Abstract

The topic of this seminar is graphene and its possible applications in the field of electronics, most notably as a candidate material for the transistor. The metal-oxide-semiconductor field-effect transistor (MOSFET) is one of the most used active components of integrated chips, but being silicon based it faces a problem, because silicon may in the near future not meet the requirements of the electronics industry any longer. Yet still in its infancy, graphene might present a viable alternative.
## Contents

1. **Introduction** 3

2. **The MOSFET** 3
   - 2.1 Structure and operation 3
   - 2.2 Characteristics and application 5

3. **Graphene** 5
   - 3.1 Dispersion relation 6
   - 3.2 The Klein paradox 8

4. **Modern graphene transistors** 11

5. **Conclusion** 12
1 Introduction

Graphene is an exciting material that has been in the spotlight of condensed-matter physics ever since the Manchester group (Novoselov, Geim et al [1]) published their ground-breaking paper in 2004. Its interesting electrical and mechanical properties suggest a huge potential for application in many different fields, but especially in electronics. Although graphene is a single sheet of carbon atoms in a honeycomb lattice it is easily produced using ordinary sticky tape and some graphite. By applying and removing the tape to the graphite the graphene sheets are exfoliated onto the tape. This was done by the Manchester group which spearheaded the work for which they received the Nobel prize in 2010.

2 The MOSFET

2.1 Structure and operation

A transistor is a semiconductor device used in electrical circuits for logical operations and signal amplification. Its widespread use in modern electronics along with the highly automated and low-cost semiconductor device fabrication makes it one of the key active components of the integrated chip, and perhaps one of the most far-reaching inventions of the twentieth century. There are many different types of transistors, but by far the most popular is the MOSFET or metal-oxide-semiconductor field-effect transistor.

Figure 1: The MOSFET transistor, called PMOS if the body is a p-type semiconductor (or NMOS conversely). [2]

It is a 4-terminal device, with the terminals being the source and drain electrodes, separated by the channel region, the gate electrode whereby the channel conductivity is controlled, and the body made from a doped semiconductor (Figure 1). The body is most commonly connected to the source (effectively making the transistor a 3-terminal device), but is otherwise used to control the strength of the inversion layer because the gate-to-body bias positions the conduction band energy levels, while the source-to-body bias positions the electron Fermi level. The MOSFET owes its name in part to the configuration of an oxide insulator (e.g. SiO$_2$) sandwiched between the metal gate and the semiconductor body (though in modern MOSFETs other materials can be used, such as highly doped silicon for the gate). This is the part of the transistor that is used to control the conductivity of the channel region and is...
called the MOS capacitor. The MOSFET is a field-effect transistor, which means the conductivity of the channel is controlled with an electric field and it doesn’t rely upon the contact of different semiconductors for operation. In order to understand the electronic properties of the MOS capacitor, let us consider the band diagrams shown in Figure 2.

At a junction of two conducting materials with different work functions (and different Fermi levels) a current will flow until the Fermi levels are equal throughout the junction at thermodynamic equilibrium. Shown in Figure 2 is energy band bending at the metal-semiconductor junction, where the metal work function is higher than the semiconductor work function \( \phi_m > \phi_s \). Consequently if an electron travels from the semiconductor to the metal, the energy of the system, along with the difference between the Fermi levels \( \phi_m \) and \( \phi_s \), decreases. The result of this process is a surface charge at the junction. When one of the materials forming the junction is a semiconductor the result is a depletion or accumulation zone, depending on the type of majority charge carriers in the substrate, and the relative values of the Fermi levels of either material. This means we can influence the surface charge by manipulating the Fermi levels with the gate voltage. If we apply positive voltage \( V_g \) to the metal gate from the example in Figure 2, the Fermi level in the metal \( E_{F,m} \) will lower by \( q \cdot V_g \) which will in turn cause a higher degree of band bending and a larger depletion layer in the semiconductor. If we raise the voltage further we eventually reach the threshold voltage \( V_{th} \) where the surface charge concentration is equal to the body doping concentration. This means \( \phi_{SB} = E_{F,s} - E_V \). If the gate voltage is higher than the threshold voltage, the

Figure 2: The energy band diagram. \( \phi_{m,s} \) are metal/semiconductor work functions, \( E_{F,m} \) and \( E_{F,s} \) their Fermi levels, \( E_V \) the top of the valence band, \( E_C \) the bottom of the conduction band, \( \chi_s \) electron affinity, \( \phi_{SB} \) the surface potential and \( V_{BB} \) the degree of bending \( (V_{BB} = |\phi_m - \phi_s|) \).
majority carriers in the body are unable to screen the electric field, so the minority carriers are attracted to the junction, forming an inversion layer. These carriers are conductive parallel to the surface of the junction, effectively forming a channel. For the MOSFET two terminals made of the opposite type of semiconductor than the body are placed on either side of the capacitor, called the source and drain electrodes, through which the current flows.

2.2 Characteristics and application

If a material is to succeed silicon in MOSFET production, it has to showcase two important properties. Firstly, a trend in the semiconductor industry is that the transistor channel length (L in Figure 1) halves every 18 months (Moore’s law). This has been the case since the 1970s when mass production first started with transistor channel lengths of $\sim 10 \mu m$ until now when they are some three orders of magnitude smaller (5th gen. Intel). This is desired as more MOSFETs can be fitted in a single chip, making integrated circuits more and more complex. In that light, scaling alone has provided the needed performance improvements in the industry. But it appears silicon-based devices are approaching their limits \cite{5}, as short-channel effects such as drain-induced barrier lowering are hindering MOSFET functionality, so an alternative is to be found. The electronic properties of candidate materials must be robust against short-channel effects. Secondly an important MOSFET structure in integrated circuits is the complementary metal-oxide-semiconductor (CMOS) which consists of an n-type and a p-type MOSFET (denoting the body semiconductor type). During operation one of the MOSFETs is always off, which makes for a very low static power consumption, low waste heat and high density of logic functions in a chip. This means switching is a very desirable property of the MOSFET, so the presence of a band gap in candidate materials would be a priority.

3 Graphene

Graphene is a two-dimensional lattice of carbon atoms arranged in a hexagonal (honeycomb) pattern (Figure 3). The understanding of its electronic properties pertains to other carbon allotropes and is important as such. For example, graphene can be thought of as composed from benzene rings stripped from their hydrogen atoms, or, wrapped up, as nanotubes or fullerenes. \cite{6} The $sp^2$-hybridization leads to the carbon atoms forming three bonds each at an angle of 120° to one another, of the length 1.46 Å. The third (nonhybridized) p-orbital is perpendicular to the three bonds and is half-filled. The hexag-
onal lattice is essentially a triangular lattice with a basis of two atoms per unit cell, therefore the reciprocal lattice is also a triangular lattice and the first Brillouin zone is a hexagon. Due to the lattice symmetry only two of the six vertices (known as Dirac points) are non-equivalent.

Figure 3: The hexagonal lattice with two primitive vectors $a_1$ and $a_2$ and the lattice constant $a$. Sites that differ for a primitive vector are equivalent (either A or B).

3.1 Dispersion relation

Using the tight-binding approximation the Hamiltonian can be written as

$$H = -t \sum_{\langle i,j \rangle,\sigma} (a_{\sigma,i}^\dagger b_{\sigma,j} + h.c.)$$

$$- t' \sum_{\langle\langle i,j \rangle\rangle,\sigma} (a_{\sigma,i}^\dagger a_{\sigma,j} + b_{\sigma,i}^\dagger b_{\sigma,j} + h.c.),$$

(1)

where $t \approx 2.8 \text{ eV}$ is the nearest-neighbor electron hopping energy and $t'$ the next-nearest-neighbor electron hopping energy ($0.02t \lesssim t' \lesssim 0.2t$). The creation-annihilation operators $a_{\sigma,i}^\dagger$ ($b_{\sigma,i}^\dagger$) and $a_{\sigma,i}$ ($b_{\sigma,i}$) create or annihilate states with spin $\sigma$ at A (B) sites. Since the Hamiltonian is of a quadratic form it can be diagonalized with an appropriate Bogoliubov transformation.

Figure 4: a) Dispersion relation with parameters $t = 2.8 \text{ eV}$ and $t' = 0.4 \text{ eV}$. b) Energy bands at the plane $k_y = 0$. The non-zero next-nearest-neighbor hopping energy $t'$ breaks the electron-hole symmetry of the bands.

The derived energy dispersion (Figure 4) is of the form

$$E = \pm t\sqrt{3 + f(k)} + t'f(k),$$

(2)
where

\[ f(k) = 2 \cos (k_x a) + 4 \cos \left( \frac{1}{2} k_x a \right) \cos \left( \frac{\sqrt{3}}{2} k_y a \right), \]

\[ k = (k_x, k_y), \quad a_1 = a \left( \frac{1}{2}, \frac{\sqrt{3}}{2} \right) \quad \text{and} \quad a_2 = a \left( -\frac{1}{2}, \frac{\sqrt{3}}{2} \right). \]

Since the carbon p-orbitals that contain the free electron are half-filled, the Fermi level passes through the Dirac points. By expanding the dispersion relation close to the \( K \) point \( (k = K + q, |q| \ll |K|) \), equation (2) can be written in a simplified form as

\[ E(q) = \pm v_F |q| + \mathcal{O}(|q/K|^2). \quad (3) \]

The quantity \( v_F \) is the Fermi velocity and is of the order of \( 10^6 \ \text{m/s} \) which is similar to the Fermi velocity in metals like lithium \( (1.29 \times 10^6 \ \text{m/s}) \) or sodium \( (1.07 \times 10^6 \ \text{m/s}) \). Equation (3) is a linear dispersion relation which means electrons with energies near the Fermi level behave as massless fermions. With the electron effective mass \( m^* \) approaching zero, the electrical conductivity \( \sigma_e \), defined in the Fermi gas model as

\[ \sigma_e = \frac{n e^2 \tau}{m^*}, \quad (4) \]

where \( n \) is the number density of electrons, \( e \) electron charge and \( \tau \) the average electron scattering time, approaches infinity. High conductivity is of course a desirable characteristic of any material that is to be used for a conducting channel in electronic devices, but presents a fundamental problem for transistor applications as a consequence of linear energy dispersion (linear energy dispersion is gapless, but an energy gap is required).

Table 1: Band gaps \( (E_g) \) and electron mobilities \( (\mu) \) of some semiconductors and graphene. [7]

<table>
<thead>
<tr>
<th>Material</th>
<th>( E_g ) [eV]</th>
<th>( \mu ) [cm(^2)/Vs]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ge</td>
<td>0.66</td>
<td>3900</td>
</tr>
<tr>
<td>Si</td>
<td>1.12</td>
<td>1400</td>
</tr>
<tr>
<td>GaAs</td>
<td>1.42</td>
<td>8500</td>
</tr>
<tr>
<td>InN</td>
<td>1.97</td>
<td>3200</td>
</tr>
<tr>
<td>Graphene</td>
<td>0</td>
<td>15000</td>
</tr>
</tbody>
</table>

If a solution is to be found, a compromise must be struck, sacrificing one for the other. There are several proposed methods of creating an energy gap in graphene, such as bilayer graphene, graphene nanoribbons (GNRs) and uniaxial strain. The dispersion relation in bilayer graphene is different to single-layer graphene due to additional non-negligible hopping energies between atoms in different layers and has an energy gap under the right conditions (if an electric field is applied perpendicular to the bilayer). Ideal GNRs are ribbons of graphene which have a finite width and a well-defined edge. The band gap is opened due to the confinement of electrons in the direction perpendicular to the GNR length. Finally, strain induced in epitaxial graphene (due to interactions with the
substrate, or conversely by stretching the substrate itself) can produce a band gap, but is still subject to debate. All three methods are, however, currently not applicable for practical purposes as they either present an engineering problem (it is difficult to produce GNRs of uniform width and global uniaxial strain of upwards of 20% is required to open a band gap in single-layer graphene, which is difficult to achieve), or by inducing an energy gap lower the mobility below that of silicon (bilayer graphene). Table 1 shows the band gaps and mobilities of some of the most widely used semiconductors. Such high values as 10000-15000 cm$^2$/Vs are routinely measured in graphene, but drop drastically when a band gap is opened (Figure 5). This is because the electron mobility is inversely proportional to the effective electron mass, which is in turn proportional to the energy band gap. 

**3.2 The Klein paradox**

The Klein paradox describes the peculiar scattering behaviour of relativistic particles at potential barriers of sufficient height. In non-relativistic quantum tunneling the case is that electrons can pass through high potential barriers even if it is forbidden classically. With the increasing barrier height (in the classical limit), the transmission coefficient approaches zero exponentially. Relativistic particles, however, can pass through high barriers with a high transmission probability. Moreover, even absolute transparency can be achieved for electrons at specific angles of incidence. Graphene provides a unique medium where quantum electrodynamics phenomena can be tested experimentally, because Dirac electrons can be described using the Dirac equation due to the similarities to relativistic fermions.

![Figure 5: III-V compounds from left to right are InSb, InAs, In$_{0.53}$Ga$_{0.47}$As, InP, GaAs, In$_{0.49}$Ga$_{0.51}$P and GaN. Also shown are simulation results for carbon nanotubes (CNTs) and GNRs. The higher graphene value is a simulation result and the lower an experimental measurement (see [5] for original sources).](image-url)
heights needed to for the Klein paradox to be observable are much lower. The formulation of the problem is as follows. Let the barrier \( V(x) \) be defined as

\[
V(x) = \begin{cases} 
V_0, & 0 < x < D, \\
0, & \text{otherwise},
\end{cases}
\]

and the effective hamiltonian as

\[
H = -i\hbar v_F \sigma \nabla, \tag{5}
\]

where \( \sigma = (\sigma_x, \sigma_y) \) are the Pauli matrices (note the Hamiltonian is equivalent to that in the massless Dirac equation). The wave function ansatz (\( \psi_1 \) and \( \psi_2 \) for electrons and holes respectively) has the form

\[
\psi_1 = \begin{cases} 
(e^{ik_x x} + re^{-ik_x x})e^{ik_y y}, & x < 0, \\
(a e^{iq_x x} + be^{-iq_x x})e^{ik_y y}, & 0 < x < D, \\
t e^{ik_x x + ik_y y}, & x > D,
\end{cases}
\]

and

\[
\psi_2 = \begin{cases} 
s(e^{ik_x x + i\phi} + re^{-ik_x x - q\phi})e^{ik_y y}, & x < 0, \\
s'(ae^{iq_x x + i\theta} + be^{-iq_x x - i\theta})e^{ik_y y}, & 0 < x < D, \\
st e^{ik_x x + ik_y y + i\phi}, & x > D,
\end{cases}
\]

where \( k_F = 2\pi/\lambda \) is the Fermi wavevector, \( k_x = k_F \cos \phi \) and \( k_y = k_F \sin \phi \) the wavevector components outside the barrier, \( q_x = \sqrt{(E - V_0)^2/\hbar^2v_F^2 - k_y^2} \), the refraction angle \( \theta = \arctan q_y/q_x \), and finally \( s = \text{sgn}(E) \) and \( s' = \text{sgn}(E - V_0) \). The refraction and transmission coefficients are obtained through applying the condition that the wavefunctions be continuous over the barrier. The result is angle-dependent and shown in Figure 6 (for the explicit expression see ref. 7, equation 3).

Figure 6: Transmission probability through a 100-nm wide potential barrier as a function of incident angle \( \phi \) for barrier height \( V_0 = 200 \text{ meV} \) (red) and \( V_0 = 285 \text{ meV} \) (blue). [9]

Now consider the energy spectrum of a potential well. In the same way the potential well has bound electron states, a potential barrier has bound positron (hole) states. If the positron energy inside the spectrum aligns with the energy continuum outside, this leads to high-probability tunneling. [9] Since electrons and holes are described
by elements of the same spinor function in the Dirac equation (charge-conjugation symmetry), it follows naturally that the hole in the barrier moves in the opposite direction to the electron, conserving momentum. Essentially an incident electron transforms into a hole within the barrier which transforms back into an electron on the other side. Another interpretation of the paradox is through pseudospin. In one dimension the energy dispersion for energies near the Fermi level is two intersecting linear functions \( E \propto \pm k \), each attributed to their respective lattice sites A or B. An electron with energy \( E \) propagating right in Figure 7 is on the same line as a hole with energy \(-E\), moving in the opposite direction. Thus electrons and holes from the same branch have the same pseudospin \( \sigma \), pointing parallel to the electron or antiparallel to the hole velocity.

In a normal conductor the electron transport and transmission probability is highly dependant upon the distribution of all possible scatterers which affect electrons by some potential (Figure 8). This is not true, however, for Dirac electrons, because if the pseudospin is to be conserved the electrons (changing into holes and back) have to travel through these potential barriers unaffected.

\[9\]

The Klein paradox is of course an unwanted property of a transistor material as the transistor cannot be pinched off due to the transparency of a gate-influenced barrier to incoming electrons. The solution lies in modifying the energy dispersion relation (by doping, for instance), rendering the electrons effectively massive and therefore describable again by the Schrödinger and not the Dirac equation.
Figure 9: Back-gated MOSFET (top), top-gated MOSFET with exfoliated graphene or graphene grown on metal and transferred to a SiO$_2$-covered Si wafer (middle), top-gated MOSFET with epitaxial graphene (bottom). [5]

4 Modern graphene transistors

The configuration of the first graphene MOSFET produced in 2004 is shown in Figure 9 (top). The graphene between the source and drain electrodes provides a conducting channel, below it is a SiO$_2$ dielectric, insulating it from the doped silicon, which serves as a bottom gate. The transistor provided a proof of concept, but not much else as the bottom gate made for unwanted transistor capacitance. In 2007 the first top-gate graphene MOSFET was produced with much more acceptable characteristics, and became a stepping stone for further advancements. Drain current characteristics for a graphene MOSFET are shown in Figure 10. Note the high leakage current (Dirac point) due to the negligible energy band gap.

Figure 10: Drain current characteristics for two MOSFETs with channel lengths $\sim 2.5 \mu m$ (red) and $\sim 10 \mu m$ (blue). [5]

Unlike Si-based transistors graphene MOSFETS conduct both at negative and positive top-gate voltages. The majority carriers are influenced highly by the top-gate voltage. At high positive voltages the carriers are electrons and at negative voltages, holes. Between the extremes is the Dirac point, or the “off” state of the transistor. The on-off ratios for large-area-
Graphene MOSFETs are currently below 20, which is not yet sufficient for application in logical circuits (especially compared to the orders-of-magnitude higher on-off ratios in contemporary MOSFETs).

5 Conclusion

Graphene as a material has several properties that suggest it is the prime candidate for the future of electronics. It is one atomic layer thick, and is as such inherently robust against short-channel effects. The electron mobility of zero-gap graphene far surpasses its semiconductor counterparts, which makes it ideal for high-frequency applications. But it still has many drawbacks, such as zero band gap, lower mobilities if a band gap is induced, difficulties with fabrication (in the case of GNRs) and so on. Further research is needed.

References


