in a smaller voltage gain and voltage swing. This is different from the Si inverters in which there is a maximum power that then reduces to zero on either side of the Q point due to the rail-to-rail operation,which eliminates the static power dissipation. For this reason graphene inverters should be used in applications not suitable for silicon devices such as transparent circuits on flexible substrates or, in case of digital applications, ultra-fast logic devices in which the power dissipation is not an issue.

Figure 2.3.3a shows the AC components of the input and output voltages when the graphene integrated complementary inverter is used as an analogue amplifier. It was possible to achieve a maximum voltage gain $|A_v|_{max} = 5.3$ at an input frequency $f = 50$ kHz and a supply voltage $V_{DD} = 2.5$ V. This is the highest gain reported for large-area monolayer graphene under ambient conditions, i.e. better than the results obtained with mechanically exfoliated graphene [14, 13]. The two signals shown in figure 2.3.3b are mismatched by $V_0 = 0.14$ V (in all fabricated devices a positive Dirac voltage was found which stems for p-doping by ambient impurities).

One of the fabricated graphene inverters was used as a digital logic gate and figure 2.3.3b shows the measured square-wave signals at an input frequency $f = 50$ kHz biased with $V_{DD} = 2.5$ V (at the highest gain point Q there is a small mismatch $V_0 = 0.11$ V). To compensate the mismatching is important that logic



(a) (b)

Figure 2.3.4: (a) Comparison between analogue waveforms measured at the highest gain point Q where there is no matching and digital waveforms at the DC operating point M where there is matching of input and output signal [30], (b) transfer curve and voltage gain together with the operating points Q and M. Everything was measured in air.

gates exhibit high voltage gain. If $|A_v|_{max}$ is large enough it is sufficient to operate inverters close to the highest gain point Q in the over-unity gain part of the transfer curve in order to have the input and output signals on the same level. Figure 2.3.4a shows the difference, on the same device at the same supply voltage $V_{DD} = 2.5$ V, between the operating points Q and M illustrated in figure 2.3.4b. The AC components, at a frequency $f = 50$ kHz, are biased at the Q point in which $A_v = -|A_v|_{max} = -4.7$ and there is no matching. The square-waves are biased at the operating point M in which the voltage gain is smaller $A_v \approx -3.5$ but there is the digital signal matching. This was the first demonstration of the matching between the input and output signals for a graphene digital logic gate under ambient conditions. The fact that the output and input signals take the same logic levels is important since this can allow the cascading of different inverters which is a prerequisite for the realization of realistic digital circuits.

At the matching point M the voltage gain is slightly smaller and therefore it did not affect very much the voltage swing of the input and output signals. The voltage swing at this point is $V_{in,p-p} = V_{out,p-p} = 0.56$ V while $V_{in,p-p} = V_{out,p-p} = 0.6$ V at the highest gain point Q, as illustrated in figure 2.3.3b. The voltage swing at the M point is 22 times larger than the thermal voltage $V_T = k_B T/e$ allowing unambiguously detection of the Boolean levels by the next logic gate. However graphene cannot compete directly with conventional silicon CMOS logic inverters which are capable of reaching a voltage swing of $\sim 100\%$ of

the supply voltage (compared to 22.4% for the fabricated graphene device which cannot be turned off in both logic states). Instead graphene logic gates could compete with emitter coupled logic (ECL) gates in which the voltage swing is only 15% of the supply voltage. The ECL is the fastest logic family and is made of over-driven transistors to reach ultra-fast operation at the expense of the static power dissipation which is similar to that of the graphene inverters. ECL gates are comprised of SiGe bipolar CMOS (BiCMOS) or InP heterojunction bipolar transistors (HBT). They are used for signal processing at frequencies above 100 GHz which cannot be reached by the standard CMOS technology. Replacing the typical materials that are used in the state of the art of ECL with graphene which has a higher mobility it would be possible to reach even higher frequencies. Moreover graphene logic gates are simpler to fabricate than ECL gates and are not limited by the trade-off between high speed and breakdown voltage which is a serious problem of HBT. However there are several problems that has to be resolved before graphene transistors could be used instead of HBTs in ECL logic gates [30]. First, the contact resistance in graphene transistors should be reduced below 10 $\Omega \cdot \mu$m to exploit the high intrinsic mobility of this material. Second over-unity voltage gain must be demonstrated at extremely high frequencies (the highest reported bandwidth for graphene amplifiers is 6 GHz [14]).

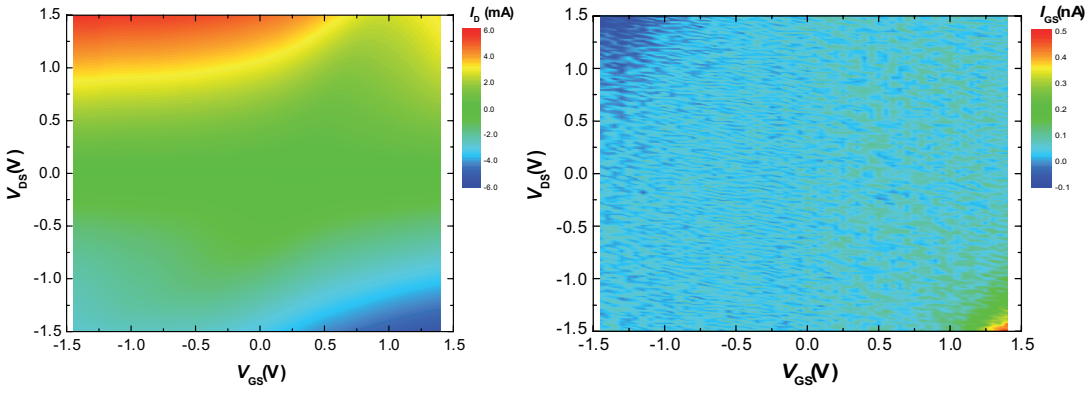
The matching of the input and output signals at room temperature was achieved due to the high values of the voltage gain stemming from several characteristics of the fabricated GFETs such as full-channel gating, good modulation of the carriers inside the channel, good mobility, low output conductance and manageable contact resistance typically 9 k$\Omega \cdot \mu$m.

The extrinsic carrier mobility was found from the transfer curves (the source-drain current vs. the gate voltage) of GFETs through transconductance g_m as

$$\mu = \frac{Lt_{ox}}{\varepsilon_{ox}\varepsilon_0}\frac{g_m}{V_{DS}W} \tag{2.3.4}$$

where V_{DS} is the source-drain voltage applied and the dielectric constant and the thickness of the oxide are $\varepsilon_{ox} = 5.65$ and $t_{ox} = 5$ nm [24]. In the fabricated GFETs were obtained good values of the transconductance $g_m \sim 3$ mS (150 μS/μm) resulting in typical device mobility at room temperature $\mu \sim 300\,\mathrm{cm^2V^{-1}s^{-1}}$, from which an intrinsic mobility $\mu_{intr} \sim 1000\,\mathrm{cm^2V^{-1}s^{-1}}$ was obtained extracting the contact resistance. This value is similar to the intrinsic mobility of typical top-gate FETs realized using mechanically exfoliated graphene [21].

The fabricated GFETs were showing a weak saturation behavior, as it can



(a) (b)

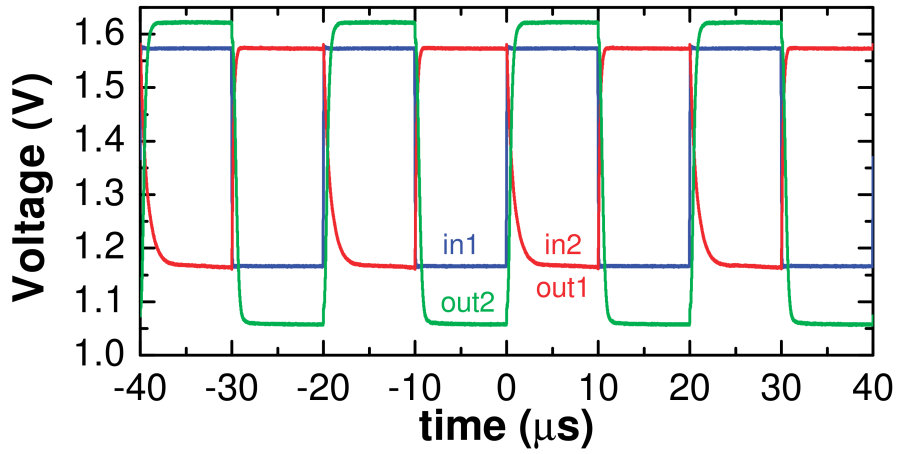
Figure 2.3.5: (a) Drain current I_D and (b) leakage current I_{GS} as a function of gate voltage V_{GS} and source-drain voltage V_{DS} of a typical fabricated GFET measured in ambient conditions.

be seen in figure 2.3.5a, which allowed to obtain relatively low values of the output conductance $g_d \sim 1$ mS (50 μS/μm) under ambient conditions. In terms of transconductance these results are comparable to the results obtained in transistors fabricated using mechanically exfoliated graphene deposited on exfoliated h-BN flakes which is not a scalable technology.

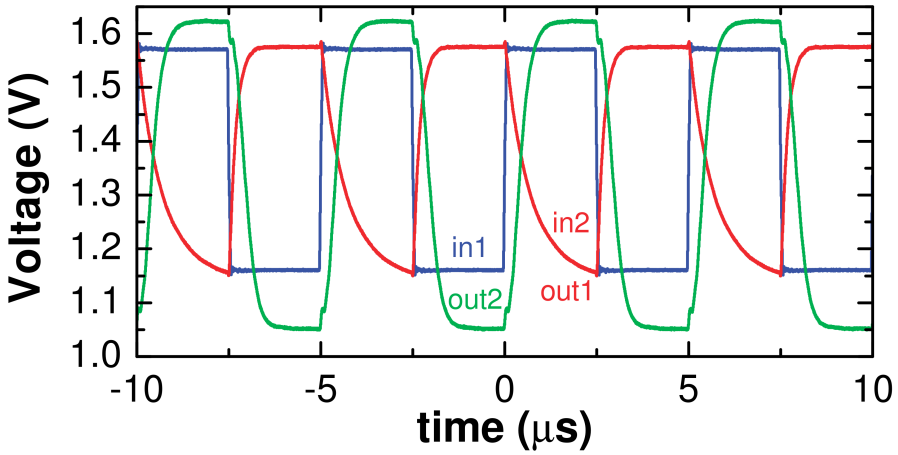
Figure 2.3.5b shows the typical leakage current that can be measured in fabricated GFETs. The highest values were obtained at the largest DC biasing, i.e., in the upper-left and bottom-right corners of figure 2.3.5b. However the GFETs in inverters were operated in the upper-right corner of the plot where there is no leakage through the gate oxide. This confirms that the obtained results were not influenced by the gate current which is 6 orders of magnitudes smaller than the drain current. Graphene devices were sensitive to variability induced by the fabrication but it was demonstrated that this does not influence logic operation as long as high values of voltage gain are preserved. Higher gains in analogue and digital applications can be achieved in the future increasing the transconductance g_m, reducing the output conductance g_d and lowering the contact resistances.

2.3.2 Cascading of graphene logic gates

The fact that it was possible to demonstrate the matching between the logic levels of the input and output signals at room temperature does not imply that it is also possible to cascade graphene logic gates in realistic applications. In fact, due to fabrication issues, it is not possible to realize graphene inverters with the same transfer characteristics and this has an effect on the highest gain operating



(a)



(b)

Figure 2.3.6: Square-wave signals of two large-area graphene inverters in cascade connection measured under ambient conditions at frequencies (a) $f = 50$ kHz and (b) $f = 200$ kHz [30].

points Q_1 and Q_2 of the two inverters which will not be the same. Due to this it is not possible to bias two inverters such that both operate at the same time at their highest gain points. It was found that mismatch between the Q points can be compensated if both inverters exhibit $|A_v| > 3$.

Figure 2.3.6a demonstrates the successful cascading of two inverters under ambient conditions at a supply voltage $V_{DD} = 2.5$ V and frequency $f = 50$ kHz. This was obtained at the DC operating point that lies in between the highest gain points of the two inverters because both inverters exhibit, $|A_v| > 4$ at their Q points. This allows to maintain the over-unity voltage gain in between the two Q points and as a consequence to realize the cascade connection, with all logic states at the correct levels. The final logic states 0 and 1 at the output of

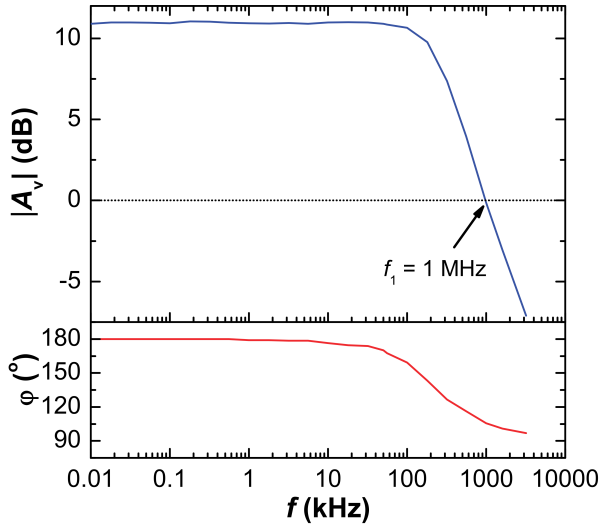


Figure 2.3.7: Frequency response of the magnitude of the voltage gain and phase shift between input and output signals of a typical fabricated graphene inverter measured in air [30].

the second inverter (green plot in figure 2.3.6a) are correctly interpreted because the corresponding voltage levels span the larger range than at the input. It can be noticed that it was possible to achieve a perfect matching at the first stage while there is a larger voltage swing at the second stage. This is due to the fact that, in graphene inverters the saturation voltage levels are not well defined as in conventional CMOS inverters. However, this does not influence the operation of the cascaded graphene logic gates since the high voltage gain does not allow the creation of intermediate levels during the signal propagation.

Using the same supply voltage but increasing the frequency of the square-wave signal to 200 kHz it can be seen in figure 2.3.6b that the digital signals start to deteriorate. For this reason the frequency response of one of the fabricated wafer-scale graphene inverters at $V_{DD} = 2.5$ V was studied (figure 2.3.7). The unity-gain frequency is $f_1 = 1$ MHz and the bandwidth of the device is mainly limited by the output resistance $r_d \approx 1\,\text{k}\Omega$ and parasitic capacitance of the cable $C_c \approx 0.6\,nF$ to $f_{-3dB} = 1/(2\pi r_d C_c) \approx 270$ kHz. The same parameters also influenced the bandwidth of the graphene voltage amplifiers realized with mechanically exfoliated graphene (section 2.2.2). With wafer-scale graphene inverters it was possible to achieve a larger value of f_{-3dB} due to the lower output resistance (because of the smaller L/W ratio) since the parasitic capacitances of the cable were almost the same.

Cascading of graphene logic gates has never been demonstrated so far at any temperature. Here it was demonstrated that such operation is possible, i.e., that

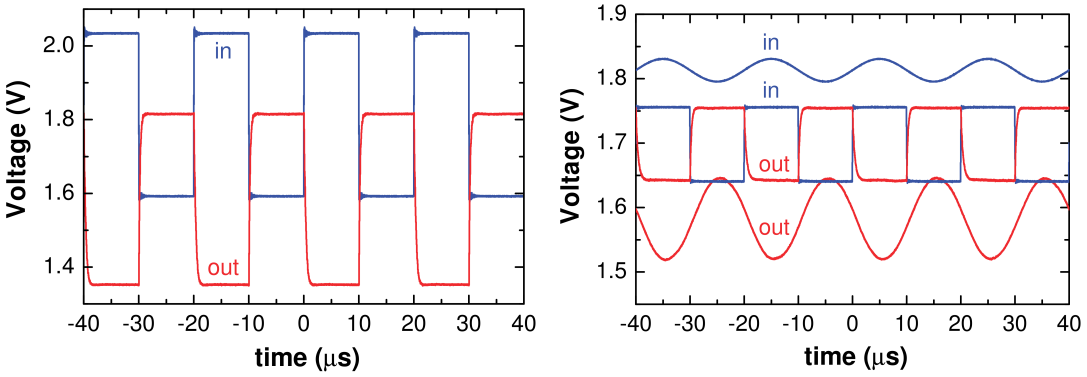


Figure 2.3.8: (a) Square-wave signals in devices with large Dirac voltages measured under ambient conditions (a) at the point Q and (b) at the matching operating point. The sine-wave signals at the the point Q are also shown [30].

it was possible with one stage to trigger the next stage under ambient conditions. This result paves the way for the realization of realistic graphene digital circuits in which static power dissipation is not an issue.

2.3.3 Large Dirac voltages

As it was said before the matching of the input and output signals of the fabricated graphene inverters depends not only on the voltage gain but also on the position of the Dirac points of the GFETs. In the previous section it was demonstrated that it was possible to match the digital signals and to cascade two inverters comprised of GFETs with charge neutrality points close to zero, i.e., $V_0 < 0.2$ V. In this section inverters with large Dirac voltages ($V_0 > 0.2$ V) are investigated.

Figure 2.3.8a shows digital waveforms in a graphene inverter biased at $V_{DD} = 2.5$ V at an input frequency $f = 50$ kHz, and $V_0 = 0.23$ V measured at the highest gain point Q in air. At this operating point the voltage swing is $V_{in,p-p} = V_{out,p-p} = 0.5$ V but there is no signal matching since the Dirac voltage is too large. As it was illustrated in figure 2.3.4b at the point M in order match the signals. Since the point M is not very close to point Q the signal matching can be realized only at the expense of the loss of the voltage gain, as shown in figure 2.3.8b. At the point Q $A_v = -3.5$ which can be calculated from the sine-wave the input and output signals shown in figure 2.3.8b. Comparing the square-wave signals of the inverter operating at the points Q and M a reduction in the voltage swing can be noticed, $V_{in,p-p} = V_{out,p-p} = 0.12$ V. In such a condition it is possible to have the signal matching in air but the voltage swing is suppressed.

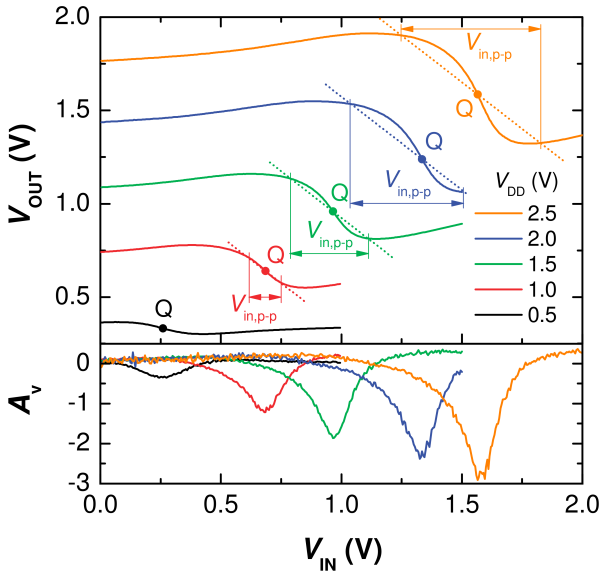
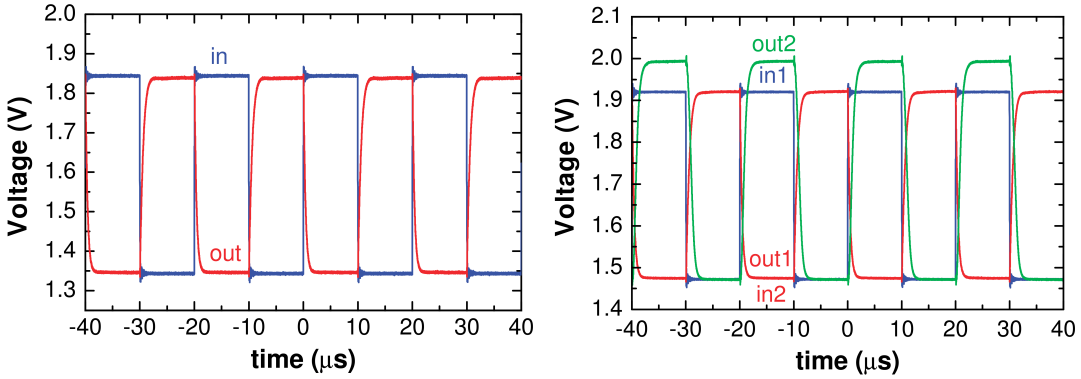


Figure 2.3.9: Vacuum transfer curves and voltage gain of the inverter with GFETs exhibiting large Dirac voltages in air [30].

With such samples it is necessary to shift the Dirac point to zero to remove the signal mismatch at the highest gain point and obtain sufficiently large voltage swing. It was find out that this can be done by keeping the devices in vacuum at a pressure of 10^{-5} mbar for approximately 1 day before starting the measurements. Even though the graphene channel is fully covered, desorption of impurities in vacuum takes place due to the diffusion of molecules at the interface between graphene and aluminum. After 1 day in high vacuum the charge neutrality point reaches approximately $V_0 = 0$. The Dirac point will stay at zero even if the pressure is afterward increased to 200 mbar.

Figure 2.3.9 shows the DC transfer curve of the previously investigated device ($V_0 = 0.23$ V) at different supply voltages but this time at a pressure $p = 200$ mbar. As it can be seen at the Q point the input and output signals are equal in vacuum. However, it was also found that the voltage reduces by 18% in vacuum; i.e., it reduced from ~ 3.5 to approximately ~ 2.9 at $V_{DD} = 2.5$ V. This could be explained by different shift of the Dirac points of the two GFETs in the inverter.

The matching between the digital signals at low pressure is illustrated in figure 2.3.10a where $V_{DD} = 2.5$ V and $f = 50$ kHz. It can also be noticed that the voltage swing is now large enough to allow to distinguish between the logic levels ($V_{in,p-p} = V_{out,p-p} = 0.5$ V) since the voltage gain is only slightly reduced in vacuum. Moreover the highest gain points of two different inverters are close enough which allows to cascade them at low pressure, shown in figure 2.3.10b ($V_{DD} = 2.5$ V). This demonstrates that it is possible not only to match and



(a) (b)

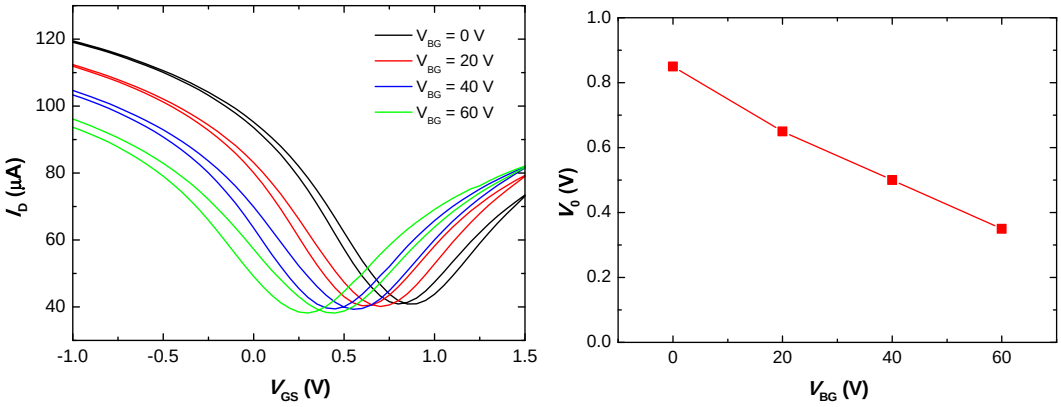
Figure 2.3.10: (a) Digital signal matching and (b) cascading of graphene logic gates performed at low pressure [30].

cascade different graphene inverters with small Dirac voltages but also with large Dirac voltages. The only difference is that in the first case this was obtained under ambient conditions while in the second case at reduced pressure.

2.4 Graphene ultra-high speed circuits

All previously fabricated devices pave the way for the realization of realistic graphene digital circuits. Due to the over-unity voltage gain in graphene inverters it was possible to realize graphene digital logic gates capable of matching the input and output signals which led to cascading of different stages, that is the main prerequisite for digital electronics. Moreover it has to be underlined that all these results were obtained under ambient conditions which is important in realistic scenarios. But as it was described previously graphene digital inverters cannot compete directly with Si CMOS logic gates due to static power dissipation issues.

Recently digital systems operating at extremely high frequency (EHF), typically $f > 100$ GHz, have become relevant due to the rapid progress of wireless, fiber optics and space communications which need signal processing at ultra-high data transfer rates (typically >100 Gbit/s) [12]. A special class of digital circuits based on III-V HBTs has been developed in order to perform data conversion at the transmitting or receiving side of a serial EHF lines. After data conversion the signals can be processed at smaller clock rates by a low power conventional silicon CMOS logic. Graphene could emerge as a contender in EHF electronics due to its large charge carrier mobility and probably replace InP which currently dominates the high-speed electronics. For instance, the fastest ECL gates have similar



(a) (b)

Figure 2.4.1: (a) Drain current as a function of the top-gate voltage for different back-gate voltages of a single GFET biased with $V_D = 50$ mV, (b) the linear shift of the Dirac voltage V_0 as a function of the back-gate voltage.

power dissipation and voltage swing as the fabricated graphene logic gates. However several improvements have to be made in order to replace InP with graphene in EHF electronics.

2.4.1 Oxide capacitance

In order to fabricated of high-frequency grahene logic circuits it is necessary to reduce as much as possible all parasitics components and increase as much as possible the control of the charge carriers inside the channel of the GFETs. High of the transconductance should also lead to high voltage gain. With that purpose I spent three months at the University of Illinois at Urbana-Champaign to study the oxide capacitance of the thin aluminum oxide layer used in our top-gated GFETs.

The top-gate oxide capacitance of the AlO_x/Al gate stack can be found from the transfer curves of the GFETs measured at the different values of the back-gate voltage V_{BG}, as shown in figure 2.4.1a. It can be seen that the Dirac voltage shifts to lower values of V_{GS} by increasing V_{BG}. In particular the Dirac voltage V_0 exhibits a linear dependence on the back-gate voltage as shown in figure 2.4.1b. From the slope of the linear plot and the capacitance equation $Q = CV$, the top-gate capacitance is

$$C_G = C_{ox} = -C_{BG}\frac{\triangle V_{BG}}{\triangle V_0} \tag{2.4.1}$$

which gives $C_G/C_{BG} = 121.2$. The fabricated device has a back-gate oxide made of 300 nm of SiO_2 with a dielectric constant $\varepsilon_{r,SiO_2} = 3.9$ which leads to a back-

gate capacitance per unit area $C_{BG} = \varepsilon_0 \varepsilon_{r,SiO_2}/t_{SiO_2} = 11.5\text{nF/cm}^2$. The top-gate capacitance per unit area is then $C_{ox} = C_G = 121.2 C_{BG} = 1.39\,\mu\text{F/cm}^2$.

The measurements of the capacitance were performed at AC frequencies

$$C = \frac{I}{2\pi f V_{AC}} \tag{2.4.2}$$

where I is the amplitude of the AC current passing through the structure, f is frequency of the AC signal and V_{AC} is the magnitude of the measured AC voltage. In practice the capacitance is obtained from the AC impedance of the device by applying an AC voltage and measuring the resulting AC current and voltage at the impedance. Such measurements take into account series or parallel resistances associated to the capacitance.

Usually the frequency of the AC signal is in the between 10 kHz and 10 MHz and the DC voltage is used to drive the MOS structure from the accumulation region to the inversion passing from the depletion region. In case of GFETs the results of the CV measurements are different since graphene is a semi-metal and it undergoes an ambipolar transition without the depletion of the carriers.

In order to perform the measurements correctly it is important to calibrate the CV meter in order to de-embed capacitances of the cables, probes, and chuck of the probe station. This was done by measuring an open and short circuit. Then the back of the chip is connected to ground together with the source and drain contacts and the bias is applied at the gate electrode. In this way, the gate capacitance of the GFET was obtained, as illustrated in figure 2.4.2, using an AC signal at a frequency of 200 kHz with an amplitude of 0.1 V in parallel configuration. In fact at this frequency the reactance of the capacitor was relatively large (>10 kΩ) and therefore the parallel resistance becomes more significant than the series resistance. The gate capacitance is given by

$$C_G = \frac{C_q C_{ox}}{C_q + C_{ox}} \tag{2.4.3}$$

and it represents the series of the oxide capacitance C_{ox} of the top-gate and the quantum capacitance of graphene C_q (section 1.3.3). The value of the graphene quantum capacitance is low and comparable to the capacitance of a thin gate oxide close to the Dirac point. However, far away from the Dirac point it assumes very large values. For this reason the quantum capacitance affects the gate capacitance only in the proximity of the Dirac point, while away from it the gate capacitance is dominated only by the oxide capacitance C_{ox}. The oxide capacitance shown in figure 1.3.12 is obtained by dividing the measured gate capacitance C_G by the

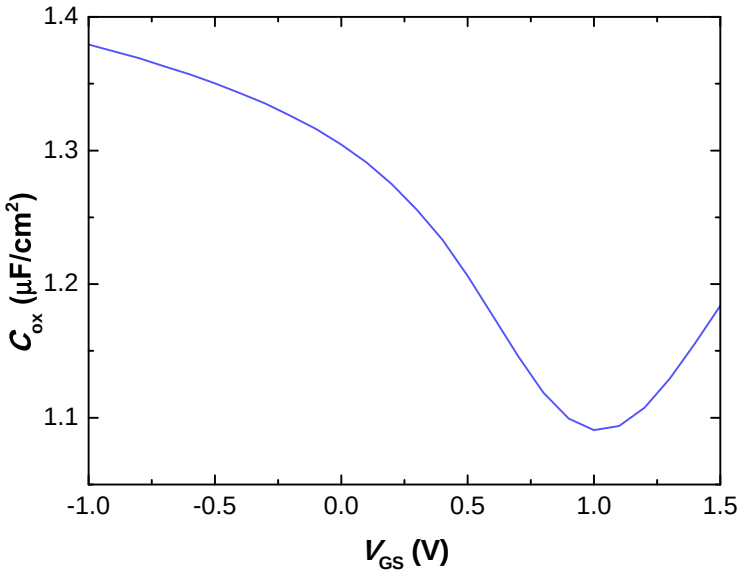


Figure 2.4.2: CV measurements performed on a graphene transistors with a very thin aluminum oxide gate electrode.

product of the channel width W and length L. Its value is accurate only far away from the Dirac point where the influence of C_q is negligible. The oxide capacitance obtained in this way is approximately 1.4 μF/cm^2 which is very close to the value obtained by the Dirac shift voltage method (figure 2.4.1a and 2.4.1b).

The obtained result can be compared to the typical value of the oxide capacitance of the standard 65 nm Si MOSFET in which the capacitance per unit area is $C_{ox} = 1.77\ \mu$F/cm^2 [26]. The very thin aluminum oxide layer allows to reach the value of the oxide capacitance of the Si technology and exceed the typical oxide capacitance reported for graphene transistors [26]. Therefore this type of gate represents a good choice for the realization of realist graphene digital circuits since it allows a good control over the carriers in the channel.

2.4.2 Graphene ring oscillators

Even though of the cut-off frequencies f_T above 100 GHz [8] have been reported for a single GFET there are still no demonstrations of high-speed graphene digital circuits, mainly because f_T is just a measure of the internal transistor delays rather than delays in realistic digital circuits. The most important circuits in digital electronics are the ring oscillators (ROs) which are made by connecting an odd number of inverters so that the output of one inverter is the input of the next inverter, figure 2.4.3a [12]. This creates a loop which makes the RO unstable and leads to oscillations at high frequency if the inverters are identical.

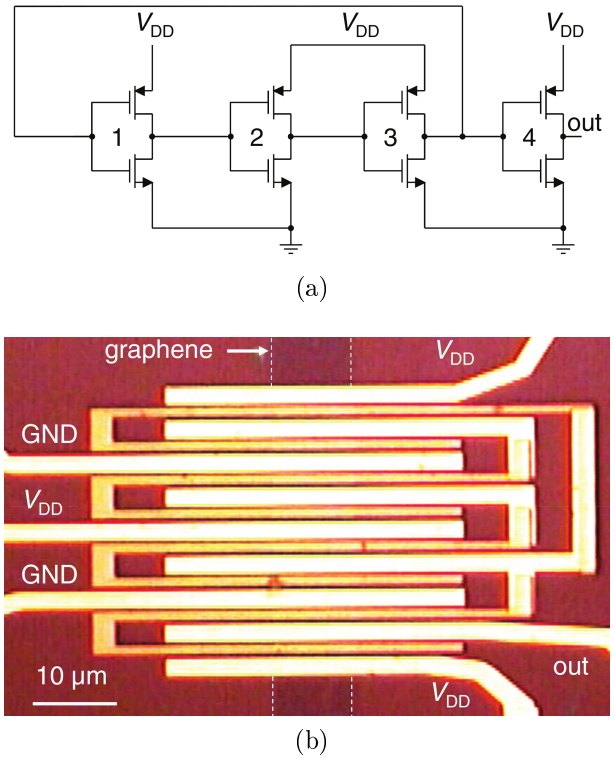


Figure 2.4.3: (a) Circuit diagram and correspond (b) optical microscope image of one of the fabricated ring oscillators with channel length $L = 1\ \mu$m [12].

In particular each inverter must exhibit an over-unity voltage gain and be able to match the input and output signals. Also the on-state resistance of the GFETs in an inverter should be as low as possible to allow a fast charge and discharge of the gate capacitance of the next stage. The oscillation frequency f_O in ROs is smaller than the cut-off frequency f_T since it represents a direct measure of the delay in realistic scenarios. For this reason ROs represent the standard electronic circuits used to test the ultimate performances (in terms of frequency) of the digital logic families.

In the past ROs have been fabricated with other two-dimensional materials such as carbon nanotubes (CNTs) [1, 7] and exfoliated MoS_2 [42] but not with graphene. The oscillation frequency in CNTs and MoS_2 was limited by large values of the on-state resistances to $f_O = 52$ MHz (CNT) and $f_O = 1.6$ MHz (MoS_2). Graphene has the possibility to reach higher frequencies due to its larger mobility which results in small on resistance and it can also be fabricated by simpler methods on a wafer scale.

The circuit schematic of one of the fabricated integrated graphene ring oscillators is shown in figure 2.4.3a [12]. The first three inverters are cascaded in a loop, forming the RO. The output of the third inverter is connected to the input

of the fourth inverter, which is a of buffer stage used to decouple the RO from the measurement equipment connected to the output. Differently from conventional integrated circuits, in which there is only a single line for the ground and the supply voltage, here there are two ground and three V_{DD} lines. This was mainly done in order to avoid an additional lithographic step (for the realization of the overlap area between the DC lines) which could introduce additional contamination.

The ROs were fabricated starting from graphene monolayers grown by chemical vapor deposition on Cu foils, using CH_4 as a gas precursor, and then transferred to SiO_2/Si substrates with the metalized back which was used it as a global back-gate. Graphene stripes with a width $W = 10\,\mu$m were defined by (EBL) and reactive ion etching, followed by the fabrication of gate contacts by EBL. The top-gates were created by evaporating 45 nm of Al, 2 nm of Ti and 13 nm of Au. Gold termination was used to realize a good ohmic contacts between the gates of one stage and the source and drain contacts of the next stage. This is visible on the right-hand side of the image in figure 2.4.3b. Without gold termination aluminum will oxidize also on the top side and prevent a good contact between the inverters in the RO. The internal (as opposed to external) connections between the inverters reduce significantly the parasitic capacitances and therefore increase the oscillation frequency of the circuit. The evaporation of the gates was made under tilt so that the evaporated material forms smooth slopes instead of abrupt edges. This helps the formation of good ohmic contacts since it increases the contact area between the electrodes of different inverters. In the last step source and drain contacts were patterned. They consist of 75 nm of Au in order to decrease the contact resistance and increase the voltage gain. In this way the contact resistance was reduced to approximately 2 $k\Omega \cdot \mu$m. As it can be seen in figure 2.4.3b this time it was not possible to fully cover the graphene channel since the top surface of the gates is terminated with a conductive layer, and that introduced unwanted access resistances which reduce the voltage gain. Graphene ROs were realized using graphene transistors with different channel lengths, 1 μm, 2 μm and 3μm, keeping constant the access lengths $L_a = 0.5\,\mu$m.

As demonstrated in the previous sections, the complementary operation of the inverters is obtained between the Dirac points of the GFETs in the inverters when a supply voltage $V_{DD} > 0$ is applied. All fabricated GFETs exhibited $V_0 > 0.2$ V under ambient conditions, stemming for p-doping of the channel. Non-zero Dirac voltages imply mismatch between the input and output signals at the highest gain point of the inverter which, as a consequence, reduce the voltage swing in multistage graphene digital circuits. For this reason a positive back-gate voltage

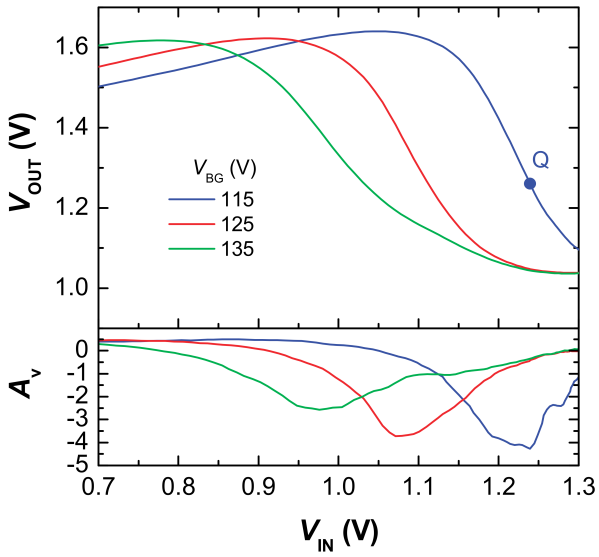


Figure 2.4.4: Output voltage and voltage gain in function of the input voltage of the buffer inverter [12].

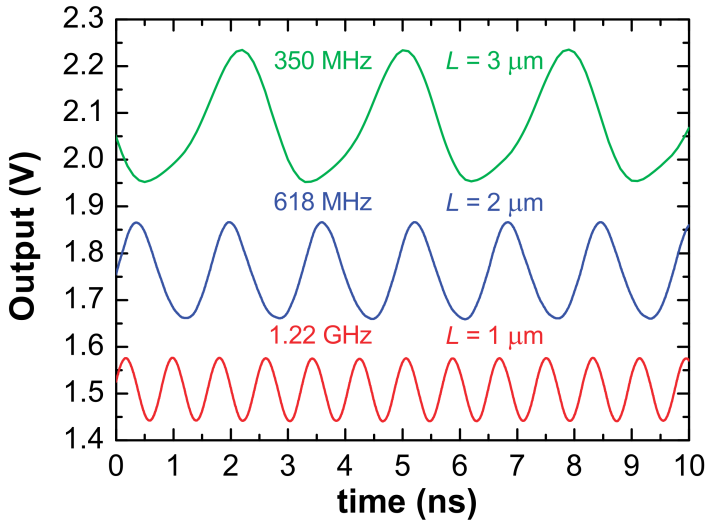
(for $V_0 > 0$) has to be applied in order to shift the Dirac point back to zero and match the signals of different inverters. ROs with three stages were chosen to realize since they oscillate at the highest frequency. However, they also require the highest voltage gain ($|A_v| \geq 2$).

In order to check whether inverters in fabricated ROs satisfy this requirement it was not necessary to fabricate separate inverters since the gain can be measured on the last stage which is not a part of the RO. The input of the fourth inverter can be accessed by supplying voltage V_{DD1} to the inverters that form the ring, which is different from the supply V_{DD} of this inverter. Due to the circuit symmetry the signal that arrives at the top-gate of the last inverter is $V_{IN} \approx V_{DD1}/2$, and the transfer curve can be obtained varying V_{DD1}. Figure 2.4.4 shows the measured transfer curve and the voltage gain of the buffer stage biased with $V_{DD} = 2.5$ V. This plot also demonstrates that the back-gate voltage is capable of shifting the Dirac point to zero achieving the signal matching and voltage gain $|A_v| > 4$ at the point Q. Therefore the fabricated ROs satisfies conditions for oscillation.

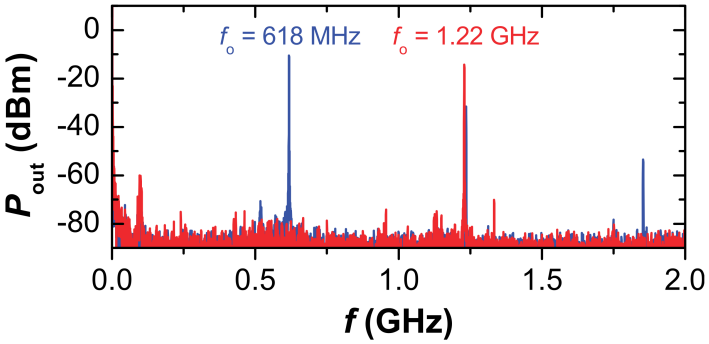
The oscillation frequency of ROs depends on the gate delays of the inverters in the loop as

$$\frac{1}{f_O} = 2\sum_{i=1}^{n} \tau_i \tag{2.4.4}$$

where τ_i is the gate delay of the i-th inverter. Assuming that all transistors are identical with the same rise and fall delays τ, the expression for the oscillation frequency of a RO with identical inverters (without the buffering stage) simplifies



(a)



(b)

Figure 2.4.5: (a) Output signals and (b) power spectra of the best performing fabricated buffered ring oscillators showing the best performances, measured for different gate lengths at the same supply voltage $V_{DD} = 2.5$ V. The second harmonic of the medium RO is located exactly in the position of the first harmonic of the small RO [12].

to

$$f_O = \frac{1}{2\pi\tau} = f_{O,max}. \tag{2.4.5}$$

Using a model in which the inverters do not reach the steady state before they are triggered again, the gate delay τ can be expressed as

$$\tau \propto CG_D^{-1} \tag{2.4.6}$$

where G_D is the sum of the extrinsic drain conductances of the transistors that form the inverter and C is the parasitic capacitance that loads the inverter. This load comes mainly from the gate capacitance C_G of the next stage, i.e., $C \approx 3C_G$, where $C_G \approx WLC_{ox}$.

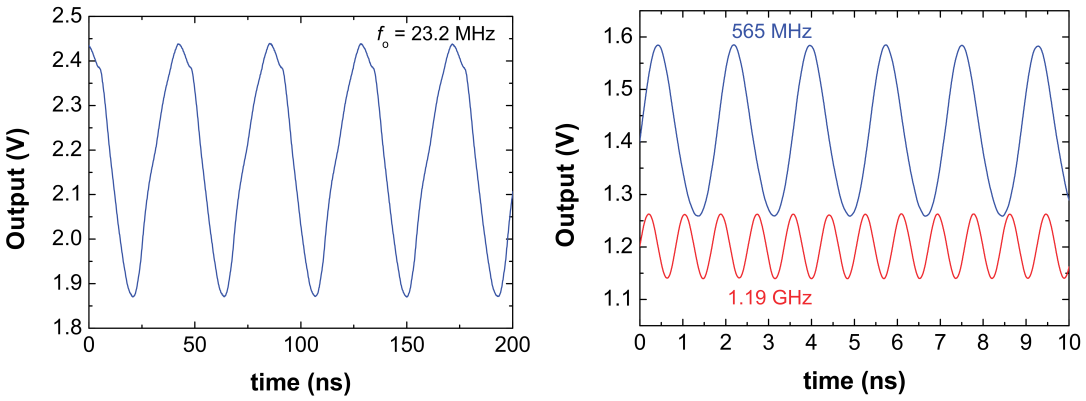
The fabricated ROs have a buffer stage which introduces an additional capacitive load to the last inverter of the ring which has a fan-out $N = 2$ (FO2) which reduces the oscillation frequency f_O. In order to minimize the parasitic capacitive load of the buffer stage the output was connected to the oscilloscope via an active probe with a capacitance of 0.8 pF (bandwidth of 4 GHz). The measurement performed under ambient conditions on different ring oscillators with the buffering stage are shown in figure 2.4.5. This is the first demonstration of graphene digital integrated circuits operating at gigahertz frequencies. All output signals in this figure were obtained by applying a supply voltage $V_{DD} = 3.5$ V. During the measurements the back-gate voltage was used to shift the Dirac points of the logic gates to zero so that the oscillations can start. In particular it was measured in large ROs ($L = 3\,\mu$m and $W = 20\,\mu$m) a maximum oscillation frequency $f_O = 350$ MHz and a voltage swing $V_{p-p} = 0.284$ V at $V_{BG} = 34$ V, in medium ROs ($L = 2\,\mu$m and $W = 10\,\mu$m) $f_O = 618$ MHz and $V_{p-p} = 0.208$ V at $V_{BG} = 5$ V and in small ROs ($L = 1\,\mu$m and $W = 10\,\mu$m) $f_O = 1.22$ GHz and $V_{p-p} = 0.136$ V at $V_{BG} = 50$ V.

Figure 2.4.5b shows the respective power spectra of the small and medium ROs. Many graphene ring oscillators of different dimensions were fabricated and the oscillation frequencies were in the range $284\,\text{MHz} < f_O < 350\,\text{MHz}$ for large ROs, $504\,\text{MHz} < f_O < 750\,\text{MHz}$ for medium ROs and $1\,\text{GHz} < f_O < 1.28\,\text{GHz}$ for the small ROs. The variation of f_O is due to small differences of the overlap area of the internal connections between inputs and outputs of the inverters which influences f_O through the total drain conductance G_D.

As the channel length of the GFETs the ROs is reduced the capacitances and resistances of the channel are reduced and as a consequence f_O is increased. Knowing the value of the oxide capacitance C_{ox} and the dimensions of the transistor (W and L) it is possible to find the gate capacitances which influence the gate delay. They are $C_G = 0.84$ pF, $C_G = 0.28$ pF and $C_G = 0.14$ pF for large, medium and small ROs respectively. However while f_O is increasing by decreasing the channel length L the amplitude of the output signal is decreasing. This is mainly due to the signal filtering by the low-pass filter present at the output of the buffer stage. This last stage is loaded by the capacitance of the measurements equipment C_L that can be expressed as

$$C_L = C_A + C_{OUT} \tag{2.4.7}$$

where C_A is the capacitance of the active probe and C_{OUT} is the parasitic capacitance of the on-chip pad, connected to the output of the last stage. It was found



(a) (b)

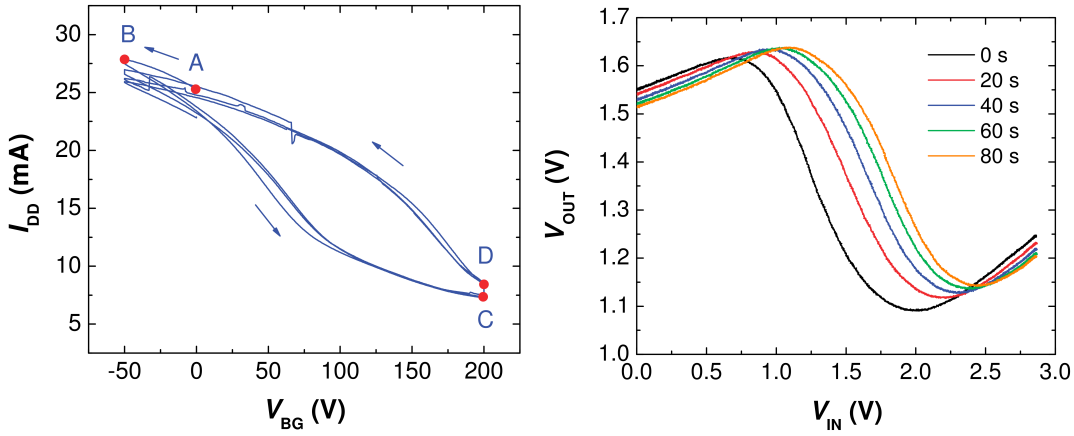
Figure 2.4.6: (a) Output signal of unbuffered large RO and (b) buffered medium and small ROs at the same supply voltage showing the highest voltage swings measured [12].

$C_L = 20$ pF. The bandwidth of the buffer inverter is then

$$f_{-3dB} = \frac{1}{2\pi \left(R_{ON} \parallel R_{OFF}\right) C_L} \sim 30\,\text{MHz}. \tag{2.4.8}$$

Unbuffered large ROs (comprised of only three inverters) were also fabricated and measured to evaluate the effect of the measurement equipment on the oscillation frequency. The output signal in this case is shown in figure 2.4.6a at $V_{DD} = 3.5$ V and $V_{BG} = 160$ V. Typical oscillation frequencies were in the range $17\,\text{MHz} < f_O < 25\,\text{MHz}$ which is well below the range of large buffered ROs. This is due to the fact that the last inverter of a RO is not loaded by the gate capacitance of the buffer stage, $C_L \approx 3C_G = 2.52$ pF, but by the much larger capacitance $C_L = 20$ pF, given by equation (2.4.7). However, as the oscillation frequency is very close to the bandwidth, the voltage swing is larger than in buffered ROs, $V_{p-p} = 0.57$ V (see figure 2.4.6a). In this case the voltage swing is 16.2% of the supply voltage which is smaller than the result obtained in section 2.3.1 but still larger than the swing in ECL gates. By comparison the highest voltage swing was reduced to 12.2% and 5% of V_{DD} in medium and small ROs respectively (2.4.6b). For this reason it is important to minimize the parasitic components of the measurement equipment in order to obtain signals of reasonably large voltage swings.

All ring oscillators reported in figure 2.4.5a were operated under ambient conditions at back-gate voltages which means that the single GFETs exhibited small Dirac voltages. It was also found that ROs with $V_{BG} < 100$ V exhibited a



(a) (b)

Figure 2.4.7: (a) Total current (sum of the currents of four inverters) as a function of the back gate voltage of a RO with high Dirac voltages and (b) drift of the transfer curves of the corresponding buffer inverter [12].

long-term stability. This was not the case in samples with large Dirac voltages in which it was not possible to keep the Dirac point at zero voltage with a constant V_{BG}. Figure 2.4.7a shows a typical back-gated transfer curve recorded during the measurements on a ring oscillator with a large Dirac voltage. The measurements started by applying $V_{BG} = 0$ (point A). Then the back-gate voltage was swept to point B ($V_{BG} = -50$ V) and then to point C ($V_{BG} = 200$ V) were the oscillations started. During the oscillation the total current of the four inverters increased to point D detuning the RO due to the fact that the Dirac points moved to higher values. This demonstrates the short-term stability of the ROs with large Dirac voltages. In order to restore the oscillation the entire cycle was repeated. The shift of the Dirac voltage can be observed in figure 2.4.7b where transfer curves of the buffer inverter where measured at different intervals of time. Higher the back-gate voltage required for the oscillation shorter the stability of the RO (faster the shift of the transfer curve). This shift introduces the mismatch between the input and output signals of the inverters which stops the oscillation after a certain time.

All measurements were performed at room temperature and the samples requiring low back-gate voltages exhibited long-term stability. The devices with large Dirac voltages (operated at $V_{BG} > 100$ V) required reduction in exposure to air in order to achieve an acceptable level of stability. This could be done by measuring the ROs in vacuum since under such conditions Dirac points of the GFETs shift to zero, as demonstrated in section 2.3.3. Howe very this was not attempted because the vacuum is not a standard operating environment of the electronic circuits. In order to increase the stability of the digital circuits

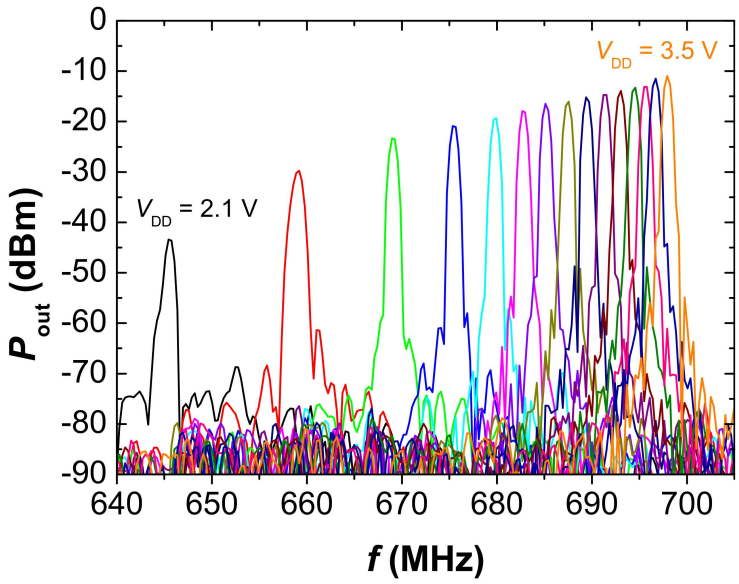


Figure 2.4.8: Power spectrum of the output signal of a medium ring oscillator at different values of the supply voltages[12].

measurements were performed at room temperature under a N_2 flux.

In conventional ROs the supply voltage V_{DD} strongly influences the oscillation frequency f_O and this can represent a serious problem in complex digital circuits. In fact a large number of transistors and increased power consumption place increased demand on the voltage supply causing it to fluctuate. This have negative effect on the operation of digital logic gates and also reduce their noise performance [12]. Figure 2.4.8 shows the power spectrum of the output signal of a medium buffered graphene RO at supply voltages in the range from 2.1 V to 3.5 V causing a variation of f_O from 645 MHz to 700 MHz. V_{DD} did not change only the oscillation frequency but also the voltage swing which is higher at higher supply voltages due to larger voltage gain of the inverters. The increase of the oscillation frequency is due to increase of the total drain conductance of the GFETs at higher V_{DD} since $f_O \propto G_D$. It was found that f_O has an average variation with the supply voltage of approximately $5.6\% f_0/V$. The fabricated graphene ROs are not so sensitive to the fluctuations of the supply voltage as Si CMOS ROs. These fluctuations with the supply voltage are 7, 21, and 51 time larger in case of Si, CNTs and MoS_2 respectively. The insensitivity to power supply fluctuation is important in applications where the frequency stability is required such as for clock generation in high-speed digital electronics.

The obtained results demonstrate the first wafer-scale graphene ROs operating at room temperature under ambient conditions. The fabricated devices oscillated at the highest reported frequency for any low-dimensional transistor material with

voltage swing exceeding that of InP ECL gates, the fastest logic family. Although this is an important step in the applications of graphene in digital electronics, it should be noted that GFETs cannot compete with silicon logic FETS in terms of static power dissipation. For instance, $I_D/W \approx 100\,\mathrm{nA}/\mu\mathrm{m}$ at $V_{DD} = 0.75$ V in 22 nm [17] while $I_D/W \approx 270\,\mu\mathrm{A}/\mu\mathrm{m}$ at $V_{DD} = 2.5$ V in the fabricated GFETs where I_D is the static drain current. The only advantage of graphene ROs with respect Si CMOS ROs is represented by the smaller sensitivity to the variations of the supply voltage, even though this is a consequence of the reduced voltage swing.

2.4.3 Stand-alone graphene frequency mixers

The use of graphene in electronics has been limited to niche applications such as frequency mixing for which an over-unity gain is not required. However without the possibility to amplify signals it is not possible to generate oscillating signals. For that reason the graphene mixers which have been realized so far used an external local oscillator (LO) in order to make the frequency conversion. The fabricated graphene ROs could be used in analogue applications for eliminating a need for an external LO. In this way graphene ROs could be used in the modulation and generation of oscillating signals forming a stand-alone graphene frequency mixers.

The result of super-imposing RF signal can be seen in figure 2.4.9 showing the power spectrum of the output signal at a supply voltage $V_{DD} = 2.5$ V with $V_{BG} = 166$ V where the signal frequencies are $f_{RF} = 25$ MHz and $f_{LO} = 292$ MHz. As can be seen in the figure, there is the contribution of the signal of the RO at the frequency f_{LO} and the product between the RF and LO signals that results in the intermediate frequencies $f_{LO} \pm f_{RF}$. The device can work also as a harmonic mixer if larger amplitude of the RF signal is used. In this case intermediate frequencies $lf_{LO} \pm mf_{RF}$ will be generated, where l and m are integers. In figure 2.4.9 a conversion loss of 19.6 dB at an oscillator power of -18.5 dBm and radio-frequency power of -34.3 dBm is obtained. The conversion losts could be reduced by applying the RF signal through an additional inverter, which will also suppress the influence of the oscillator signal on the radio-frequency port.

In this way the first stand-alone graphene mixers were realized demonstrating the versatility of the fabricated ROs. This shows that graphene can find immediate applications in RF (microwave) electronics. RF circuits consist of voltage amplifiers, oscillators and mixers. The demonstration of the first graphene ROs operating under ambient conditions opens a possibility to realize all-graphene

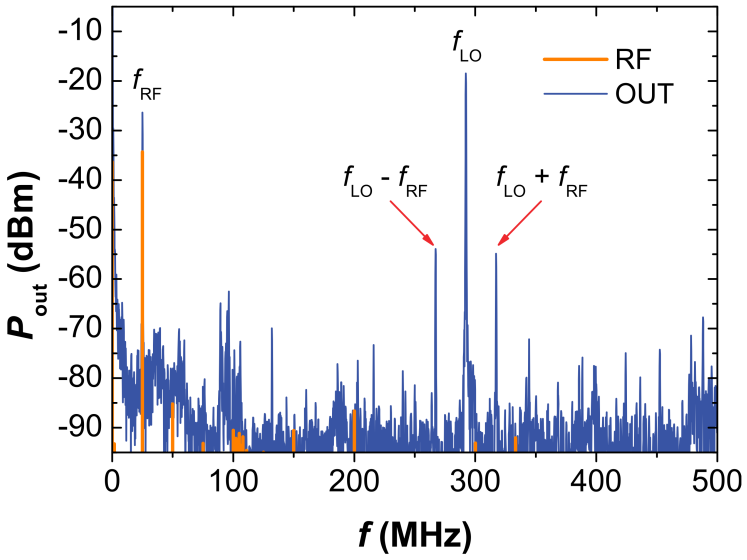


Figure 2.4.9: Power spectrum of the input radio-frequency signal (RF) and output signal (OUT) of the fabricated stand-alone graphene frequency mixer [12].

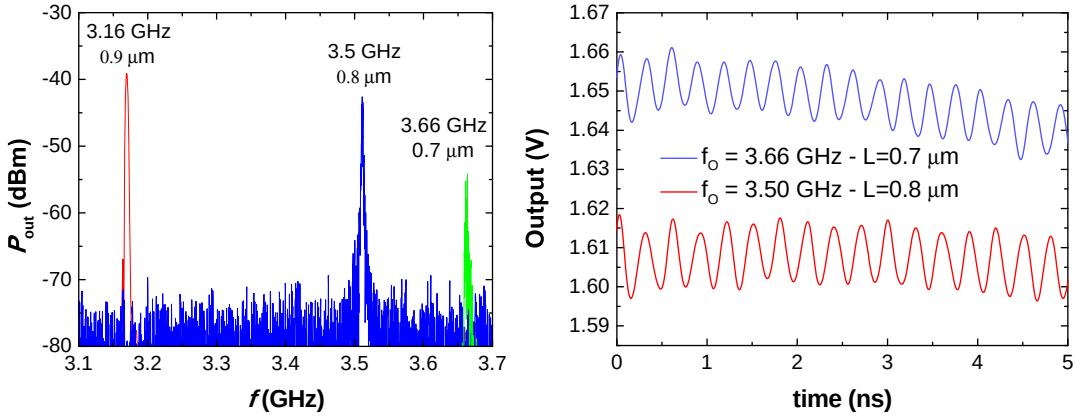
microwave circuits.

2.4.4 Scaling of graphene ROs

In order to further increase the oscillation frequency of graphene ROs it is necessary to scale GFETs to sub-micron gate lengths. High frequency operation requires low resistances R and capacitances C since $f_O \propto 1/\tau \propto 1/RC$. The gate delay is the result of different contributions and can be expressed as

$$\tau = R_0 C_G = \left(R_S \frac{L}{W} + 2\frac{R_S L_a}{W} + \frac{R_C}{W} \right) W L C_{ox} \tag{2.4.9}$$

where R_0 is the resistance at the Dirac point, C_G is the gate capacitance, R_S is the sheet resistance, L_a is the access length and R_C is the contact resistance. As the dimensions of the channel length of the GFETs that form the inverters are reduced the capacitances and resistances of the channel become smaller and as a consequence f_O is increased. In order to achieve even higher frequencies the length of the line interconnections were reduced in order to obtain smaller parasitic resistances and capacitances. The ungated part of the channel was reduced by fabricating devices with smaller access length which increases both the voltage gain and the oscillation frequency. The reduction of the channel width reduces the parasitics capacitances but it also increases the channel and contact resistance. For this reason the best way to increase the speed of ROs is to decreases the channel length.



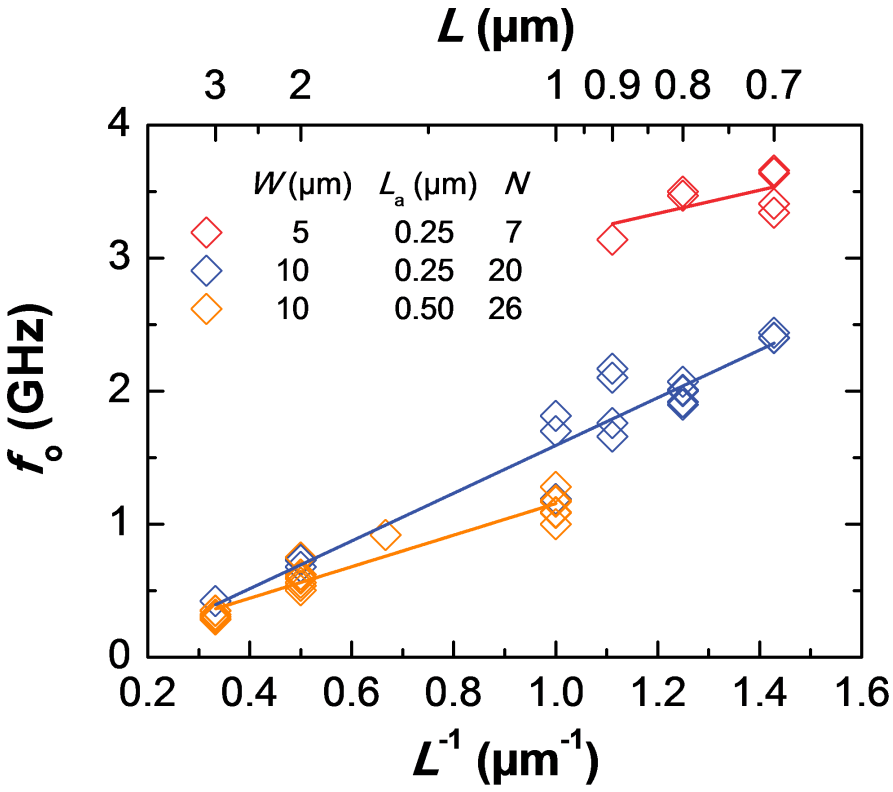
(a) (b)

Figure 2.4.10: (a) Power spectra of ROs with sub-micron graphene channel lengths of 0.9 μm, 0.8 μm and 0.7 μm, (b) output signal of the two smallest submicron buffered ROs.

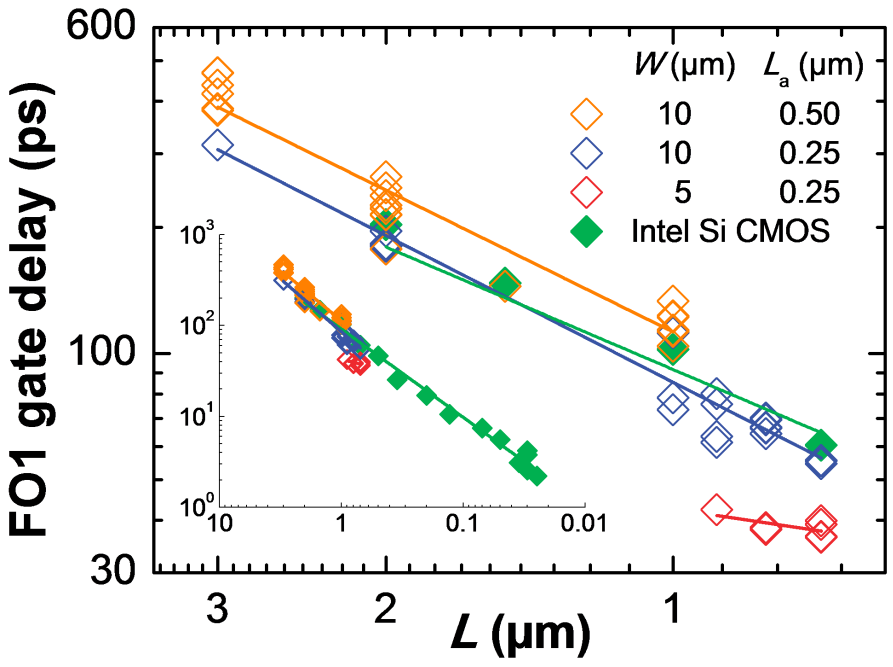
With all these changes the fabricated buffered graphene ROs reached higher oscillation frequencies shown in figure 2.4.10a. In particular, reducing the channel width to $W = 5\,\mu$m, the access length to $L_a = 0.25\,\mu$m and applying a supply voltage $V_{DD} = 2.5$ V, it was possible to achieve $f_O = 3.16$ GHz for $L = 0.9\,\mu$m, $f_O = 3.5$ GHz for $L = 0.8\,\mu$m and $f_O = 3.66$ GHz for $L = 0.9\,\mu$m. However, in all these cases the voltage swing at the output is very small, $V_{p-p} \sim 10$ mV. Such a small swing is a consequence of the limited bandwidth of the buffer inverter as discussed before. Here this bandwidth is higher (~ 300 MHz) because the load capacitance C_L was reduced by reducing the on-chip pad area. However, even such increased bandwidth is not high enough to eliminate the suppression of the measured high frequency signals.

Figure 2.4.11a show how the frequency of the fabricated buffered graphene ROs scale with $1/L$ for three different types of design. It would be expected that oscillation frequency f_O scales with $1/L^2$ due to the fact the both the resistance and the capacitance of the channel scale L. However, since the parasitic components due to the contact resistance (that scales with 1/W) and the parasitic resistances and capacitances of the interconnects do not scale, f_O only scales with $1/L$.

FO1 (which is the fan out that describes the number of gates attached at the output) gate delay of the realized ROs is plotted as a function of the channel length of the GFETs in figure 2.4.11b. A direct comparison with conventional Si CMOS gates is also shown. The fabricated devices follows the same trend as Si CMOS and most of them exhibit almost identical gate delay at the same



(a)



(b)

Figure 2.4.11: (a) The oscillation frequency f_O as a function of $1/L$ and the channel length L and (b) FO1 gate delay in function of L.

channel lengths. However, ROs with $W = 5\,\mu$m exhibit even smaller gate delays compared to Si CMOS, although they cannot compete with Si CMOS in terms of static power consumption.

2.5 Conclusions

From its discovery graphene has been considered to be the material of the future that could replace silicon in state-of-the-art MOSFETs. However, due to the absence of a band-gap, graphene has not been used in realistic electronic circuits, but mainly in niche analog applications which do not require signal amplification. Previously to this work none of the fabricated graphene devices were capable of signal amplification. In order to effectively prove that graphene could replace silicon, realistic graphene electronic circuits exhibiting such amplification have to be realized.

The extremely high mobility of charge carriers in graphene at room temperature makes this material suitable for applications in high-speed electronics. GFETs exhibiting an over-unity intrinsic voltage gain are required for the realization of more complex graphene circuits and for this reason graphene voltage amplifiers were firstly investigated. They were obtained by using a very thin top-gate dielectric that allowed the realization of graphene complementary inverters, on single graphene flakes, with a high voltage gain (>10 dB) at room temperature. The AlO_x gate dielectric and the complete coverage of the channel increased the control over the charge carriers and eliminates the access resistances, allowing to reach high voltage gains A_v. This result reported the highest voltage gain obtained at room temperature for any kind of graphene amplifiers realized until that time.

Such graphene devices could be used as the main building block of both analog and digital electronics. However graphene cannot compete directly with conventional silicon CMOS logic inverters which are capable of reaching a voltage swing of 100% of the supply voltage, compared to $\sim 20\%$ for the fabricated graphene devices. It is also difficult to reach a saturation region of the drain current in GFETs, in which the output conductance is minimized, allowing to have large voltage gains. It was shown that the demonstrated technology is not scalable since it is based on graphene flakes obtained by mechanical exfoliation and it suffers from large static power dissipation due to the impossibility to turn-off the graphene transistors.

The fabrication of graphene amplifiers represents only the first step in the

realization of realistic graphene digital circuits. The amplifiers were also used as digital logic gates at room temperature by applying the square-wave signal at the input. High values of the voltage gain in digital applications are required in order to distinguish between the two logic levels and to match the input and output signals. In this case it was not possible to obtain the matching between the input and output signals even though it was possible to distinguish the two logic levels. It was found that the mismatch between the signals is equal to the voltage at the Dirac point of the unbiased GFETs. Typically, charge neutrality points were far away from zero, showing a p-type doping in air, due to the adsorption of ambient impurities on graphene prior to the fabrication. In order to match the signals it was necessary to fabricate GFETs with Dirac points close to zero.

All fabricated graphene voltage amplifiers were realized using mechanically exfoliated graphene which is not a scalable technology. In order to solve the problem of the scalability of complementary graphene inverters it was necessary to use large-scale graphene films grown by chemical vapor deposition (CVD). Graphene monolayers, provided by the group of Prof. Eric Pop from the Stanford University, where grown by CVD on Cu foils with a CH_4 precursor. They were transferred to SiO_2/Si substrates and integrated into graphene digital complementary inverters. Such a wafer-scale material allowed to achieved a weak saturation regime of the drain current and large voltage gain ($A_V = -5.3$). This is the highest gain reported for large-area monolayer graphene under ambient conditions in top-gated GFETs. It was demonstrated that by using GFETs with $V_0 < 0.2$ V in digital logic gates, with sufficiently large voltage gain, signal matching can be obtained. The voltage swing of the input and output signals was $V_{in,p-p} = V_{out,p-p} = 0.56$ V. The same result was achieved even in GFETs with $V_0 > 0.2$ V but performing the measurement in vacuum. The lower pressure reduced only the voltage gain due to a different shift of the Dirac points of the GFETs that form the inverter. The fabricated graphene inverters were capable of reaching a voltage swing of 22.4% of the power supply, since GFETs cannot be turned-off in both logic states. Graphene logic gates could compete with high-frequency emitter coupled logic (ECL) gates in which the voltage swing is only 15% of the supply voltage. Replacing the typical materials that are used in the state-of-the-art of ECL gates with graphene, which has a higher mobility, it would be possible to reach even higher frequencies.

At the DC operating point that lies in between the highest gain points of two inverters (in which they exhibit$|A_v| > 4$) the over-unity voltage gain is maintained in both inverters and as a consequence cascade connection can be realized, with

the correct logic levels. In this way it was possible to trigger one stage with another under ambient conditions, a result that has never been achieved before at any temperature.. This paves the way for the realization of realistic graphene digital circuits in which static power dissipation is not an issue. The very thin aluminum oxide layer allowed to reach the same oxide capacitance as in the 65 nm gate length Si CMOS which exceeds the typical oxide capacitance reported for graphene transistors. This provides a very good control over the carriers in the channel.

The most important circuits in digital electronics are the ring oscillators (ROs) which are made by connecting an odd number of inverters so that the output of one inverter is connected to the input of the next inverter. ROs represent the perfect device to test the ultimate performances limits of a given technology. The previously demonstrated graphene logic gates were used in the realization of graphene ROs. This was the first demonstration of graphene digital integrated circuits operating at gigahertz frequencies. It was also demonstrated that as the channel length of the GFETs in the ROs is reduced the capacitances and resistances of the channel are reduced and as a consequence f_O is increased. Reducing the parasitic capacitances and resistances it was possible to reach, at the channel length $L = 0.7\,\mu$m, a maximum oscillation frequency of 3.66 GHz. It was demonstrated that f_O only scales with $1/L$ even though both the resistance and the capacitance of the channel scale with L. This is because the contact resistance (that scales with $1/W$) and the parasitic resistances and capacitances of the interconnects do not scale. It was found that the fabricated ROs exhibited even smaller gate delays compared to Si CMOS at the same gate length. Another advantage of graphene ROs was that they had a smaller sensitivity to the power supply fluctuations compared to silicon CMOS ROs.

Graphene digital circuits cannot directly compete with Si CMOS circuits due to larger power consumption. However, graphene could replace InP in high-frequency electronics due to its higher mobility but for this to happen further technological advances are needed.

Bibliography

[1] A. Bachtold, P. Hadley, T. Nakanishi, and C. Dekker. Logic circuits with carbon nanotube transistors. *Science*, 294(5545):1317–1320, November 2001.

[2] Alexander A. Balandin. Thermal properties of graphene and nanostructured carbon materials. *Nature materials*, 10(8):569–581, August 2011.

[3] R. Behzad. *Design of analog CMOS integrated circuits*. McGraw-Hill, 2001.

[4] P. Blake, E. W. Hill, A. H. Castro Neto, K. S. Novoselov, D. Jiang, R. Yang, T. J. Booth, and A. K. Geim. Making graphene visible. *Appl. Phys. Lett.*, 91(6):063124, 2007.

[5] A. H. Castro Neto, N. M. R. Peres, Kostya S. Novoselov, and Andre K. Geim. The electronic properties of graphene. *Reviews of Modern Physics*, 81(1):109–162, January 2009.

[6] James G. Champlain. A first principles theoretical examination of graphene-based field effect transistors. *Journal of Applied Physics*, 109(8):084515, 2011.

[7] Zhihong Chen, Joerg Appenzeller, Yu-Ming Lin, Jennifer Sippel-Oakley, Andrew G. Rinzler, Jinyao Tang, Shalom J. Wind, Paul M. Solomon, and Phaedon Avouris. An integrated logic circuit assembled on a single carbon nanotube. *Science*, 311(5768):1735, March 2006.

[8] Rui Cheng, Jingwei Bai, Lei Liao, Hailong Zhou, Yu Chen, Lixin Liu, Yung-Chen Lin, Shan Jiang, Yu Huang, and Xiangfeng Duan. High-frequency self-aligned graphene transistors with transferred gate stacks. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29):11588–11592, July 2012.

[9] T. Fang, A. Konar, H. Xing, and D. Jena. Carrier statistics and quantum capacitance of graphene sheets and ribbons. *Appl. Phys. Lett.*, 91(9):092109–3, August 2007.

[10] Aaron D. Franklin, Shu-Jen Han, Ageeth A. Bol, and Vasili Perebeinos. Double Contacts for Improved Performance of Graphene Transistors. *IEEE Electron Device Letters*, 33(1):17–19, January 2012.

[11] Andre K. Geim. Graphene: status and prospects. *Science (New York, N.Y.)*, 324(5934):1530–1534, June 2009.

[12] Erica Guerriero, Laura Polloni, Massimiliano Bianchi, Ashkan Behnam, Enrique Carrion, Laura Giorgia Rizzi, Eric Pop, and Roman Sordan. Gigahertz integrated graphene ring oscillators. *ACS nano*, 7(6):5588–5594, 2013.

[13] Erica Guerriero, Laura Polloni, Laura Giorgia Rizzi, Massimiliano Bianchi, Giovanni Mondello, and Roman Sordan. Graphene audio voltage amplifier. *Small*, 8(3):357–361, February 2012.

[14] Shu-Jen Han, Keith A. Jenkins, Alberto Valdes-Garcia, Aaron D. Franklin, Ageeth A. Bol, and Wilfried Haensch. High-frequency graphene voltage amplifier. *Nano letters*, 11(9):3690–3693, September 2011.

[15] Naoki Harada, Katsunori Yagi, Shintaro Sato, and Naoki Yokoyama. A polarity-controllable graphene inverter. *Applied Physics Letters*, 96(1):012102, 2010.

[16] Allen Hsu, Han Wang, Ki Kang Kim, Jing Kong, and Tomas Palacios. Impact of Graphene Interface Quality on Contact Resistance and RF Device Performance. *IEEE Electron Device Letters*, 32(8):1008–1010, August 2011.

[17] C.-H. Jan, U. Bhattacharya, R. Brain, S.-J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, and P. Bai. A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications. *International Electron Devices Meeting*, pages 3.1.1–3.1.4, December 2012.

[18] John P. Uyemura. *CMOS Logic Circuit Design*. Kluwer Academic Publishers, 2001.

[19] M.C. Lemme, T.J. Echtermeyer, M. Baus, and H. Kurz. A graphene field-effect device. *Electron Device Letters, IEEE*, 28(4):282–284, 2007.

[20] Song-Lin Li, Hisao Miyazaki, Akichika Kumatani, Akinobu Kanda, and Kazuhito Tsukagoshi. Low operating bias and matched input-output characteristics in graphene logic inverters. *Nano letters*, 10(7):2357–2362, July 2010.

[21] Song-Lin Li, Hisao Miyazaki, Akichika Kumatani, Akinobu Kanda, and Kazuhito Tsukagoshi. Low operating bias and matched input-output characteristics in graphene logic inverters - SI. *Nano letters*, 10(7):SI, July 2010.

[22] Wei Li, Yiran Liang, Dangmin Yu, Lianmao Peng, Kurt P. Pernstich, Tian Shen, A. R. Hight Walker, Guangjun Cheng, Christina A. Hacker, Curt A. Richter, Qiliang Li, David J. Gundlach, and Xuelei Liang. Ultraviolet/ozone treatment to reduce metal-graphene contact resistance. *Applied Physics Letters*, 102(18):183110, 2013.

[23] Albert D. Liao, Justin Z. Wu, Xinran Wang, Kristof Tahy, Debdeep Jena, Hongjie Dai, and Eric Pop. Thermally Limited Current Carrying Ability of Graphene Nanoribbons. *Physical Review Letters*, 106(25):256801, June 2011.

[24] Chun-Chieh Lu, Yung-Chang Lin, Chao-Hui Yeh, Ju-Chun Huang, and Po-Wen Chiu. High Mobility Flexible GrapheneField-Effect Transistors with Self-Healing Gate Dielectrics - SI. *ACS nano*, 6(5):SI, 2012.

[25] J. Martin, N. Akerman, G. Ulbricht, T. Lohmann, J. H. Smet, K. von Klitzing, and A. Yacoby. Observation of electron-hole puddles in graphene using a scanning single-electron transistor. *Nature Physics*, 4(2):144–148, November 2007.

[26] I. Meric, M. Y. Han, A. F. Young, B. Ozyilmaz, P. Kim, and K. L. Shepard. Current saturation in zero-bandgap, top-gated graphene field-effect transistors. *Nat Nano*, 3(11):654–659, November 2008.

[27] J. S. Moon, M. Antcliffe, H. C. Seo, D. Curtis, S. Lin, A. Schmitz, I. Milosavljevic, A. A. Kiselev, R. S. Ross, D. K. Gaskill, P. M. Campbell, R. C. Fitch, K.-M. Lee, and P. Asbeck. Ultra-low resistance ohmic contacts in graphene field effect transistors. *Applied Physics Letters*, 100(20):203512, 2012.

[28] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov. Electric field effect in atomically thin carbon films. *Science*, 306(5696):666–669, 2004.

[29] Dharmendar Reddy, Leonard Franklin Register, Gary D. Carpenter, and Sanjay K. Banerjee. Graphene field-effect transistors. *Journal of Physics D: Applied Physics*, 45(1):019501, January 2012.

[30] Laura Giorgia Rizzi, Massimiliano Bianchi, Ashkan Behnam, Enrique Carrion, Erica Guerriero, Laura Polloni, Eric Pop, and Roman Sordan. Cascading wafer-scale integrated graphene complementary inverters under ambient conditions - SI. *Nano letters*, 12(8):SI, August 2012.

[31] Frank Schwierz. Graphene transistors. *Nature nanotechnology*, 5:487–496, 2010.

[32] Frank Schwierz. Graphene transistors: status, prospects and problems. *Proceedings of the IEEE*, 101(7):1567–1584, July 2013.

[33] Roman Sordan, Floriano Traversi, and Valeria Russo. Logic gates with a single graphene transistor. *Applied Physics Letters*, 94(7):073305, 2009.

[34] B. N. Szafranek, G. Fiori, D. Schall, Daniel Neumaier, and Heinrich Kurz. Current Saturation and Voltage Gain in Bilayer Graphene Field Effect Transistors. *Nano letters*, 12(3):1324–1328, March 2012.

[35] S. M. Sze and K. K. Ng. *Physics of semiconductor devices*. Wiley-Interscience, third edition, 2006.

[36] Yuan Taur and T. H. Ning. *Fundamentals of modern VLSI devices*. Cambridge University Press, second edition, 1998.

[37] S. A. Thiele, J. A. Schaefer, and Frank Schwierz. Modeling of graphene metal-oxide-semiconductor field-effect transistors with gapless large-area graphene channels. *Journal of Applied Physics*, 107(9):094505, 2010.

[38] Scott E. Thompson, Senior Member, Robert S. Chau, Tahir Ghani, Kaizad Mistry, Sunit Tyagi, and Mark T. Bohr. In Search of Forever, Continued Transistor Scaling One New Material at a Time. *IEEE Transactions on Semiconductor Manufactoring*, 18(1):26–36, 2005.

[39] Floriano Traversi, Valeria Russo, and Roman Sordan. Integrated complementary graphene inverter. *Applied Physics Letters*, 94(22):223312, 2009.

[40] Han Wang, Allen Hsu, Justin Wu, Jing Kong, and Tomas Palacios. Graphene-based ambipolar RF mixers. *IEEE Electron Device Letters*, 31(9):906–908, 2010.

[41] Han Wang, Daniel Nezich, Jing Kong, and Tomas Palacios. Graphene frequency multipliers. *IEEE Electron Device Letters*, 30(5):547–549, 2009.

[42] Han Wang, Lili Yu, Yi-Hsien Lee, Yumeng Shi, Allen Hsu, Matthew L. Chin, Lain-Jong Li, Madan Dubey, Jing Kong, and Tomas Palacios. Integrated Circuits Based on Bilayer MoS2 Transistors. *Nano letters*, 12(9):4674–4680, September 2012.

[43] J. D. Wood, S.W. Schmucker, A. S. Lyons, Eric Pop, and Joseph W. Lyding. Effects of polycrystalline Cu substrate on graphene growth by chemical vapor deposition. *Nano letters*, 11:4547–4554, 2011.

[44] Fengnian Xia, Damon B. Farmer, Yu-Ming Lin, and Phaedon Avouris. Graphene field-effect transistors with high on/off current ratio and large transport band gap at room temperature. *Nano letters*, 10(2):715–718, February 2010.

[45] Fengnian Xia, Vasili Perebeinos, Yu-ming Lin, Yanqing Wu, and Phaedon Avouris. The origins and limits of metal-graphene junction resistance. *Nature nanotechnology*, 6(3):179–184, March 2011.