Part VI

From quantum mechanics to conduction modeling

This part is an **integration of the theoretical elements given in Part1**. It provides first a good summary of quantum mechanics for those students who has not this part in their background; furthermore it goes again through the modeling of conduction toward the nanoscale with a refined point of view. References are provided for further details.

CREDITS: This chapter is derived by the MSc thesis of Ing. Fabrizio Mo.

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

CHAPTER 17

Brief review of quantum mechanics

The purpose of this chapter is to review the fundamental topics of quantum mechanics that are necessary to understand the next chapters related to molecular device modeling. These topics are nowadays often covered in the electronic engineer background, and for example they a re subject of bachelor's level courses at Politecnico di Torino. For this reason this chapter is intended to be a review of the main results to be kept in mind in the following of this work, rather than a complete introduction to quantum mechanics. Very good textbooks about quantum mechanics fundamentals, in my opinion, are the following: [210] (in Italian) that formally introduces all the topics without losing track of the physical insights, [211] that provides an excellent understandable introduction to these topics, even if sometimes the mathematical formalism is omitted, [44] that provides a complete, simple and understandable introduction to the fundamental topics without requiring particular prerequisite, [45] that provides an extreme simple and application-oriented introduction to quantum mechanics, [212] that provides a very rigorous and formal introduction to basic quantum mechanics, and [213] that provides a formal introduction to some advanced topics sometimes required in rigorous modeling of molecular devices. Many other valid and excellent textbooks are anyway present in literature (just to mention one: the milestone "Principles of Quantum Mechanics" by Paul A. M. Dirac). Throughout all this chapter I will always refer to the previously cited books, even if not always explicitly.

17.1 Wave-particle duality

The first question I will address in this introductory section is: "why quantum mechanics?". This entire work deals with molecular devices used as sensors. Molecules are the basic entities of matter and they can be composed of very few atoms (think for example to hydrogen molecule H_2 that is composed of only two hydrogen atoms), as well as a very large number of atoms. In both cases the characteristic dimensions are very small when compared to classical entities (like a rugby ball or even a small marble). For example the benzene ring has a characteristic dimension of less than 1 nm (it is around 5 Å), and again the fullerene C_{60} (a molecule with sixty carbon atoms) has a diameter of approximately 7 Å. In such microscopic systems novel physical effects arise. These novel effects and phenomena are not present in classical mechanics, that deals with the motion of macroscopic bodies, and in fact classical mechanics fails in accurately describing the motion and the interaction of such microscopic systems. The physical theory that, instead, correctly explains in a satisfactory manner the behavior of microscopic systems, like atoms and molecules, is quantum mechanics. Thus it must be considered the starting point in modeling molecular-based devices.

Quantum mechanics born at the beginning of the twentieth century, precisely to explain

physical phenomena, especially at the microscopic scale, that were not explainable at all with classical physics. Actually, important unsolved problems by the end of nineteenth century were for example related to the interaction between electromagnetic radiation and matter, that did not result in accordance with the classical laws of electromagnetism (i.e. Maxwell's equations). Among them I only cite the blackbody radiation spectrum, the photoelectric effect, the Compton effect, the discrete spectral lines in atomic or molecular absorption or emission of radiation and natural radioactivity. The explanation of all these, classically inexplicable, phenomena was possible only after that in 1900 Max Planck assumed that energy exchange between electromagnetic radiation and matter is not continuous, but it can happen only by discrete amounts, called energy "quanta". In Planck's model an energy quantum has an energy E proportional to the electromagnetic radiation frequency f by means of the Planck's constant h ($h = 6.62607015 \cdot 10^{-34}$ J·s):

$$E = hf \tag{17.1}$$

The Planck's assumption was in agree with the concept of photon (name coined by Lewis in 1926) introduced by Einstein in 1905, intended as an energy quantum of electromagnetic radiation. The interaction between electromagnetic fields and matter was thus rethought in terms of photons, meaning that only single photons interact with matter. This allowed to explain the photoelectric effect, the Compton effect and the blackbody radiation. The important conceptual consequence in this reasoning is that the electromagnetic radiation behaves like waves in some experiments and like a particle-based entity in others. This led scientists at the beginning of the twentieth century to accept the "dualistic" nature of the electromagnetic radiation. The wave nature of electromagnetic field is associated to physical quantities such as frequency or wavelength, while the particle-based nature is associated to a well defined energy and momentum. The connection among the two pictures is provided in equation (17.1). Moreover the modulus of the momentum associated to a photon, that is a particle with null mass since it is a quantum of electromagnetic energy, is provided by the following relativistic relation (introduced by Einstein in his work on special relativity):

$$p = \frac{E}{c} \tag{17.2}$$

where c is the light speed.
By putting together eq. (17.1) and eq. (17.2) it is possible to get (remember that $c = \lambda \cdot f$):

$$p = \frac{h}{\lambda} = \hbar k \tag{17.3}$$

where $\hbar$ is the reduced Planck's constant (i.e. $\frac{h}{2\pi}$) and k is the wavenumber (i.e. the modulus of the wave vector $\vec{k}$). The last relation can be then rewritten in vector form like (the wave vector is conventionally defined with the same direction of the propagation direction of the wave):

$$\vec{p} = \hbar\vec{k} \tag{17.4}$$

The just introduced quantities are the key points for the comprehension of many classically inexplicable phenomena, like the already mentioned photoelectric effect, Compton effect and blackbody radiation spectrum.
Instead the nature of atomic and molecular spectra was explained in a satisfactory way only after many other progresses were achieved, with a complete theory of microscopic world: the quantum mechanics. Nevertheless the intermediate achievements, even if sometimes they were *ad hoc* theories, are nowadays important in understanding the basic assumptions of quantum mechanics and its change of perspective w.r.t. classical physics. An important achievement I would like to briefly mention is the Niels Bohr's pioneering work on absorption and emission spectra of atoms, dated 1913. He assumed true the atomic structure

demonstrated by Rutherford, with a central nucleus around which electrons rotate. Classically such a system cannot exist since it would be unstable, indeed electrons are charge particles and in a circular orbit around the nucleus (i.e. accelerated motion) they should emit electromagnetic radiation loosing their own energy until they should collapse toward the nucleus. Bohr started from the obvious prove that atoms are essentially stable (since they exist), and assumed that well determined orbits were possible such that electrons (for some unknown reasons) can rotate around the nucleus without radiate. Even if this assumption was in conflict with classical laws he quantified it. Indeed he supposed that such orbits were circular and that the only "permitted" orbits were the ones whose the modulus l of the electron angular momentum w.r.t. nucleus (that was considered point-like) was an integer multiple of the reduced Planck's constant:

$$l = n\hbar \quad , \quad n = 1, 2, 3, ... \tag{17.5}$$

The orbits for which this angular momentum quantization law holds were supposed to give rise to "stationary" states, thus leading to stable structures. Then Bohr used the concept of photon and extended the Planck's assumption supposing that an electron in a steady state E by means of the absorption of a photon of suitable energy, i.e. frequency f, in a unique process, could be excited in a steady state E' (with $E' > E$). Consequently, involving the conservation of energy, he stated that the energy difference should be equal the energy provided by the photon (Bohr's formula):

$$\Delta E = E' - E = hf \tag{17.6}$$

The reasoning was analogous for the emission of photons by atoms and molecules (with a transition toward ground state from excited states). This model admits only transitions between stationary states, that have a quantized angular momentum l. Consequently the "permitted" energy levels of an atom are discrete, or at least a subset of the possible atomic energy levels are discrete, and they correspond to stationary states. This last concept constitutes the current conceptual importance of Bohr's model. The admission of having quantized quantities (see eq. 17.5) such as the energy levels or the angular momentum in microscopic systems is a revolutionary concept, that is naturally demonstrated by means of quantum mechanics. The difference between this model (or the Wilson-Sommerfeld model for quantized action integrals) and an authentic quantum theory, is in the fact that this first attempt introduces quantization *ad hoc* in a classical theory, while in quantum mechanics it arises naturally by solving the equations and enforcing boundary conditions.
The difficulties related to a complete and satisfactory explanation of the microscopic phenomena were emphasized by the fact that in some experiments the same physical quantities that showed a discrete nature, showed also a continuous nature, like the energy emitted by electrons in some well-defined conditions (e.g. the braking radiation of electrons). The first attempt in explaining why a fundamental property of a physical system, such as its energy, can sometimes appear discrete and sometimes instead continuous, was proposed by Louis de Broglie in 1924. He proposed an intrinsically "dualistic" nature, up to that time accepted for the electromagnetic radiation as mentioned above, also for "material" systems, i.e. microscopic systems of material particle. In particular de Broglie assumed that microscopic systems, up to that time seen as particles, could present also undulatory aspects. This is referred as the "wave-particle duality", and this is one of the fundamental conceptual achievements in quantum mechanics. More precisely de Broglie postulated that a particle of momentum $\vec{p}$ (with modulus p) can be associated to a wave, characterized by means of a wavelength λ, that analogously to what happens with electromagnetic waves and photons, was given by:

$$\lambda = \frac{h}{p} \tag{17.7}$$

This relation is known as "de Broglie relation". It can be rewritten in vector form like:

$$\vec{p} = \hbar\vec{k} \tag{17.8}$$

where $\vec{k}$ is the wave vector with a modulus $k = 2\pi/\lambda$. In light of this, the different behavior of an electron can be explained as follows. A free electron that is braked (e.g. by Coulomb interaction) behaves like a traditional particle and thus it can posses (and emit) a continuous range of energy. Instead if it is bonded to an atom, it shows its wave nature, and the only allowed energy states are those corresponding to stationary waves. Indeed, from the standpoint of de Broglie, the Bohr's assumption corresponds to state that the only permitted orbits are those that include an integer number of electron wavelengths. And exactly like an acoustic wave in a room is said to be steady if the dimension of the room is an integer multiple of its wavelength, if the electron orbit length (i.e. circumference in case of circular orbits like in the Bohr's model) is an integer multiple of the the electron de Broglie wavelength, it is a steady state, with no loss (neither gain) of energy (as assumed by Bohr). Now one may wonder: "what is thus an electron? a wave or a particle?". The answer is both or neither. Or better: in the wave-particle duality picture an electron is seen as a physical entity, that in some cases shows a particle-like behavior while in others a wave-like properties. In some experiments it is is well described by the classical concept of particle (with a given momentum and energy) and in others it is well described by the classical concept of wave (with a given wavelength and wave vector). The connection between the two descriptions is provided by the de Broglie relation (equations 17.7 and 17.8). In summary, the wave-particle duality states that with each particle is associated a wave field (sometimes called a matter field - not to be confused with waves of matter like acoustical or mechanical waves!). This is analogous to associate a photon to its electromagnetic field. The wave field describes the dynamical condition of a particle in the same sense that the electromagnetic field corresponds to photons which have precise momentum and energy. As already mentioned the connection between the wave field and the particle (e.g. electron) is provided by the de Broglie relation (eq. 17.7 and 17.8). Well, at this point a question may be: "what is the meaning of this wave field?". The answer is sometimes still controversial. According with the widely accepted and traditionally taught Copenhagen interpretation of quantum mechanics, the wave field has to be intended as a position probability field, as better discussed in section 17.3.1. In mathematical terms the wave field is represented by the so called "wave-function" (see section 17.3 and 17.3.1, 17.3.2 particularly) related to the electron or the microscopic entity. The wave-function ψ is a complex function of space and time that defines the wave field for the considered microscopic system:

$$\psi(x, y, z; t): \quad \mathbb{R}^4 \ \mapsto \ \mathbb{C}$$

Thus more precisely the quantity that is associated to the meaning of position probability of the particle is the wave-function intensity $|\psi(x, y, z; t)|^2$, indeed a probability must be a real quantity (refer again to section 17.3.1). In this optics it is expected that whenever the motion of a particle is disturbed in such a way that the wave field associated with it cannot propagate freely, interference and diffraction phenomena should be observed, exactly like it happens for elastic waves or electromagnetic waves. Nevertheless the physical meaning of this wave field is associated to the probability of position of the particle, thus a perturbation in the wave-function ψ leads to a different position probability of the particle. That is, whenever the particle motion is somehow perturbed and the wave field modified, there is a different probability of finding the particle in the different regions of space, according to interference phenomena (typical of waves!) that modify the wave-function ψ. In this, the mentioned wave field or matter field is extremely different from the classical concept of matter waves in which e.g. the molecules of air are subjected to a well defined undulatory motion because of the propagation of an acoustical wave in air.
The first experimental evidence of the wave-like behavior of electrons was thanks the experiments of Davisson and Germer and of G. P. Thomson starting from the 1927. In the

Thomson experiments a well collimated electron beam (of enough low energy to neglect electron-electron interaction) was sent through a thin crystalline film (see fig. 17.1 left) and then made impacting on a photographic plate. If the electron had behaved as classical particles a blurred image would have been observed focused around the center of the photographic plate, since each electron would undergo in general a different scattering by the atoms in the crystal. However the obtained result was analogous to the one obtained with X-rays diffraction by a polycrystalline substance. As evident in figure 17.1 (right), the result clearly showed a wave like propagation and diffraction of the electron beam by means of the crystalline film. Notice that in the moment in which an electron impacted on the photographic plate it generated a small point, like it were a "small particle", even if during its propagation it underwent to a wave-like propagating behavior (indeed there were favoured "trajectories" and forbidden ones). By doing all the calculations starting from the spacing between crystal planes, and considering the laws for evaluating interference, the recovered value for the electron wavelength was in agreement (with the limits of the experimental error) with the one postulated by de Broglie (eq. 17.7).

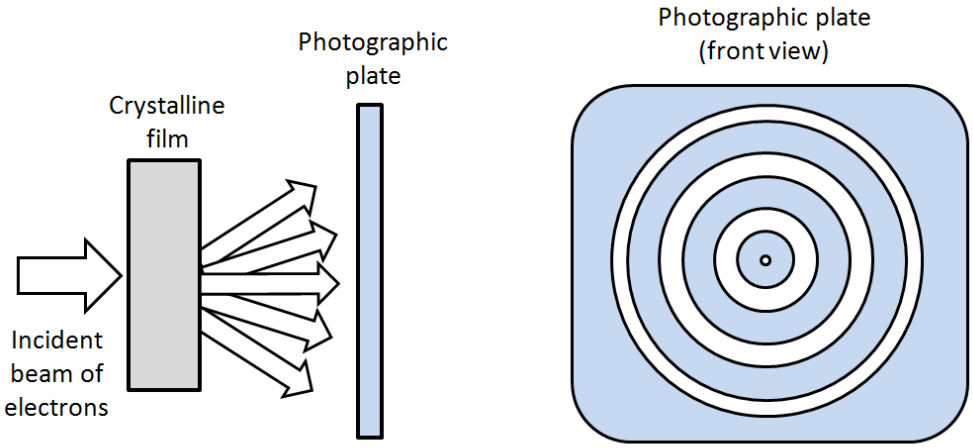


Figure 17.1: Experimental arrangement for observing electron diffraction through crystalline material (left); and diffraction patterns of electrons on the photographic plate (right).

In the C. Davisson and L. Germer experiments an electron beam was sent at a given angle toward a crystal face and the diffracted electrons were observed at a symmetrically located detector (see fig. 17.2). It was found that the electron current registered by the detector had a maximum every time the so-called Bragg condition (derived for X-rays) was fulfilled, i.e. only for given angles θ such that:

$$2d\ sin\theta = n\lambda$$

where d is the separation between successive atomic layers, n is an integer multiple and λ is the de Broglie wavelength. Again this experiment allowed to highlight the wave-like propagation behavior of electrons and their undulatory diffraction patterns as a result of the interaction with the crystal layers.
Other famous experiments are the ones that make electron pass through a single or a double slit in a shielding panel. In the single slit experiment a small circular micro-hole is present on a opaque screen behind which a photographic plate is present. Again in this case, if the electrons were classical particles, moving on rectilinear trajectories, a strong sign should be present on the photographic plate in correspondence of the slit, with blurred signs around it due to interaction of the electrons with the slit edge. Thus a sort of a Gaussian distribution of the intensity should be present. Instead, exactly like in the previously mentioned experiments, even in this case concentric rings (similar to the one of figure 17.1 right) are present,

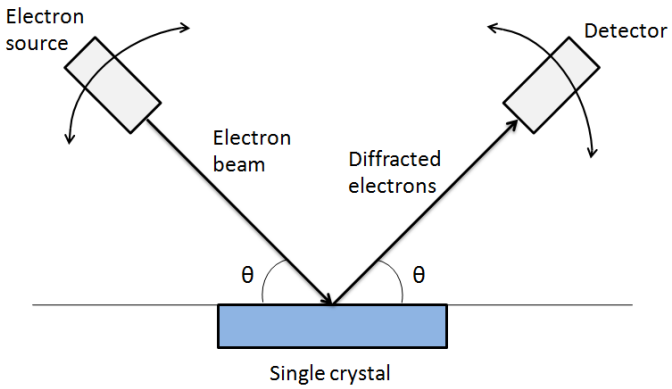


Figure 17.2: Davisson and Germer experimental arrangement for observing Bragg scattering of electrons.

showing the presence of "allowed" zones in which a large amount of electrons impact and "forbidden" zones in which no electron impacts. The diffraction pattern is again similar to the one obtained with diffracted light rays by a circular aperture.
The two slits experiment is analogous but two close slits are present. When only one of the two slits is opened the same result of the previous experiment is recovered, instead when both the slits are opened simultaneously an interference pattern is present on the photographic plate.
Once more the only possible explanation to these phenomena is that electrons (and in general whatever microscopic system - indeed analogous experiments with analogous results were carried on neutrons, protons and atoms) can posses a wave nature and thus they can interfere. Nevertheless notice that the wave nature of microscopic systems is not a collective phenomenon due to the interactions among them. Indeed the experiments described above lead to same results even if single electrons are sent one by one toward the photographic plate, and remember that before it was assumed that the electron energy was low enough to make negligible the interactions among them. Moreover also connecting a single electron with a single wave is generally wrong. Indeed, if it were so, a single electron would produce a diffraction figure on the photographic plate, eventually more marked if the number of electrons is increased. Instead as already mentioned a single electron produces on the photographic plate a single small spot, indicating that it is well localized (*a posteriori*) in the moment in which it impacts on it, showing its particle nature. Only when an enough large amount of electrons is considered the diffraction figure takes its shape. Again the impact regions and thus the propagating nature of electrons (or microscopic systems in general) are somehow not classical but linked with wave-like properties. And again everything becomes clear if the meaning of the wave field associated to electrons is considered: the intensity of the wave field, i.e. $|\psi(x, y, z; t)|^2$, must be interpreted as intimately related to the position probability amplitude, as mentioned above and discussed later in section 17.3.1. Thus to each electron is associated a well defined probability of impacting on a precise point of the photographic plate, and the waves are probability waves. Once again it should be clear now that wondering if an electron is a particle or a wave has no sense: it is a microscopic physical entity and thus it obeys to the physical laws that govern this world.

17.2 Particles, wave packets and indetermination principle

In the previous section it was reviewed the concept of wave-particle duality. Few words were said also on its physical interpretation and then some experiments (among the many) were briefly described to point out the correct interpretation of this revolutionary concept. In this section the main conceptual implications of this fact are considered.
A classical particle is a material object with a given spatial extension (that sometimes can be approximated like a dimensionless material point) and with a given trajectory. Instead a classical wave is a physical entity that shows opposite characteristics. A pure monochromatic wave (i.e. with a well define wavelength and thus in the de Broglie representation with a well defined momentum) has an infinite extension and it has no sense speaking about localization or trajectory. A "localization" of an undulatory system is possible only if wave packets are formed. A wave packet can be decomposed in the superposition of monochromatic waves of different frequencies (or wavelengths), and the greater is the number of superimposed monochromatic waves the smaller is the physical extension of the wave packet. This is essentially a Fourier decomposition (i.e. transform) of the wave packet. Indeed the wave packet of whatever waveform, or alternatively the pulse, can be considered localized if it presents an amplitude that is enough large only in a given region of space, such that it can be neglected outside of that region. In the wave-particle dual representation the wave field amplitude (or intensity) is linked to the probability of position of the system. Thus if the wave amplitude is non-negligible only in a given region of space, that is if the wave-function intensity $|\psi(x,y,z;t)|^2$ is non-negligible only in a given region of space, it means the probability of finding the system represented by ψ outside that region is essentially null, and thus the system is localized.
In other words a monochromatic wave is non-localized since its amplitude is constant in all the (infinite) space, see figure 17.3 (top). Nevertheless a monochromatic wave has a well defined wavelength λ, and thus from eq. (17.7) a well defined momentum p. Moreover a phase velocity can be associated to a monochromatic wave. By definition the phase velocity is the velocity that an observer should have to see a constant phase of the wave. From classical wave theory it is well known that a monochromatic one-dimensional (x-direction) wave can be described in mathematical terms like (harmonic wave):

$$wave(x,t) = A\ sin(kx - \omega t)$$

where $\omega = 2\pi f$ is the angular frequency and A the amplitude, or alternatively by the complex expression:

$$wave(x,t) = Ae^{i(kx-\omega t)}$$

where i is the imaginary unit.
The phase velocity ($v_{phase} = \frac{\omega}{k}$) of this monochromatic wave is obtained by differentiating the phase of the wave and by setting it to zero (since by definition the observer sees a constant phase its derivative is set to zero). Exploiting the equations (17.1) and (17.7) it is obtained (m is the particle mass or effective mass if the effective mass approximation is considered):

$$v_{phase} = \frac{\omega}{k} = \lambda f = \frac{h}{p}\frac{E}{h} = \frac{E}{p} = \frac{p}{2m} = \frac{1}{2}v_{particle} \tag{17.9}$$

where the kinetic energy of the particle is used. In general the total energy of the particle is given by the sum of kinetic T and potential energy U:

$$E = T + U(x) = \frac{1}{2}mv_{particle}^2 + U(x) \tag{17.10}$$

Here for simplicity $U(x) = 0$, thus:

$$E = \frac{1}{2} m v_{particle}^2 = \frac{p^2}{2m} \tag{17.11}$$

The phase velocity of a monochromatic wave field is equal to one-half the particle velocity, thus they are different. Nevertheless this has no particular implications since it is not possible to measure the phase velocity of a monochromatic wave directly (see [211] for details). The main point here is that a monochromatic wave has an amplitude (and an intensity) that is constant in all the space, thus it is non-localized, indeed (considering again the 1D example of before):

$$|\psi(x,t)|^2 = |wave(x,t)|^2 = \left|Ae^{i(kx-\omega t)}\right|^2 = |A|^2 = constant\ in\ x$$

Now it should be evident that a monochromatic wave field $\psi(x,t) = Ae^{i(kx-\omega t)}$ does not provide information about the localization in space of a particle.

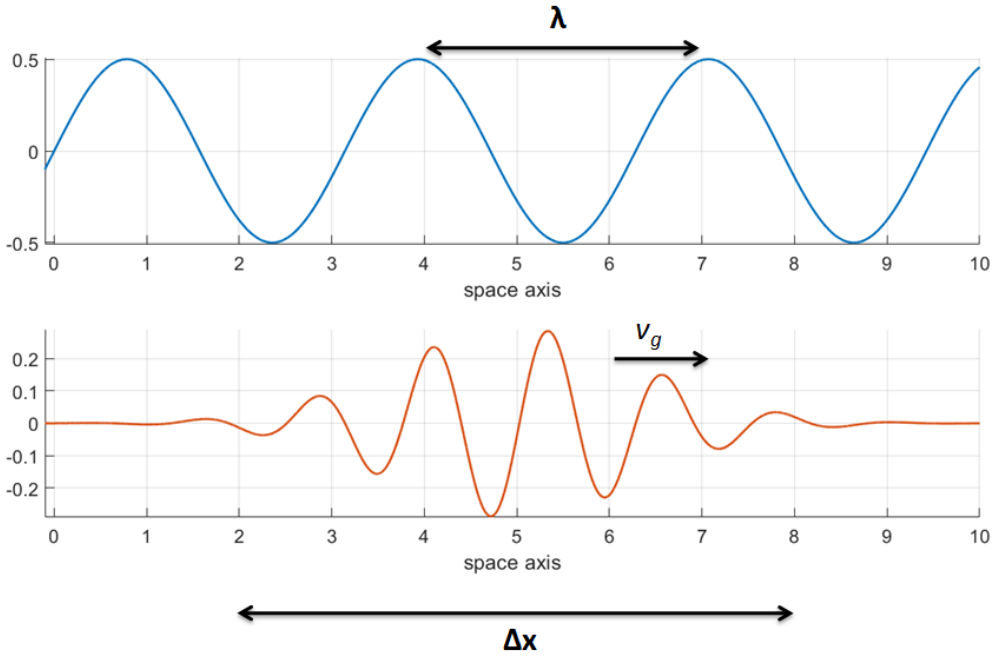


Figure 17.3: A monochromatic continuous wave train corresponding to an unlocalized particle (top); and an example wave packet corresponding to a localized particle within the distance Δx (bottom); λ is the monochromatic wavelength, v_g is the group velocity of the pulse envelope.

Instead a localized particle should be related to a wave field that has a large amplitude (or intensity) in the region of space in which the particle is localized and null or very small outside it. As already mentioned previously, a wave field with a large intensity only in a small region of space must be built by means of interference processes, by superposition of many monochromatic waves. Indeed a pulse can be localized, see figure 17.3 (bottom). It can be expanded as superposition of monochromatic waves by exploiting the Fourier transform (or series), thus it is effectively a wave-packet, intended as a set of monochromatic waves. In this regard think for example that the Fourier transform of a pulse (with a well define time -or spatial- support) presents a wide bandwidth in frequency domain that includes many different frequencies, while the spectrum of a monochromatic wave contains (ideally) only one frequency, i.e. it is a delta Dirac. From the basic physics and from the courses about signal processing and telecommunication theory, it is well known that the velocity at which a wave-packet travels (think e.g. to amplitude modulation schemes) is not

the phase velocity but instead the so-called group velocity, that is given by $v_{group} = \frac{d\omega}{dk}$ (a derivation can be found in [211] or in any basic classical wave physics or signal processing book). Since the wave-packet (i.e. the pulse) is associated to the particle position, its envelope velocity must be equal to the particle velocity, that indeed is:

$$v_{group} = \frac{d\omega}{dk} = \frac{dE}{dp} = \frac{d}{dp}\left[\frac{p^2}{2m}\right] = \frac{p}{m} \equiv v_{particle} \tag{17.12}$$

Now it is evident that a localized particle is associated to wave-field that is not monochromatic but given by a wave packet, whose group velocity corresponds to the particle velocity. A crucial conceptual point is that when an electron shows its particle-like nature, e.g. through a position measurement, nothing is known about its wave-like nature since this corresponds to a well localized wave packet, thus containing a large range of frequencies or wavelengths. In terms of Fourier transform a small space support (small region of space in which the probability of finding the particle is non-negligible) corresponds to a wide frequency bandwidth, with many frequencies (i.e. wavelengths) within it. With the de Broglie relation (eq. 17.7) this corresponds to a wide range of momenta p. Consequently an electron with well defined position has a non-defined momentum p. Vice versa whenever an electron shows its wave-like nature, e.g. through a momentum measurement, nothing is known about its particle-like nature since this corresponds to an unlocalized monochromatic wave, with a given value of p and thus a well-defined wavelength λ. This is the basic intuition behind the indetermination principle formulated in 1927 by Werner Heisenberg. It states that in a microsystem is not possible perform simultaneous measurements of position and momentum with arbitrary high accuracy. This limit is not related to measurement setup issues, but to the intrinsic nature of microscopic systems. This is a peculiar situation that is not preset in classical physics. Indeed in classical mechanics it is always possible to know with the desired accuracy both the position and the momentum of classical particles (within the measurement error and accuracy). Thus in classical physics it is possible to speak about trajectories, intended as the curve followed by classical particles for which the position and the velocity (or momentum) are simultaneously known. Instead in quantum mechanics, because of the wave-particle duality, in order to have a space localized wave packet it is necessary to superpose several monochromatic fields with different wavelengths λ, or with different wavenumbers k. Once again notice that if the Fourier transform of the wave packet (i.e. the pulse) is considered, the waveform result decomposed as a superposition of monochromatic waves, each with suitable amplitude (i.e. Fourier transform coefficient). Then recall also that the Fourier transform of a space coordinate x corresponds to a wavenumber k_x, and in 1D case the subscript can be omitted without confusion:

$$x \overset{\mathfrak{F}}{\longleftrightarrow} k$$

At this point if the wave packet extends over a region Δx, the values of the wavenumbers of the interfering waves that compose the wave packet and have appreciable amplitude fall within a range Δk such that:

$$\Delta k \sim \frac{2\pi}{\Delta x} \quad \Rightarrow \quad \Delta x\ \Delta k \sim 2\pi$$

Notice that the last expression follows directly from the theory of Fourier analysis. According with eq. (17.7) with $\lambda = \frac{2\pi}{k}$, different wavelengths λ or wavenumbers k mean that there are several values of p such that $\Delta p = \hbar \Delta k$, and therefore the previous expression becomes:

$$\Delta x\ \Delta p \sim h \tag{17.13}$$

The last expression constitutes a possible mathematical relation of the above introduced Heisenberg's indetermination principle (see section 17.5 for a more precise expression). Notice that it is a direct consequence of the Fourier transform analysis, as already pointed

out. The physical meaning of eq. (17.13) is the following: Δx and Δp should be interpreted as the variance (or eventually standard deviation) of the position and the momentum of the considered microscopic system. Since their product is essentially constant and proportional to the Planck's constant h if one of the two physical quantities is known with whatever precision (i.e. with null variance and standard deviation) then the other one is completely unknown. In other words if a particle is within the spatial region $x - \frac{1}{2}\Delta x$ and $x + \frac{1}{2}\Delta x$ (that is Δx is the uncertainty on its position), its associated wave field is represented by superposing monochromatic waves of momenta between $p - \frac{1}{2}\Delta p$ and $p + \frac{1}{2}\Delta p$, where Δp is related to Δx by means of eq. (17.13). The information about the localization in space is obtained at the expense of the information about its momentum, or velocity, thus as already said in quantum mechanics there is no more the concept of trajectory. It is not possible to accurately determine both the position and the momentum simultaneously ($\Delta x = 0$ would imply $\Delta p \to \infty$ and vice versa). The Heisenberg's uncertainty principle holds true in general for each pair of canonically conjugated variables (i.e. for all that variables that are connected by means of a time derivative operation), and not only for position and momentum. This again is a consequence of the Fourier transform properties:

$$\frac{d}{dt} \overset{\mathfrak{F}}{\longleftrightarrow} i\omega \tag{17.14}$$

where $\omega = 2\pi f$ and i is the imaginary unit (sometimes depending on the sign convention in the exponent in the Fourier transform definition the previous relation is written with a minus sign: $-i\omega$). Considering that $f = E/h$ it is easily recovered the previous result.
In addition to the position and momentum uncertainty relation and to all the canonically conjugated variables uncertainty relation, an interesting uncertainty relation needs a brief comment: the time-energy uncertainty relation. Indeed starting from eq. (17.14) it is possible to write:

$$\Delta t \; \Delta\omega \sim 2\pi \tag{17.15}$$

Again the physical meaning is suddenly clear when the Fourier theory is considered. As well known, if the time duration of a wave packet, i.e. the time support of a pulse, is very short in time, then to represent that pulse in terms of Fourier transform (or series) as superposition of monochromatic waves, a wide frequency bandwidth is necessary. Thus to define the exact time at which a particle passes through a given point, the pulse should in fact have a short duration Δt. But to build up such a pulse different harmonic waves with different frequencies are to be used. Then these harmonic waves will have an appreciable amplitude in the frequency range $\Delta\omega$ centered around the frequency ω. Multiplying both members of eq. (17.15) by $\hbar$, recalling that $E = hf = \hbar\omega$ and $2\pi\hbar = h$ the energy-time uncertainty relation is found:

$$\Delta t \; \Delta E \sim h \tag{17.16}$$

The uncertainty relation (17.16) needs a remark on its physical interpretation and another relative to its formal correctness. The physical meaning of equation (17.16) should be interpreted as related to unstable or excited systems. Equation (17.16) states essentially that in order to determine the energy E of a system with an uncertainty $\sim \Delta E$ a time of the order of Δt is necessary. This means that in order to measure the energy of a stable system with a perfectly defined energy it should be necessary an infinite time. Thus a correct interpretation of eq. (17.16) is in terms of lifetime of excited states. For example if an electron in an atom for some reason is excited and goes in an excited state, then after some time it will suffer a radiative transition by means of which it looses its excess energy and goes into a stationary state, namely the ground state. Unfortunately it can be shown that there is no mean of predicting exactly how long the electron will remain in the excited state before it makes the transition. At most it is possible to speak about transition probability per unit time that the electron jumps back to the ground state. Therefore the average time

interval the electron stays in the excited state is named lifetime of that state (it is inversely proportional to the transition probability), and it can be known with an uncertainty of the order of Δt. Hence the energy of the excited state cannot be known exactly, but only within an uncertainty range of the order of ΔE. The shorter is the lifetime (and thus Δt) the larger is the uncertainty on the excited state energy ΔE. For the ground state, whose lifetime is infinite (since the system is stable and cannot suffer of a transition toward a state of lower energy) the lifetime and thus Δt is infinite, while its energy can be known with the desired accuracy ($\Delta E \to 0$).
Coming to the formal correctness of eq. (17.16), a good discussion is provided in [210]. In extreme summary the fact is that despite the name "principle" the Heisenberg's indetermination principle can be demonstrated and indeed it a theorem. As already mentioned the demonstration holds for all the canonically conjugated variables. Nevertheless the total energy of a system E and the time t are not canonically conjugated variables (the total energy of a system cannot be determined as the time derivative of the time - it has no meaning). Formally this trouble comes out because the time in quantum mechanics is not considered a "physical observable" but instead like a free parameter. In quantum mechanics, as will be clarified in section 17.3.4, to each physical observable a specific operator is associated, but since the time is a parameter it is not an operator. Thus if one uses the same procedure used for the demonstration of the Heisenberg's indetermination principle a conceptually wrong relation is recovered. Nevertheless it is possible to formally demonstrate the time-energy uncertainty relation proceeding with slightly different arguments. Thus a similar expression to eq. (17.16) can be formally showed using non-relativistic (let's say "standard") quantum mechanics, and its physical interpretation is conventionally the one I have reported above.

17.3 The Schrödinger's equation

Classical mechanics was developed for describing the detailed motion of macroscopic bodies under the assumptions that these bodies are localized in space and can be observed without appreciably disturbing their motion. These assumptions are in general made implicitly rather than stated in a precise way, but they are at the basis of the classical formalism. As briefly summarized in the previous sections 17.1 and 17.2, it is no more possible talking about localization in a classical way when microscopic entities are considered. Moreover measurements on some physical observable related to microscopic systems show the evidence that such quantities (like the energy levels of atoms and molecules) are quantized, and this is again unexpected with classical mechanics. Quantum mechanics is instead capable of catch all these physical insights. The first issue to address in a novel physical theory of mechanics is how to represent the state of the system under study and then how to determine the dynamics of the system state, i.e. its time evolution. The state of the system is represented by means of the so called "wave-function" $\psi(x, y, z; t)$:

$$\psi(x, y, z; t): \quad \mathbb{R}^4 \mapsto \mathbb{C}$$

this complex function represents the wave field associated to each microscopic particle or system. It is a function of space (in the three variables x, y, z) and time t. Its meaning and its properties will be described more in details in sections 17.3.1 and 17.3.2 respectively.
The time evolution of the system described by ψ can be determined by means of the Schrödinger's equation, firstly introduced by Erwin Schrödinger in 1926:

$$i\hbar\frac{\partial}{\partial t}\psi(\vec{r}; t) = \widehat{H}\psi(\vec{r}; t) \tag{17.17}$$

where:

– i is the imaginary unit

- $\hbar$ is the reduced Planck's constant (already introduced)
- $\vec{r} = (x, y, z)$ is the position vector in the 3D space
- $\frac{\partial}{\partial t}$ indicates the derivation operation w.r.t. the time t
- $\psi(\vec{r}; t)$ is the wave-function that represents the system state at the time instant t
- $\widehat{H}$ is the so called Hamiltonian (quantum mechanical) operator

Before going on few words concerning the Hamiltonian operator are needed. In quantum mechanics to each physical observable (such as the energy, the position, the momentum, the angular momentum and so on...) a mathematical operator is associated. By making use of such mathematical operator (that can be e.g. a differential operator) it is possible to make previsions and calculations about that specific physical quantity and then predict the set of its possible values. A more detailed introduction to quantum mechanical operator (and their origins) is reported in section 17.3.4. In classical rational mechanics the system total energy E, given by the sum of the kinetic energy T and the potential energy U, is equal to the so called Hamiltonian H of the system (this holds if the body constraints are time-independent, otherwise the Hamiltonian H is a scalar function of the system total energy). Thus $H = T + U = E$ corresponds to the total energy of the system. Since as already mentioned in quantum mechanics to each physical observable a mathematical operator is associated, to the total energy of the system an operator, i.e. the Hamiltonian operator, is associated. Generally to distinguish the operator from the observable the symbol $\widehat{\ }$ is positioned on top. Thus $\widehat{H}$ is the quantum mechanical operator associated to the total energy of the system, i.e. the system Hamiltonian, and it can be used to find the possible set of energy values (from here on indicated with E) for that specific system. In section 17.3.4 this concept will be further discussed.

The Schrödinger's equation is the most important equation in (non-relativistic) quantum mechanics. As already pointed-out it provides the dynamics of the system, i.e. its time evolution. Starting from the knowledge and the hypothesis about the system under study it is possible to derive an explicit expression for the Hamiltonian operator $\widehat{H}$, that is known. Then by solving the Schrödinger's equation the wave-function is obtained, that is: the system state (represented by ψ) is found by solving the Schrödinger's equation. More precisely equation (17.17) is referred as the time-dependent Schrödinger's equation, and indeed it provides the time evolution of the system (ψ in eq. (17.17) is function of time). Starting from the knowledge of the system state at initial time instant t_0, i.e. starting from $\psi(\vec{r}; t_0)$, the state $\psi(\vec{r}; t)$ at whatever time instant t can be found by solving the Schrödinger's equation (17.17). The general solution of eq. (17.17) will be addressed in section 17.3.5.

The Schrödinger's equation was the fundamental starting point for the development of a complete quantum theory. Indeed the quantization of some physical properties (such as the energy) arises naturally as a consequence of the enforced boundary conditions used in the solution of the Schrödinger's equation for a specific system (and the same quantity -e.g. the energy- results quantized or continuous depending exactly on the specific boundary conditions), contrarily to the initial attempts briefly discussed in section 17.1, in which the quantization was introduced *ad hoc*.

The Schrödinger's equation is often compared for importance and meaning to the well known Newton's second law of dynamics, i.e. $\vec{F} = m\vec{a}$ (where $\vec{F}$ is the total force acting on the body of inertial mass m and $\vec{a}$ is its acceleration); and analogously to it there exist no demonstration of the Schrödinger's equation. This equation was the result of the exceptional intuition of Erwin Schrödinger and at most it is possible to provide some "arguments of plausibility" that are aimed in intuitively justify it. In this regard a good treatment is provided in [210] (in Italian) or in [214] (in English). Here I only point out that the purpose of the Schrödinger's equation is to provide the wave field $\psi(\vec{r}; t)$ that is associated to a microscopic system, and thus some similarities with a conventional wave equations are expected. In particular embedded within the Hamiltonian operator $\widehat{H}$ there is a space second order derivative (see section 17.3.4 for details), and in eq. (17.17) a

first order derivative in time appears. In a classical, let's say standard, wave equation it is expected to have also a second order derivative in time, nevertheless this is what distinguish the Schrödinger's equation from a standard wave equation. The reason behind the first order time derivative is linked to the validity of superposition principle for the wave-function. Indeed as clarified in section 17.3.2 a fundamental physical requirement for the wave-function ψ is that the superposition principle holds, i.e. if ψ_A and ψ_B are possible states for a microscopic system, then a linear combination of them must be again a state for that system. This requires firstly the Schrödinger's equation to be linear (as it is, since derivative operators are linear), and then to have non-dynamic coefficients (if dynamic coefficients were present different wave-functions corresponding to different coefficient values would be solutions of different equations thus making the superposition principle no more valid). The latter requirement implies a first order time derivative instead of a second order one in eq. (17.17).

17.3.1 Wave-function physical meaning

The first implication of the wave-particle duality (section 17.1) is that it is not possible to talk about trajectories for microscopic system because of the Heisenberg's uncertainty principle (section 17.2) that states the impossibility of knowing at the same the position and the momentum of a microscopic particle. Moreover the revolutionary fundamental concept of the wave-particle duality is that to each particle is associated a wave field, that is somehow represented by means of the complex field $\psi(\vec{r};t)$. The purpose of this section is to definitively clarify the physical meaning of the wave-function, i.e. of $\psi(\vec{r};t)$, that is associated to a microscopic particle.
The wave-function $\psi(\vec{r};t)$ precisely represents the amplitude of the wave field (sometimes called matter field [211]) that is associated to a particle. From the mathematical standpoint a "field" is a function of more variables whose domain and codomain are the same, e.g. in physics very often four variables are used (three for the space coordinate and one for the time coordinate) thus for example $f : \mathbb{R}^4 \mapsto \mathbb{R}^4$. A scalar field is instead a scalar function of more variables. For example the temperature, that links a scalar value to each point in space at each time, is a scalar field: $temperature : \mathbb{R}^4 \mapsto \mathbb{R}$. In the case of the wave-function:

$$\psi(x,y,z;t) : \quad \mathbb{R}^4 \mapsto \mathbb{C}$$

It means that a complex scalar value ψ is associated to each point of space at each time instant; thus ψ is a (complex) scalar field. The complex value ψ has to be intended as the amplitude of the wave field that is associated to a given microscopic system. From classical theory of waves it is well known that the intensity of a wave (linked to its energy or power) is proportional to the square of its amplitude. By analogy in quantum mechanics the intensity of the wave field is $|\psi(\vec{r};t)|^2 = \psi(\vec{r};t)\psi^*(\vec{r};t)$, since the wave-function already corresponds to the wave field amplitude. In this case the wave-function is in general a complex function of space and time, and its modulus squared must be considered, since the intensity of a wave field is by definition a real scalar quantity.
The physical meaning of the wave-function and of its modulus squared, i.e. the physical meaning of the intensity of the wave field associated to microscopic entities, becomes then clear if the experiments on electrons described in section 17.1 are considered. In the moment in which it is accepted that the electrons impacting on the photographic plate are described by a wave field, it is possible to associate the intensity of the electrons tracks on the photographic plate to the wave field intensity, i.e. $|\psi(\vec{r};t)|^2$. Where the wave field is more intense, $|\psi(\vec{r};t)|^2$ is larger, and thus $|\psi(\vec{r};t)|^2$ must be somehow proportional to the distribution of the electrons on the photographic plate. Notice at this point that $|\psi(\vec{r};t)|^2$ does not describe a collective property of the electrons, indeed as mentioned in section 17.1 the same interference patterns are obtained also sending one electron at a time. Moreover each electron hits the photographic plate in one specific point (leaving a well define point-

like sign), thus the wave field cannot be interpreted as a matter wave, in which there is a collective motion of the particles involved (think for example to a mechanic wave in a string, the particles of the string oscillate in a well defined way but the phenomenon is obviously extremely different from the one noticed for electrons). Consequently this wave field should be considered like a wave field of probability. The probability should then be interpreted as probability of position, i.e. of finding the microscopic system in that specific point at that specific time. Consequently the intensity of the wave field, that is its amplitude squared, i.e. $|\psi(\vec{r};t)|^2$, should be somehow proportional to the probability of finding the microscopic system (e.g. the electron) in that place at that time. More precisely the wave-function modulus squared is conventionally interpreted as the position probability density (i.e. probability per unit volume) of the system that it describes. For this reason the wave-function is also called "amplitude of position probability".
To be more quantitative: the probability of finding the microscopic system described by the wave-function $\psi(\vec{r};t)$ in the (3D) neighbourhood $d\vec{r}$ around the point $\vec{r}$ at the time instant t is given by: $|\psi(\vec{r};t)|^2 d\vec{r}$. Where in this notation (similar to one of [210]) $d\vec{r} = dxdydz$. In summary, $|\psi(\vec{r};t)|^2 d\vec{r}$ is interpreted as the probability of finding the system between $\vec{r} = (x, y, z)$ and $\vec{r} + d\vec{r} = (x + dx, y + dy, z + z)$ at the time instant t. The total position probability can be recovered integrating over the entire volume V the probability density:

$$P = \int_V |\psi(\vec{r};t)|^2 \, d\vec{r} \tag{17.18}$$

where P corresponds to the probability of finding the system described by ψ within the volume V, and as already mentioned $|\psi(\vec{r};t)|^2$ is the probability per unit volume.
This interpretation of the wave-function squared modulus like a probability density is universally accepted and it is the conventional interpretation provided in all basic courses of quantum mechanics [210]. It is usually referred as the "Copenaghen" interpretation of quantum mechanics, mainly due to the works, done in Copenaghen, by Bohr and Heisenberg during the 1927 and after the second world war by Wolfgang Pauli, even if the first idea of associating to the intensity of the wave field the meaning of probability density is due to Max Born. Other interpretations of quantum mechanics are possible, and even if their treatment is outside the scope of this work, I would like to briefly mention the so called "statistic" interpretation of quantum mechanics. In this interpretation the probability associated to the wave field has to be interpreted like applied to a set of systems or particles but not to a single system or particle. In this interpretation the amplitude of the wave field has a more abstract physical meaning related to the statistics of obtaining a given configuration (e.g. a given diffraction pattern on the photographic plate) that is represented by the wave-function. The difference w.r.t. the Copenaghen interpretation can appear subtle at a first read, but the main diversity is exactly that in the Copenaghen interpretation the wave field meaning of probability is associated also to a single particle. In the statistic interpretation the probability arises from the fact that there is an intrinsic limit in the knowledge of microscopic system, that nevertheless behave in a deterministic way. Notice that the limit is intrinsic and not related to instrumentation limits (see section 17.2). If there were not this intrinsic limit then a deterministic description would be possible. Instead in the Copenaghen interpretation the probabilistic meaning of the wave field is intrinsic. To better clarify this point consider the two slits experiment presented at the end of section 17.1. In this experiment when a single slit is open the diffraction pattern of figure 17.1 is recovered, instead when both are open a different shape is obtained, that evidently presents interference patterns. The electrons in such an environment are described by two wave-functions ψ_A and ψ_B depending if the electron passes through one slit (A) or the other (B). The two wave-functions ψ_A and ψ_B alone corresponds exactly to the diffraction patterns of figure 17.1, correctly aligned to the slit A or B. When both slits are opened then the electrons, that can pass through one or the other slit, are described in general by a wave-function ψ_C that is the superposition of the wave-functions of before: $\psi_C = \psi_A + \psi_B$. This effectively corresponds to the interference pattern that appears on the photographic

plate when both the slits are opened. Thus if ψ_A and ψ_B are solution of the Schrödinger's equation (17.17) for that system, also ψ_C must be solution of the same equation, i.e. the superposition principle must hold. This is exactly the reason why the Schrödinger's equation is linear and with non-dynamic coefficients, as discussed previously. And since the Schrödinger's equation is linear and with non-dynamic coefficients ψ_C is actually solution for the Schrödinger's equation. The electron state before it impacts on the photographic plate, in the case in which both the slits are open, is represented by $\psi_C = \psi_A + \psi_B$, i.e. it is the sum of two states. In the Copenaghen interpretation the electron, before impacting onto the plate, is actually in two different states simultaneously, i.e. in both states $\psi_A + \psi_B$. In the moment in which a detector is used to understand toward which slit the electron is passing, then there happens the so called "collapse" of wave-function, that means that electrons are no more in the state ψ_C but they collapse in one of the two states ψ_A or ψ_B that constitute ψ_C. The collapse of the wave-function in the Copenaghen interpretation is actually intended as a change of state of the electrons, that from the state ψ_C collapse to the state ψ_A (or ψ_B) because of the measurement process. The measurement procedure in microscopic system always interact strongly with the system and force this transition of state. Instead in the statistic interpretation of quantum mechanics, writing $\psi_C = \psi_A + \psi_B$ is a way to quantify that some electrons pass through A, and thus are represented by ψ_A and others trough B (represented by ψ_B), but since it is not possible to know toward which aperture they pass the description is given by $\psi_C = \psi_A + \psi_B$. The state ψ_C is then a sort of collective state, that is the result of the experiment only when an enough large number of electrons are considered, while the single electron will not be in two states (ψ_A and ψ_B) at the same time but only in one of them (ψ_A or ψ_B). In this picture the electron wave-function collapse is not intended as a state transition from ψ_C to ψ_A or ψ_B but is intended as the revelation of the actual state of that specific electron, that can be either ψ_A or ψ_B.

As mentioned many other interpretations are possible and they are not discussed here. The difference between these two should have been useful in highlight the widely accepted Copenaghen interpretation. The point is that all such interpretations are a boundary matter between physics and metaphysics (or philosophy), and indeed they are the subject of many science philosophy debates.

Independently of what is the correct interpretation of quantum mechanics, and thus of what are the features of the world we are living in, the practical importance of this theory is that it provides the means to predict the behavior of microscopic systems and to design novel devices based on them. In the following of this work the Copenaghen interpretation will be considered.

So far only the modulus of the wave-function was considered. Since it is a complex function it has also a phase. The wave-function phase is useful in interference phenomena, for example when more states are superimposed, to correctly find the total state (like done previously with the two slit experiment). The interference patterns between wave-functions are indeed the results of the phase mismatch between different wave-functions.

Notice again that the wave-function (in the Copenaghen interpretation) describes a single microscopic system and not a statistic set of systems. Nevertheless also in the Copenaghen interpretation it is possible to consider the wave-function as representative of statistic properties of a set of identical systems, at least if the number of considered identical systems is enough large. For example if a set of $N_A \sim 6 \cdot 10^{23}$ (Avogadro's number) of identical systems is considered, and the wave-function has a probability of 10% to find the single system in a given volume, then it is possible to say that volume contains the 10% of the considered systems.

A final remark is now provided by means of an example. An atom is a quantum system, and it is known form basic chemistry that it is constituted by a nucleus of protons and neutrons around which the electrons are present, like they were in "orbit" around the nucleus. Since it is not possible to talk about trajectories (and thus orbits), and keeping in

mind the wave-function meaning explained above, the wave-functions that are solutions of the Schrödinger's equation for an atom, are to be intended as the amplitude of the position probability wave field. Thus their moduli squared, when integrated over the space, provides a surface in the 3D space that corresponds to the high probability region. The higher is $|\psi|^2$ the higher is the probability of finding the electron in that space region. In general the wave-function modulus squared are the so called atomic orbitals (usually indicated with $1s$, $2s$, $2p_x$, $2p_y$, $2p_z$, etc...), usually presented in the basic chemistry courses. It is well known that the $1s$ orbital has the shape of a sphere around the atomic nucleus. Well, that sphere is the region in which it is most probable to find the electron that is in that specific state (with that specific energy, angular momentum modulus and orientation and spin). In this optics atomic and molecular orbitals correspond simply to high probability regions, in which it is highly probable to find that electrons.
Finally notice that the wave-function describes the microscopic system and it contains all the information that is possible to get about the system.

17.3.2 Wave-function properties

Once understood the physical meaning (Copenaghen interpretation) of the wave-function as the amplitude of the position probability wave field associated to the microscopic system (see section 17.3.1), a set of properties that the wave-function must satisfy is immediately derived. In particular the probability of finding a given particle in all the space must be 100% (if the particle exists it must be inside the universe). Then it follows immediately that:

$$P = \int_{All\ space} |\psi(\vec{r};t)|^2 \, d\vec{r} = 1 \tag{17.19}$$

This condition is called the "normalization condition" [211]. Indeed it imposes a limitation on the possible forms of the wave-function ψ since it is not always possible to satisfy eq. (17.19). In particular ψ must decrease rapidly when the coordinates x, y, z are large in order to ensure the integral to exist. More precisely the wave-function must be a square-integrable function (or quadratically integrable function), that means:

$$\int_{-\infty}^{+\infty} |\psi(\vec{r};t)|^2 \, d\vec{r} < \infty \tag{17.20}$$

The usual notation for indicating a square-integrable wave-function over the space region V is: $\psi \in L^2(V)$. If it is supposed to have an electron under test, it can be considered like certainly confined in the laboratory in which the experiments takes place, thus said V the volume of the laboratory, the normalization condition becomes:

$$P = \int_{V} |\psi(\vec{r};t)|^2 \, d\vec{r} = 1 \tag{17.21}$$

and the wave-function must be square-integrable in the domain V: $\psi \in L^2(V)$.

Example 2.1: Let's consider for example a 1D wave-function (where K is a real constant, $K \in \mathbb{R}$):

$$\psi(x;t) = \begin{cases} K & , \ if\ 0 < x < x_0 \\ 0 & , \ otherwise \end{cases}$$

The normalization condition implies that:

$$\int_{-\infty}^{+\infty} |\psi(x;t)|^2 \, dx = 1 \quad \Rightarrow \quad \int_{0}^{x_0} |K|^2 \, dx = 1 \quad \Leftrightarrow \quad |K|^2 \, x_0 = 1 \quad \Leftrightarrow \quad K = \frac{1}{\pm\sqrt{x_0}}$$

Thus assuming for example the positive sign solution the normalized wave-function, to

which it is possible to associate the meaning of position probability density is:

$$\psi_N(x;t) = \begin{cases} \frac{1}{\sqrt{x_0}} & , \quad if\ 0 < x < x_0 \\ 0 & , \qquad otherwise \end{cases}$$

such that:

$$\int_{-\infty}^{+\infty} |\psi_N(x;t)|^2\, dx = \int_0^{x_0} \left|\frac{1}{\sqrt{x_0}}\right|^2 dx = \frac{1}{x_0} x_0 = 1$$

where the subscript N is used to point out that ψ_N is the normalized wave-function. □

Other direct consequences of the (normalized) wave-function physical interpretation as probability density are that it must be a monodromic function (i.e. it provides a single complex scalar value and not a complex vector) since it has no sense talking about more than one probability values of finding the particle in a given point. It must be everywhere limited (the probability cannot be infinite in a point or region of space) and that it must be continuous (if a 1D wave-function in space is considered, it has no meaning talking about a left-side probability or right-side probability around a point x_0 in which a discontinuity is present, thus the left and the right limits should be equal and the wave-function is continuous). Another property that is possible to demonstrate is that the wave-function must also have continuous first derivatives. The latter property follows from a continuity equation (analogous to the well known continuity equation for the electrical current), that is intimately linked with the concept of continuity of the total probability. In few words the fact the total probability must be conserved (i.e. if at a given time instant t_0 there is a particle in a given region of space, at another time instant t_1 the particle does not disappear if no particular interactions occur, thus the integral over all the space of the wave-function even at t_1 must be unitary, that corresponds to a total probability conservation) implies that the total probability must satisfy a continuity equation, and thus must have a continuous first derivative.
The last two properties, i.e. the wave-function must be continuous with continuous first derivative, imply that the wave-function must be of class C^1: $\psi \in C^1(V)$, where V is the volume of definition of the wave-function. From the mathematical standpoint this is also a direct implication of the fact that $\psi \in L^2(V)$. Indeed it is possible to show that:

$$\psi \in L^2(V) \quad \Rightarrow \quad \psi \in H^2(V) \quad \Rightarrow \quad \psi \in C^1(V)$$

where H^2 is the so called Sobolev space. This last relation is often seen in the numerical methods courses.
The last important property to be satisfied by wave-function is the superposition principle. This is a direct consequence of the de Broglie assumption of wave-particle duality. Indeed in order to localize a particle in space it is necessary to consider a wave-packet, as described in section 17.2. A wave-packet is obtained thanks to the interference of monochromatic waves, i.e. it is constituted by a superposition of wave-functions, thus the superposition principle must hold. This point was also clarified in the previous section 17.3.1 when the two slits experiment was considered as an explicative example for the wave-function physical meaning. Two are the equivalent statements that summarize the validity of the superposition principle. First, wave-functions that differ only for the normalization constant (i.e. they are different only because of a different multiplicative constant) represent the same state. So the two wave-functions ψ and ψ_N considered in example 2.1 are representing the same physical state.
Second, if a system can be in a state described by the wave-function ψ_A and also in another state described by the wave-function ψ_B (see e.g. the two slits experiment), it can also be in whatever state that is a linear combination of the two:

$$\psi = \alpha\psi_A + \beta\psi_B \quad , \quad \alpha\ ,\ \beta\ \in \mathbb{C} \tag{17.22}$$

The last statement can be rephrased as: the set of all wave-functions that describe all possible physical states of a given quantum system constitute a vector space.
In this regard notice that if a system is in the state $\psi = \alpha\psi_A + \beta\psi_B$, the result of measurement on it can be either ψ_A or ψ_B, and it is said (as already pointed out) that the wave-function ψ collapse into either ψ_A or ψ_B.
As already pointed out in the previous sections the Schrödinger's equation is totally linear and with non-dynamic coefficients. This ensures the validity of the superposition principle (and indeed the Schrödinger's equation was "built" up to satisfy it).
In summary the properties of a wave-function are the following:

- it must be square-integrable: $\psi \in L^2(space)$.
- it must be normalized such that it can be interpreted as position probability density, the normalization condition is given by eq. (17.19).
- it must be a monodromic function (i.e. it provides a single -complex- scalar value and not a vector, the codomain is $\mathbb{C}$).
- it must be everywhere limited.
- it must be continuous.
- it must have continuous first derivatives.
- the superposition principle holds.

17.3.3 Expected values and momentum space

At the end of section 17.3.1 was pointed out that the wave-function contains all the information that is possible to get about the system. The purpose of this section is to further discuss this statement, by giving an introduction to the calculation of average values in quantum mechanics. More precisely the question addressed in this section is "what is the information that can be extracted from the wave-function?". First of all a fast review of average value and standard deviation is provided.

Expectation values review

From an experimental standpoint, from basic bachelor's level courses on physics and statistics, the expected value $\langle F \rangle$ of a quantity F can be defined by doing a series of N measurements performed on identical systems in the same conditions (and thus described by the same wave-function). The arithmetic mean of the obtained measurement results is the average value:

$$\langle F \rangle = \frac{1}{N} \sum_i n_i f_i = \sum_i \nu_i f_i \tag{17.23}$$

where it is intended that during the N measurements of F, the value f_i was obtained n_i times, and where $\sum_i n_i = N$. The frequency at which the result f_i appears is then $\nu_i = \frac{n_i}{N}$. If the number of total measurements N is very large then the frequency ν_i can be interpreted as a probability (see for example [215]): $\nu_i \rightarrow P_i$ when $N \rightarrow \infty$. Thus equation 17.23 can be rewritten as:

$$\langle F \rangle = \frac{1}{N} \sum_i f_i P_i \tag{17.24}$$

The last relation holds for system in which quantization appears and thus discrete values of the quantity F are possible. It corresponds to the case of a discrete random variable [215]. If the quantity F can assume a set of continuous values then the previous equation is modified as follows (F is considered in the same way as a continuum random variable):

$$\langle F \rangle = \int f P(f) df \tag{17.25}$$

where $P(f)$ is the conventionally called probability density function, and $P(f)df$ has the meaning of differential probability, i.e. of probability of finding a result between f and $f+df$ when a measurement on F is performed. Finally recall that the definition of standard deviation (intended as the dispersion of the obtained results f when measurements on F are performed) is the following:

$$\Delta F = \sqrt{\langle F^2 \rangle - \langle F \rangle^2} \tag{17.26}$$

where the term under square root is called variance. The bigger is ΔF the more the measured values are different among them. Notice that it corresponds to the notation used in section 17.2 for indicating the standard deviation of position, momentum, energy and time.

Expectation values in quantum mechanics

Coming now to the information that can be extracted from the wave-function, and considering the probabilistic meaning of wave-function itself, it is intuitive that a statistical information can be extracted from it. Indeed in quantum mechanics it is possible to evaluate the expected value $\langle F \rangle$ of each physical observable F starting from the knowledge on the wave-function that represents the state of the system. In fact the wave-function squared modulus $|\psi|^2$ corresponds exactly to the probability density function $P(f)$ since it is interpreted as the position probability density, and $|\psi|^2 d\vec{r}$ corresponds to the differential probability $P(f)df$ (see section 17.3.1).
Thus the expected value for the position vector $\vec{r}$ can be expressed as:

$$\langle \vec{r} \rangle = \int \vec{r} \, |\psi(\vec{r};t)|^2 \, d\vec{r} \tag{17.27}$$

in which the integral corresponds to a continuum sum over all the possible values of $\vec{r}$ (usually they coincide with all the possible points in space), each multiplied by its probability: $|\psi(\vec{r};t)|^2$. Then $|\psi(\vec{r};t)|^2 \, d\vec{r}$ is exactly the differential probability of before, i.e. the probability of finding $\vec{r}$ in the infinitesimal volume in between $\vec{r}$ and $\vec{r} + d\vec{r}$. Notice that eq. (17.27) is a vector equation that embeds three scalar equations in the three variables x, y, z. In general it is possible to show that whatever function $F(\vec{r})$ of $\vec{r}$ has an expected value that is:

$$\langle F(\vec{r}) \rangle = \int F(\vec{r}) \, |\psi(\vec{r};t)|^2 \, d\vec{r} \tag{17.28}$$

indeed $|\psi(\vec{r};t)|^2 \, d\vec{r}$ corresponds to the probability density function of position, and a function $F(\vec{r})$ involves only the probability of position. In other words the probability density function of a function of the position vector corresponds to the position probability density. This point is clarified in the following example.

Example 2.2: The average value of the function $F(x, y, z) = xy$ is given by:

$$\langle xy \rangle = \int xy \, |\psi(\vec{r};t)|^2 \, d\vec{r} \tag{17.29}$$

indeed $|\psi(\vec{r};t)|^2 \, d\vec{r}$ is the probability of finding x at the time instant t between x and $x+dx$ but also the probability of finding y between y and $y + dy$. □

Equation (17.28) can be rewritten as:

$$\langle F(\vec{r}) \rangle = \int \psi^*(\vec{r};t) F(\vec{r}) \psi(\vec{r};t) d\vec{r} \tag{17.30}$$

Usually indicated symbolically as:

$$\langle F(\vec{r}) \rangle = (\psi, F\psi) \tag{17.31}$$

where (f, g) indicates the functional scalar product between the two functions f and g, that in the simple 1D case (i.e. $f(x)$ and $g(x)$) is defined by:

$$(f, g) = \int f^*(x)g(x)dx \tag{17.32}$$

The wave-function physical meaning is the one of position probability density. For this reason it is suitable for the definition of expected values of all those quantities $F(\vec{r})$ that are function of the position $\vec{r}$. Nevertheless it is not suitable for the definition of expected values that are function of the momentum $\vec{p}$, or alternatively of the wave vector $\vec{k}$ (linked to $\vec{p}$ by means of the de Broglie relation: $\vec{p} = \hbar\vec{k}$). In order to estimate the expected values of a whatever function $F(\vec{k})$ of the wave vector (or momentum) it is necessary to consider as probability density function $P(f)$ in equation (17.25) a suitable probability density. In particular in analogy to what done for a general function of the position, the momentum probability density must be considered, i.e. the probability of finding the particle momentum between $\vec{p}$ and $\vec{p} + d\vec{p}$. The last statement correspond to finding the wave vector between $\vec{k}$ and $\vec{k} + d\vec{k}$.
Fortunately the momentum probability density can be recovered by means of the Fourier transform as it will clear in a while.

Position space and momentum space

It is well known that the Fourier transform generate a correspondence between time t and frequency ω: the Fourier transform of a signal that is a function of time $s(t)$ is a function of frequency $S(\omega) = \mathfrak{F}\{s(t)\}$, where $\mathfrak{F}$ indicates the Fourier transform. This correspondence is often indicated as:

$$t \overset{\mathfrak{F}}{\longleftrightarrow} \omega \quad \textit{or alternatively}: \quad s(t) \overset{\mathfrak{F}}{\longleftrightarrow} S(\omega)$$

Moreover recall the derivative property of the Fourier transform that associates to a derivative in time domain a multiplication by ω in frequency domain:

$$\frac{d}{dt} \overset{\mathfrak{F}}{\longleftrightarrow} i\omega \tag{17.33}$$

where sometimes, depending on the sign convention in the exponent in the Fourier transform integral, the previous relation is written with a minus sign: $\frac{d}{dt} \overset{\mathfrak{F}}{\longleftrightarrow} -i\omega$.
The Fourier transform can be generally performed w.r.t. to any variable and not only the time. In particular it should be known that the Fourier transform of a function of the position (e.g. the wave-function $\psi(\vec{r}; t)$) is a function of the wave vector. In other words the (3D) space variable $\vec{r}$ is linked by means of the Fourier transform to the wave vector $\vec{k}$:

$$\vec{r} \overset{\mathfrak{F}}{\longleftrightarrow} \vec{k} \quad \textit{or alternatively}: \quad (x, y, z) \overset{\mathfrak{F}}{\longleftrightarrow} (k_x, k_y, k_z)$$

that in the 1D case becomes simplifies in (if only x-coordinate is considered):

$$x \overset{\mathfrak{F}}{\longleftrightarrow} k_x$$

Again the meaning is that the Fourier transform of a function of position is a function of the wave vector:

$$\psi(\vec{r}; t) \overset{\mathfrak{F}}{\longleftrightarrow} A(\vec{k}; t) \qquad \textit{or alternatively}:$$
$$A(\vec{k}; t) = \mathfrak{F}\{\psi(\vec{r}; t)\} \quad \textit{and} \quad \psi(\vec{r}; t) = \mathfrak{F}^{-1}\left\{A(\vec{k}; t)\right\}$$

where $\mathfrak{F}$ indicates the Fourier transform and $\mathfrak{F}^{-1}$ the Fourier inverse transform (or anti-transform).
More quantitatively it is possible to define a 3D Fourier transform w.r.t. to space of the

wave function as [210] (the integrals are to be intended as triple integrals in the three space variables, i.e. $d\vec{r} = dxdydz$ and $d\vec{k} = dk_x dk_y dk_z$):

$$A(\vec{k};t) = \mathfrak{F}\left\{\psi(\vec{r};t)\right\} = \frac{1}{\sqrt{(2\pi)^3}} \int \psi(\vec{r};t) e^{-i\vec{k}\cdot\vec{r}} d\vec{r} \tag{17.34}$$

and consequently:

$$\psi(\vec{r};t) = \mathfrak{F}^{-1}\left\{A(\vec{k};t)\right\} = \frac{1}{\sqrt{(2\pi)^3}} \int A(\vec{k};t) e^{+i\vec{k}\cdot\vec{r}} d\vec{k} \tag{17.35}$$

notice that the previous expressions in the 1D case simplify in:

$$A(k_x;t) = \mathfrak{F}\left\{\psi(x;t)\right\} = \frac{1}{\sqrt{2\pi}} \int \psi(x;t) e^{-ik_x x} dx$$

$$\psi(x;t) = \mathfrak{F}^{-1}\left\{A(k_x;t)\right\} = \frac{1}{\sqrt{2\pi}} \int A(k_x;t) e^{+ik_x x} dk_x$$

In summary, the Fourier transform links the so called "position space", in which the state of the system is represented by means of the wave-function $\psi(\vec{r};t)$, that is function of the position vector $\vec{r}$, to the so called "momentum space", in which the state of the system is represented by means of the Fourier transform of the wave-function $A(\vec{k};t)$, that is function of the wave vector $\vec{k}$. Sometimes the momentum space is called: "k-space".
An important theorem about the Fourier transform is the so called Plancherel's theorem (for the Fourier series the analogous of the Plancherel's theorem is the so called Fischer-Riesz, that has an analogous meaning). It states that if a function (of position, in this case) $\psi(\vec{r};t)$ is Fourier transformed in a function (of the wave vector in this case) $A(\vec{k};t)$, and then the inverse Fourier transform is performed, it is obtained a function (of position) $\psi'(\vec{r};t)$ that is different from the initial function at most in a numerable set of isolated points. It means that an integral equality holds. Moreover a corollary states that if the starting function $\psi(\vec{r};t)$ is continuous (as it is for a wave-function - see section 17.3.2) then the obtained function $\psi'(\vec{r};t)$ is pointwise equal to the original one: $\psi(\vec{r};t)$. The Plancherel's theorem has an extremely important conceptual consequence. In fact it states that the correspondence between a wave-function in the position space $\psi(\vec{r};t)$ and its Fourier transform in the momentum space $A(\vec{k};t)$ is an exactly biunivocal correspondence. In other words it means that the same information that is included in the wave-function $\psi(\vec{r};t)$ in position space, is also included in its Fourier transform $A(\vec{k};t)$ in the momentum space. Thus working with one representation or the other makes no difference. In mathematical terms the position space and the momentum space are "isomorphous". Given a wave-function in position space $\psi(\vec{r};t)$, its Fourier transform $A(\vec{k};t)$ in momentum space is uniquely determined and vice versa.
A direct consequence is on the physical meaning of $A(\vec{k};t)$. Indeed if the wave-function $\psi(\vec{r};t)$ is normalized to 1, and thus it is an amplitude of position probability, then its Fourier transform $A(\vec{k};t)$ is also normalized to 1, and it has the meaning of amplitude of momentum probability. This is a consequence of the so called Parseval's theorem that states that [45] (it holds true between any couple of functions linked by means of the Fourier transform):

$$(\psi, \psi) = (A, A)$$

where the notation $(\cdot,\cdot)$ is used to indicate the functional scalar product, as pointed previously in eq. (17.32), and thus:

$$(\psi, \psi) = \int \psi^*(\vec{r};t)\psi(\vec{r};t)d\vec{r} = \int \left|\psi(\vec{r};t)\right|^2 d\vec{r}$$

$$(A, A) = \int A^*(\vec{k};t)A(\vec{k};t)d\vec{k} = \int \left|A(\vec{k};t)\right|^2 d\vec{k}$$

from which the Parseval's relation corresponds to:

$$(\psi,\psi)=(A,A) \quad \Leftrightarrow \quad \int |\psi(\vec{r};t)|^2 \, d\vec{r} = \int \left|A(\vec{k};t)\right|^2 d\vec{k}$$

Consequently a correctly normalized wave-function in position space, such that:

$$\int |\psi(\vec{r};t)|^2 \, d\vec{r} = 1$$

corresponds to a correctly normalized wave-function in k-space:

$$\int \left|A(\vec{k};t)\right|^2 d\vec{k} = 1$$

Thus $A(\vec{k};t)$ has in general the meaning of amplitude of momentum probability, and its modulus squared has the meaning of momentum probability density, since its integral over all the possible values of the wave vector is normalized to 1 (from Parseval's relation):

$$(A,A) = \int A^*(\vec{k};t)A(\vec{k};t)d\vec{k} = \int \left|A(\vec{k};t)\right|^2 d\vec{k} = 1$$

In conclusion, in position space $\psi(\vec{r};t)d\vec{r}$ is the differential probability of finding the system described by ψ in the infinitesimal volume between $\vec{r}$ and $\vec{r}+d\vec{r}$; while in momentum space $A(\vec{k};t)d\vec{k}$ is the differential probability of finding the system described by A with a momentum (or wave vector, remember always: $\vec{p}=\hbar\vec{k}$) between $\vec{k}$ and $\vec{k}+d\vec{k}$. This is exactly the momentum probability density we were looking for at the beginning of this digression about the Fourier transform.
Before going on notice that up to now the time dependence of the wave-function was not considered. It is generally possible to Fourier transform also w.r.t. time, and from time domain it is possible to go in frequency domain: $t \overset{\mathfrak{F}}{\longleftrightarrow} \omega$. By exploiting the Planck's relation: $E=hf=\hbar\omega$ it is possible to rewrite the correspondence as:

$$t \overset{\mathfrak{F}}{\longleftrightarrow} \frac{E}{\hbar}$$

in this case one talks about the energy domain. Thus in the most general case the Fourier transform can be a 4D Fourier transform that corresponds to three Fourier transforms w.r.t. space (as explained above), that connect the position vector $\vec{r}$ to the wave vector $\vec{k}$, and one w.r.t. time that connect the time domain t with the energy domain $E/\hbar$.

Expectation values in momentum space

At this point it is known how to evaluate the expectation values of position and of a whatever function of position from equations (17.27) and (17.28), and moreover it should be clear how it is possible to proceed in the case of expected values of momentum or functions of momentum. Indeed proceeding along the lines of what was done before for writing equations (17.27) and (17.28), it is possible to say that the expected value of the wave vector $\vec{k}$, expressed in the momentum space, is:

$$\langle\vec{k}\rangle = \int \vec{k}\left|A(\vec{k};t)\right|^2 d\vec{k} = \int A^*(\vec{k};t)\vec{k}A(\vec{k};t)d\vec{k} \tag{17.36}$$

that is:

$$\langle\vec{k}\rangle = \left(A,\vec{k}A\right) \tag{17.37}$$

and for a whatever function of the wave vector $F(\vec{k})$:

$$\langle F(\vec{k})\rangle = \int F(\vec{k})\left|A(\vec{k};t)\right|^2 d\vec{k} = \int A^*(\vec{k};t)F(\vec{k})A(\vec{k};t)d\vec{k} \tag{17.38}$$

that is:

$$\langle F(\vec{k})\rangle = \left(A, F(\vec{k})A\right) \tag{17.39}$$

Nevertheless sometimes may be useful to recover an expression for these expected values in the position (real) space, instead that in momentum space. This can be done by Fourier anti-transforming the previous relations, thus eq. (17.36) can be Fourier anti-transformed to position space and the following expression is recovered (the demonstration for the 1D case is provided in example 2.3 below):

$$\langle \vec{k}\rangle = \int \psi^*(\vec{r};t)\left[-i\nabla\psi(\vec{r};t)\right] d\vec{k} \tag{17.40}$$

that can be rewritten in symbol form like:

$$\langle \vec{k}\rangle = \left(\psi, \widehat{\vec{k}}\psi\right) \tag{17.41}$$

in which the symbol $\widehat{\vec{k}}$ must be intended as the differential operator: $\widehat{\vec{k}} = -i\nabla$ (that comes out by considering the Fourier anti-transform).
Proceeding analogously for a whatever function F of the wave vector $\vec{k}$, it is possible to write:

$$\langle F\rangle = \left(\psi, \widehat{F}\psi\right) \tag{17.42}$$

in which the symbol $\widehat{F}$ must be intended as the differential operator that is built by substituting $\vec{k}$ with $-i\nabla$ in the expression of F: $\widehat{F} = F(-i\nabla)$.
Notice that these operators are a direct consequence of the Fourier transform.
At this point it is possible to recover the average values in position space of a whatever physical observable of interest that is function of position (such as the potential energy $U(\vec{r})$), by means of equation (17.28), and of whatever function of momentum (such as the kinetic energy $T = \frac{p^2}{2m}$) by means of equation (17.42) by properly substituting the function $F(\vec{k})$ with its differential operator $\widehat{F} = F(-i\nabla)$ obtained by performing the following substitutions:

$$\vec{k} \;\leftrightarrow\; -i\nabla \quad and \quad \vec{p} \;\leftrightarrow\; -i\hbar\nabla \tag{17.43}$$

that arise from the derivation property of the Fourier transform (and noticing the de Broglie relation $\vec{p} = \hbar\vec{k}$).

Example 2.3: Here a proof for the previous relation is reported for the 1D case (for the 3D case see e.g. [210]):

$$x \;\overset{\mathfrak{F}}{\longleftrightarrow}\; k_x$$

the expected value of k_x in momentum space is:

$$\langle k_x\rangle = \int k_x \left|A(k_x;t)\right|^2 dk_x = \int A^*(k_x;t)k_x A(k_x;t)dk_x$$

by substituting the explicit expression of the Fourier transform for $A(k_x;t)$ and its complex conjugate:

$$A(k_x;t) = \frac{1}{\sqrt{2\pi}}\int \psi(x';t)e^{-ik_x x'}dx'$$

$$A^*(k_x;t) = \frac{1}{\sqrt{2\pi}}\int \psi^*(x;t)e^{+ik_x x}dx$$

it follows:

$$\langle k_x\rangle = \frac{1}{2\pi}\int\int\int e^{+ik_x x}\psi^*(x;t)k_x e^{-ik_x x'}\psi(x';t)dk_x dx dx'$$

noticing that:

$$\frac{\partial}{\partial x'}e^{-ik_x x'} = -ik_x e^{-ik_x x'}$$

it follows (an integration by parts is required):

$$\langle k_x \rangle = \frac{1}{2\pi} \int\int\int e^{+ik_x(x-x')} \psi^*(x;t) \left[-i \frac{\partial}{\partial x'} \psi(x';t) \right] dk_x dx dx'$$

and considering that the only term dependent on k_x is the exponential and that its integral corresponds to the definition of the delta Dirac function:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{+ik_x(x-x')} dk_x = \delta(x-x')$$

it follows:

$$\langle k_x \rangle = \int \psi^*(x;t) \left[-i \frac{\partial}{\partial x} \psi(x;t) \right] dx = \left(\psi, -i \frac{\partial}{\partial x} \psi \right)$$

The extension to the general 3D case requires the usage of a 3D delta Dirac function and the substitution of the partial derivative w.r.t. x with the nabla operator. □

17.3.4 Quantum mechanical operators and Hamiltonian operator

A fundamental assumption in quantum mechanics is that to each physical observable F a quantum mechanical operator $\widehat{F}$ is associated. It can be a differential operator, for example of the kind of the ones associated to the wave vector for the evaluation of its expected value in real space (see the discussion presented in the previous section 17.3.3). Actually, the origin of such operators has to be found exactly in the evaluation of the expected values. In particular they arise from the Fourier transform applied to the evaluation of expected values, like it was shown in the previous section. If each physical observable must be associated to an operator, then if position space expected values are considered, from equations (17.27), (17.31), (17.40) and (17.42) it is possible to derive the following associations, true in the (real) position space (where $\vec{p} = \hbar\vec{k}$):

$$\begin{aligned} \vec{r} &\rightarrow \vec{r} \\ F(\vec{r}) &\rightarrow F(\vec{r}) \\ \vec{k} &\rightarrow -i\nabla \\ \vec{p} &\rightarrow -i\hbar\nabla \\ F(\vec{k}) &\rightarrow F(-i\nabla) \end{aligned}$$

Thus the general rules for building the quantum mechanical operators in position space are:

- the operator associated to the position vector is equal to itself: $\widehat{\vec{r}} = \vec{r}$
- a general function $F(\vec{r})$ of the position vector $\vec{r}$ is thus unchanged and its quantum mechanical operator is equal to itself: $\widehat{F} = F(\vec{r})$
- the operator associated to the wave vector is equal to $-i\nabla$: $\widehat{\vec{k}} = -i\nabla$
- a general function $F(\vec{k})$ of the wave vector $\vec{k}$ is thus associated to a quantum mechanical operator obtained substituting in the expression of F the wave vector $\vec{k}$ with $-i\nabla$: $\widehat{F} = F(-i\vec{\nabla})$

Of particular interest are the following quantities:

- the potential energy $U(\vec{r})$ is a function of the position, thus its quantum mechanical operator in position space is equal to itself: $\widehat{U} = U(\vec{r})$
- the momentum $\vec{p} = \hbar\vec{k}$ is a function of the wave vector thus its quantum mechanical operator is recovered by substituting $\vec{k}$ with $-i\nabla$ in its expression: $\widehat{p} = -i\hbar\nabla$
- the kinetic energy $T = \frac{1}{2}mv^2 = \frac{p^2}{2m}$ is a function of the momentum, i.e. of the wave vector, thus its quantum mechanical operator is recovered with the same procedure:

$$\widehat{T} = \frac{(-i\hbar\nabla)^2}{2m} = \frac{-\hbar^2}{2m}\Delta$$

where $\Delta = \nabla^2 = \nabla \cdot \nabla = \left(\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial y^2}, \frac{\partial^2}{\partial z^2}\right)$ is the so called Laplacian operator (or Laplace operator).

Notice that the quantum mechanical operators are to be intended as mathematical entities that act on a specific function, i.e. the wave-function, that represent the physical states. It has no meaning of talking about an operator if it is not "applied" to a state function (in quantum mechanics the wave-functions represent the state of the system and sometimes they are simply called "states"). Notice that from the above discussion the quantum mechanical operators are differential operators in the space variables but not w.r.t. time. The time in quantum mechanics is a parameter to which no quantum mechanical operator is associated. As time goes by, the wave-functions in general change, i.e. the system state is modified. In the Schrödinger's wave mechanics the system states, i.e. the wave-functions, evolve in time. It is possible to point out a set of properties and operations concerning the quantum mechanical operators, for a reference see e.g. [210]. The only properties that I mention here and that will be useful later on, are the linearity (since they are differential operators they are linear) and that to each physical observable is associated an operator that is Hermitian, i.e. for which the following relation holds:

$$F^\dagger = F$$

where the symbol † indicates the operation of complex conjugation (and transpose if F is in matrix form - see later). The important fact is that to an Hermitian operator is associated a physical observable that can assume real values, that is: if a measure is performed on the physical observable that is represented by an operator that is Hermitian, a real value is obtained. Since physical observables can always assume only real values, then the operators associated to them are always Hermitian.

At this point it is trivial to write down an explicit expression for the Hamiltonian operator $\widehat{H}$ that appears in the Schrödinger's equation (17.17). As already mentioned the classical Hamiltonian corresponds to the sum of the kinetic and potential energy: $H = T + U$, thus the quantum mechanical operator can be recovered with the rules just stated as follows (position space):

$$\widehat{H} = \widehat{T} + \widehat{U} = \frac{-\hbar^2}{2m}\Delta + U(\vec{r}) \tag{17.44}$$

in which the linearity is exploited such that the Hamiltonian is recovered as the sum of the kinetic energy operator $\widehat{T}$ and the potential energy one $\widehat{U}$. The last expression is exactly the one to be used in the Schrödinger's equation. The Hamiltonian operator is thus the quantum mechanical operator associated to the total energy of the system, and indeed it will be linked to the energy levels of the quantum system in sections 17.3.5 and 17.3.6.

In general in quantum mechanics to each physical observable a suitable operator (that is found by applying the previously mentioned rules) is associated, so one talks about the angular momentum operator, the squared angular momentum operator, and so on... In section 17.3.6 few more words will be said on their link to the physical observables they represent.

Quantum mechanical operators in momentum space

In section 17.3.3 it was already pointed out that is possible to express the expected value of a physical quantity also in the momentum space or k-space. In this case by proceeding in an analogous way to what done previously for obtaining the operators in position space, it is possible to get the following fundamental relations for the representation of operators in the momentum space:

$$\vec{r} \;\rightarrow\; +i\nabla_k = +i\hbar\nabla_p$$

$$F(\vec{r}) \;\rightarrow\; F(+i\nabla_k)$$

$$\vec{k} \;\rightarrow\; \vec{k}$$

$$\vec{p} \rightarrow \vec{p}$$
$$F(\vec{k}) \rightarrow F(\vec{k})$$

where $\nabla_k = (\frac{\partial}{\partial k_x}, \frac{\partial}{\partial k_y}, \frac{\partial}{\partial k_z})$ and $\nabla_p = (\frac{\partial}{\partial p_x}, \frac{\partial}{\partial p_y}, \frac{\partial}{\partial p_z})$. Everything is then analogous: a function of the momentum is associated to an operator that is equal to itself, while a function of the position is associated to an operator built up in an analogous way of what presented previously. In table 17.1 the fundamental relations are summarized.

Table 17.1: Quantum mechanical operators that are associated to physical observables in both position space and momentum space.

position space representation	momentum space representation
$\vec{r} \rightarrow \widehat{\vec{r}} = \vec{r}$	$\vec{r} \rightarrow \widehat{\vec{r}} = +i\nabla_k = +i\hbar\nabla_p$
$F(\vec{r}) \rightarrow \widehat{F} = F(\vec{r})$	$F(\vec{r}) \rightarrow \widehat{F} = F(+i\nabla_k)$
$\vec{k} \rightarrow \widehat{\vec{k}} = -i\nabla$	$\vec{k} \rightarrow \widehat{\vec{k}} = \vec{k}$
$\vec{p} \rightarrow \widehat{\vec{p}} = -i\hbar\nabla$	$\vec{p} \rightarrow \widehat{\vec{p}} = \vec{p}$
$F(\vec{k}) \rightarrow \widehat{F} = F(-i\nabla)$	$F(\vec{k}) \rightarrow \widehat{F} = F(\vec{k})$

17.3.5 General solution of Schrödinger's equation

The purpose of this section is to provide a general methodology for solving the Schrödinger's equation (17.17). Considering the expression for the Hamiltonian operator of eq. (17.44), the Schrödinger's equation can be rewritten as:

$$i\hbar\frac{\partial}{\partial t}\psi(\vec{r};t) = \widehat{H}\psi(\vec{r};t) = \left[\frac{-\hbar^2}{2m}\Delta + U(\vec{r})\right]\psi(\vec{r};t) \tag{17.45}$$

It is partial differential equation of the second order in space and of the first order in time. Since $\widehat{H}$ is time-independent, the Schrödinger's equation is completely separable in the time and the space variables. Indeed the left-hand member: $i\hbar\frac{\partial}{\partial t}\psi(\vec{r};t)$ depends only on time derivative while the right-hand one: $\left[\frac{-\hbar^2}{2m}\Delta + U(\vec{r})\right]\psi(\vec{r};t)$ depends only on space derivatives. It is possible to show that this kind of separable equations present always factorized solutions of the kind (see e.g. [210]):

$$\psi(\vec{r};t) = \Psi(\vec{r})\phi(t)$$

where $\Psi(\vec{r})$ is a function of space coordinates only and $\phi(t)$ is a function of time only. Substituting this solution in the Schrödinger's equation:

$$i\hbar\Psi(\vec{r})\frac{\partial}{\partial t}\phi(t) = \phi(t)\widehat{H}\Psi(\vec{r})$$

dividing both members by $\Psi(\vec{r})\phi(t)$:

$$i\hbar\frac{1}{\phi(t)}\frac{\partial}{\partial t}\phi(t) = \frac{1}{\Psi(\vec{r})}\widehat{H}\Psi(\vec{r})$$

where the left-hand member is function of time only and the right hand one of space only. Since the two members depend on different variables (time and space respectively) the equation is true if and only if the two members are equal to a certain constant, and said E this constant the equation can be rewritten as a system of two equations:

$$\begin{cases} i\hbar\frac{1}{\phi(t)}\frac{d}{dt}\phi(t) = E \\ \\ \frac{1}{\Psi(\vec{r})}\widehat{H}\Psi(\vec{r}) = E \end{cases}$$

$$\Rightarrow \begin{cases} i\hbar\frac{d}{dt}\phi(t) = E\phi(t) \\ \\ \widehat{H}\Psi(\vec{r}) = E\Psi(\vec{r}) \end{cases}$$

The two equations can be solved separately. The first equation is a first order differential equation in time:

$$\frac{d}{dt}\phi(t) = -\frac{i}{\hbar}E\phi(t) \quad \rightarrow \quad \frac{d\phi}{dt} = -\frac{i}{\hbar}E\phi \quad \rightarrow \quad \int\frac{1}{\phi}d\phi = -\frac{iE}{\hbar}\int dt \quad \rightarrow \quad ln\phi = -\frac{iE}{\hbar}t$$

Thus the solution of the first equation, that is the term that provides the time evolution of the state $\psi(\vec{r};t)$, is the following:

$$\phi(t) = e^{-\frac{i}{\hbar}Et} \tag{17.46}$$

unless an inessential integration constant, that is now omitted. The second equation is called time-independent Schrödinger's equation, or steady state Schrödinger's equation:

$$\widehat{H}\Psi(\vec{r}) = E\Psi(\vec{r}) \tag{17.47}$$

indeed its solutions $\Psi(\vec{r})$ are corresponding to stationary states that are time-independent. The solution for the steady state Schrödinger's equation is addressed in the next section. Here the point is that the factorized solutions for the time-dependent Schrödinger's equation (17.45) have the following form:

$$\psi(\vec{r};t) = \Psi(\vec{r})e^{-\frac{i}{\hbar}Et} \tag{17.48}$$

in this kind of solutions the time dependence is factorized and known, once the constant E is fixed. As it will be addressed in the section, the constant E corresponds exactly to the total energy of the system, this is indeed the meaning of the steady state Schrödinger's equation (see section 17.3.6). The solution of the steady state Schrödinger's equation provides indeed the energy values E characteristics of the system described by $\widehat{H}$, and also the steady states $\Psi_E(\vec{r})$, i.e. the states that are stable, and thus with an infinite life-time as introduced in section 17.2. Notice that in the following the wave-functions $\Psi_E(\vec{r})$ that are solutions of the steady state Schrödinger's equation will be indicated with a subscript E: $\Psi_E(\vec{r})$. Notice also that the steady state Schrödinger's equation represents an eigenvalue problem, in which the energy E is called "system eigenvalue" and the wave-function is called "eigenfunction" (or sometimes eigenvector). The correspondence with a usual eigenvalue problem should be clear if the Hamiltonian operator is expressed in matrix form (see section 17.4).
Moreover recall that each quantum mechanical operator that is associated to a physical observable, like e.g. the Hamiltonian operator $\widehat{H}$, associated to the total energy of the system E, is said to be Hermitian.
It is possible to show that the eigenfunctions of an Hermitian operator are always orthogonal (and thus once normalized othonormal) and constitute a "complete" set of functions. A

complete set of orthonormal functions is a set of functions that can be used as basis in the Fourier transform (or series for discrete systems). In particular if E (the eigenvalues) is a continuous variable, then it is said that the spectrum of the operator $\widehat{H}$ is continuous, and the set $\{\Psi_E(\vec{r})\}_E$ can be used as a basis for the Fourier transform; instead if E (the eigenvalues) is a discrete variable, then it is said that the spectrum of the operator $\widehat{H}$ is discrete, and the set $\{\Psi_E(\vec{r})\}_E$ can be used as a basis for the Fourier series. The completeness is intended in the sense of the Plancherel's theorem for the Fourier transform or in the sense of the Fischer-Riesz theorem for the Fourier series. The meaning of this two theorems was already pointed out, and in particular it is related to the biunivocal correspondence that the Fourier transform establishes. In particular when they hold (and they hold when a complete set is used as basis for the Fourier transform or series), it is possible to show that the Fourier transform (series) exist and converge to the original function.
From a practical standpoint, the important point of this treatment is that it is always possible to express the most general solution for the Schrödinger's equation as superposition of the factorized solutions that were presented above. In mathematical terms the way to superimpose an eventually infinite set of functions is indeed by means of the Fourier transform (or series), exactly like explained for the build up of the wave packet in section 17.2. Thus the most general solution $\psi(\vec{r};t)$ of the Schrödinger's equation can be expressed as superposition of the factorized solutions found above:

$$\psi(\vec{r};t) = \sum\!\!\!\!\!\!\int dE\ C(E)\ \Psi_E(\vec{r})\ e^{-\frac{i}{\hbar}Et} \tag{17.49}$$

where $\Psi_E(\vec{r})$ are the solutions of the steady state Schrödinger's equation, E are the so called energy eigenvalues (i.e. the possible or "permitted" energy values for the system described by $\widehat{H}$), the symbol $\sum\!\!\!\!\!\!\int dE$ indicates a summation (i.e. a Fourier series expansion) if the Hamiltonian operator spectrum is discrete (E is discrete), or an integral (i.e. a Fourier transform) if it is continuous, and $C(E)$ are the Fourier series coefficients if the Hamiltonian operator spectrum is discrete while it is the Fourier transform if the Hamiltonian operator spectrum is continuous. Notice that in the most general case the Hamiltonian operator spectrum can be partially continuous over a range of E and partially discrete for other values of E (see later - section 17.3.7).
In both cases the coefficient can be evaluated accordingly to the notation:

$$C(E) = \left(\Psi_E(\vec{r})e^{-\frac{i}{\hbar}Et}, \psi(\vec{r};t)\right) = \int \Psi_E^*(\vec{r})e^{+\frac{i}{\hbar}Et}\psi(\vec{r};t)d\vec{r}$$

i.e. they are the projection on the factorized solutions $\Psi_E(\vec{r})e^{-\frac{i}{\hbar}Et}$ of the general wave-function $\psi(\vec{r};t)$. This corresponds exactly to the meaning of Fourier series coefficients, or Fourier transform (please refer to a signal processing book if not clear - e.g. [216] [217]).
In conclusion, the general solution of the time-dependent Schrödinger's equation was provided, and it was noticed that it can be expressed as a superposition (by means of the Fourier series/transform) of the factorized solutions of the Schrödinger's equation, that represent the states at a well defined energy. These factorized solutions have always a factor dependent on time multiplied by the solutions (the eigenfunctions) of the steady state Schrödinger's equation. The latter represent a well defined energy state, while the exponential term has an imaginary exponent (indeed the energy E, time t and $\hbar$ are all real), thus leading to an oscillatory behavior in time with angular frequency: $\omega = E/\hbar$. These factorized solutions are solutions at a well defined energy E, solution of the steady state Schrödinger's equation. Thus the important insight of this section is that a whatever solution of the time-dependent Schrödinger's equation, can be expressed as a superposition of the factorized solutions that are the solutions at a well defined energy. In other words, they are the steady state solutions, and thus the solution of the steady state Schrödinger's equation lets to known everything is needed about that system, since the other solutions

are simply a superposition of the steady state ones.

This is not, in principle, much different from the free response study of a Linear Time Invariant (LTI) system. Indeed in that case the system poles (think to an RC circuit for example), play the role of the steady states in quantum mechanics, and the general LTI system response can be always be recovered from the knowledge of the system poles (that e.g. determine the denominator of all the possible transfer functions).

17.3.6 Eigenvalue problems and steady state Schrödinger's equation

In the previous section it was pointed out that the general solution of the time-dependent Schrödinger's equation can be written as a superposition of the factorized solutions of the kind:

$$\psi(\vec{r};t) = \Psi_E(\vec{r})e^{-\frac{i}{\hbar}Et}$$

These factorized solutions are called steady states of the the system. Indeed they are completely determined if the wave-functions $\Psi_E(\vec{r})$ (function of space only) are known, and the $\Psi_E(\vec{r})$ correspond to the solutions of the so called steady state Schrödinger's equation:

$$\widehat{H}\Psi_E(\vec{r}) = E\Psi_E(\vec{r}) \tag{17.50}$$

that is a time-independent version of the Schrödinger's equation. Once the stationary (time-independent) states of the system are known then the generic state of the system can be recovered with eq. (17.49), but the fundamental point that I am stressing is that the knowledge about a given system is complete in the moment in which its steady states $\Psi_E(\vec{r})$ are known.
This is similar to what is generally done in an electrical circuit with reactive elements. From circuit theory and LTI system theory should be well known that once the network poles (i.e. its proper modes, its steady states) are known, then everything about that system is known, since the general behavior of the network in presence of an external stimulus can be evaluated, also in transient behavior.
Here the name steady states comes from the fact the great majority of the interesting physical properties for these states is constant in time. This is a consequence of the fact that the time appears only in the complex exponent (thus defining an oscillating behavior with angular frequency $\omega = E/\hbar$). As highlighted in section 17.3.3, in quantum mechanics we are often interested only in average quantities, and great importance is given to expected values of physical observables, that are easily recovered starting from the wave-function as explained in section 17.3.3. If a physical observable F that is not explicitly dependent on time is considered, then its expected value, if the system is in a steady state $\psi(\vec{r};t) = \Psi_E(\vec{r})e^{-\frac{i}{\hbar}Et}$, is given by:

$$\langle F\rangle = \left(\psi, \widehat{F}\psi\right) = \int \Psi_E^*(\vec{r})e^{+\frac{i}{\hbar}Et}\widehat{F}\Psi_E(\vec{r})e^{-\frac{i}{\hbar}Et}d\vec{r}$$

and the time dependence is removed:

$$\langle F\rangle = \int \Psi_E^*(\vec{r})\widehat{F}\Psi_E(\vec{r})d\vec{r} = \left(\Psi_E, \widehat{F}\Psi_E\right)$$

Thus the expected value is time-independent, reason for which these are called steady states.
The fact that whatever solution of the Schrödinger's equation can be expressed as superposition of steady states limits the system study to the research of its steady states, solution of the time-independent Schrödinger's equation. Starting from them, a complete understanding of the physical system behavior is obtained, as it will be clarified in section 17.3.7,

where the steady state Schrödinger's equation will be effectively solved in few fundamental examples.
Moreover it is now clear that the time in quantum mechanics plays a minor role, and indeed no quantum mechanical operator is directly associated to it. Instead the system energy is likely the most important quantity for a system. It provides information concerning the actual state of a system and even about its history. This is usually well known to electrical engineers, since usually the state variables of an electrical network containing reactive elements (such as capacitors and inductors) are the voltages across the capacitors and the current that flows through inductors. These two quantities are indeed linked with the stored energy (that tells the "story" of the element) in those reactive elements:

$$W_C = \frac{1}{2}CV^2 \quad and \quad W_L = \frac{1}{2}LI^2$$

An additional point is now discussed concerning the steady state Schrödinger's equation (17.50), here reported for convenience:

$$\widehat{H}\Psi_E(\vec{r}) = E\Psi_E(\vec{r})$$

This is the so called "eigenvalue" equation for the Hamiltonian operator. The similarity with a conventional eigenvalue problem of linear algebra is evident by noting that the Hamiltonian operator can be expressed in a matrix form and the state wave-function in vector form (while E is a scalar real number), see section 17.4. In this equation the energy levels E (that represents the total energy of the system) are called "eigenvalues", and the wave-functions $\Psi_E(\vec{r})$ are called "eigenfunctions". The physical meaning of this equation has to be intended as follows. The eigenvalue equation for the operator $\widehat{H}$ that is associated to the physical observable E, when solved, provides the eigenvalues E and the eigenfunctions Ψ_E. The system eigenfunctions Ψ_E are describing the possible system states in which the physical observable represented by $\widehat{H}$ assumes exactly the value E (without uncertainty), in each measurement of that physical observable when the system is in that specific state. Thus the set of eigenvalues E constitutes the set of all possible measurement results that are performed on the physical observable represented by $\widehat{H}$ on that system, and the set of eigenfunctions $\{\Psi_E\}_E$ corresponds to the set of all possible steady states for that system, i.e. the proper modes of such a system.
In terms of Schrödinger's steady state equation this means that starting from the knowledge of $\widehat{H}$, by solving the energy eigenvalue problem (i.e. eq. (17.50)), all the possible energy values E for that system are known, and all the related stationary states Ψ_E are known.
As mentioned it is possible to write an eigenvalue equation for each possible quantum mechanical operator. In general it is possible to show that, given an operator $\widehat{F}$ that represents a physical observable F, its eigenvalue equation is:

$$\widehat{F}\psi_i = f_i\psi_i \tag{17.51}$$

where f_i are its eigenvalues and ψ_i its eigenfunctions. The physical meaning explained above still holds: each time a measurement is performed on the physical observable F, the possible result can be one among the f_i, that means that the system was in the state ψ_i. This physical meaning is very evident when the general eigenvalue equation (17.51) is demonstrated. Indeed it is demonstrated by defining a quantum mechanical operator that represents the standard deviation (or variance) introduced in section 17.3.3 and by forcing it to zero, such that f_i in eq. (17.51) actually represents a possible result of a measurement (without uncertainty).
A complete demonstration of this relation is outside the purposes of this work. A detailed treatment is present in [210].
As already pointed out in the previous section, the operator $\widehat{F}$ has a discrete spectrum if the set of its eigenvalues $\{f_i\}_i$ constitutes a discrete set (i.e. is a numerable set); while it is said that the operator $\widehat{F}$ has a continuous spectrum if the set of its eigenvalues $\{f_i\}_i$

constitutes a continuous interval in $\mathbb{R}$. Moreover it is said that the operator $\widehat{F}$ has a mixed spectrum if the set of its eigenvalues $\{f_i\}_i$ is both continuous and discrete depending on the specific system state ψ_i that is considered.
It was already said that an operator associated to a real physical observable (such as the Hamiltonian operator associated to the total energy of the system) has the property of being Hermitian. It is possible to show that the eigenvalues of an Hermitian operator are always real (indeed they correspond to possible values of a measurement performed on the system), and moreover that its eigenfunctions are orthogonal (and after a normalization orthonormal) and constitute a complete set. If the set $\{\psi_i\}_i$ is a complete then it is possible to use it as a basis for the Fourier series (if it is a discrete spectrum operator) or Fourier transform (if it is a continuous spectrum operator). This implies the validity of the already discussed Fischer-Riesz theorem (for the Fourier series) and Plancherel's theorem (for the Fourier transform). This allows to write whatever wave-function ψ as superposition of the operator eigenfunctions $\{\psi_i\}_i$, exactly like expressed in equation (17.49). With the notation of this section it becomes:

$$\psi = \sum\!\!\!\!\!\!\int dE\ C(E)\ \psi_i$$

with the same meaning of eq. (17.49). This last concept, that is generally true, will be important again when the matrix representation of quantum mechanical operators will be considered in section 17.4.

17.3.7 Solution of steady state Schrödinger's equation in few easy cases

In this section the solution of the steady state Schrödinger's equation in some important cases is addressed. The goal is to find the stationary states solution for the Schrödinger's equation that are the proper modes, i.e. the eigenfunctions, for a specific system, and also its eigenvalues (i.e. the system energy levels). The knowledge of these proper modes correspond to the complete knowledge about such a system, as widely discussed in sections 17.3.5 and 17.3.6. The equation to be solved is:

$$\widehat{H}\psi = E\psi \tag{17.52}$$

where the steady states (eigenfunctions) are simply indicated with ψ in this section, E are the eigenvalues, and $\widehat{H}$ is the Hamiltonian operator:

$$\widehat{H} = \frac{-\hbar^2}{2m}\Delta + U(\vec{r})$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, and $U(\vec{r}) = U(x, y, z)$.
Very often it is possible to simply a lot the problem, that is a 3D problem, by separating it into three 1D problems. This is possible whenever the Hamiltonian is completely separable, that corresponds to the possibility of separating the potential energy term $U(x, y, z)$ into the sum of three independent contributions:

$$U(x, y, z) = U_x(x) + U_y(y) + U_z(z)$$

Indeed under this assumption the Hamiltonian is:

$$\widehat{H} = \frac{-\hbar^2}{2m}\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right] + U_x(x) + U_y(y) + U_z(z)$$

and it can be expressed as the sum of three independent contributions in the three variables:

$$\widehat{H} = \widehat{H}_x + \widehat{H}_y + \widehat{H}_z$$

where:

$$\widehat{H}_x = \frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U_x(x)$$
$$\widehat{H}_y = \frac{-\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + U_y(y)$$
$$\widehat{H}_x = \frac{-\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_z(z)$$

In order to complete the separation of variable suitable expressions for the total eigenvalues and eigenfunctions must be recovered. In particular the total energy is given by the sum of the eigenvalues in the three directions:

$$E = E_x + E_y + E_z$$

while the wave-function is factorized in the product of the three wave-functions in the three directions:

$$\psi(x,y,z) = \psi_x(x)\psi_y(y)\psi_z(z)$$

In this way the 3D problem becomes:

$$\left[\widehat{H}_x + \widehat{H}_y + \widehat{H}_z\right]\psi_x(x)\psi_y(y)\psi_z(z) = \left[E_x + E_y + E_z\right]\psi_x(x)\psi_y(y)\psi_z(z)$$

and by dividing both members by $\psi_x(x)\psi_y(y)\psi_z(z)$ (that are always non-null since the interesting solutions are the non-trivial ones) and considering that $\widehat{H}_x$ applies only to $\psi_x(x)$ (analogously for $\widehat{H}_y$ and $\widehat{H}_z$):

$$\begin{aligned}&\widehat{H}_x\psi_x(x)\psi_y(y)\psi_z(z) + \widehat{H}_y\psi_x(x)\psi_y(y)\psi_z(z) + \widehat{H}_z\psi_x(x)\psi_y(y)\psi_z(z) = \\ &= E_x\psi_x(x)\psi_y(y)\psi_z(z) + E_y\psi_x(x)\psi_y(y)\psi_z(z) + E_z\psi_x(x)\psi_y(y)\psi_z(z)\end{aligned}$$

$$\rightarrow \frac{1}{\psi_x(x)}\widehat{H}_x\psi_x(x) + \frac{1}{\psi_y(y)}\widehat{H}_y\psi_y(y) + \frac{1}{\psi_z(z)}\widehat{H}_z\psi_z(z) = E_x + E_y + E_z$$

the problem is divided into a system of three 1D problems (indeed the above equation is true if the three members at left-hand side are one-by-one equal to three members at right-hand side):

$$\begin{cases}\widehat{H}_x\psi_x = E_x\psi_x \\ \widehat{H}_y\psi_y = E_y\psi_y \\ \widehat{H}_z\psi_z = E_z\psi_z\end{cases}$$

This separation of variables is possible because of the linearity of the Hamiltonian operator, and under the assumption of being able to separate the potential energy in the sum of three independent contributions as already pointed out. This is very often possible in many problems of practical importance and relevance. Once the variables are separated the 3D problem is divided into three 1D problems that can be solved independently the one from the others. The complete 3D solution is then recovered as already pointed out above: the total energy eigenvalues are obtained by summing the unidimensional ones while the total wave-function is the product of the unidimensional ones. For this reason in the following unidimensional problems will be considered.

One dimensional problems

According with the separation of variables it is very often possible to separate a 3D problem into three 1D problems. The 1D steady state Schrödinger's equation is (the subscripts "x" are omitted since in 1D no confusion is possible):

$$\widehat{H}\psi(x) = \frac{-\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x) + U(x)\psi(x) = E\psi(x) \tag{17.53}$$

where the potential energy $U(x)$ is linked with the (conservative) forces that are acting on the quantum system by means of the classical relation:

$$\vec{F} = -\nabla U \quad that\ in\ 1D\ is: \quad \vec{F_x} = -\frac{d}{dx}U(x) \tag{17.54}$$

An important remark is needed. The important forces in the 1D problems to be addressed are the electrostatic forces. It is well known that the electrostatic field is conservative, and thus it can be expressed as the gradient of a scalar function, that is the electrostatic potential. Consequently eq. (17.54) holds between the electrostatic force (Coulomb force) and the potential energy. In this picture is evident that in "standard" (i.e. non-relativistic) quantum mechanics the forces, and thus the potentials and potential energies, are still considered and evaluated as classical. The particle is the entity that behaves as non-classical, due to the wave-particle duality. In some fields of application this is no more accurate and also the interactions must be quantized (no more classical); this procedure is called "second quantization", but it will not addressed in this work.
In the next examples the entity and the kind of the force acting on the particle will be supposed known. Consequently the potential energy $U(x)$ is always supposed known. If the electrostatic interaction is repulsive the potential energy is conventionally positive. Nevertheless the potential energy is defined up to an additive constant, that can shift it. Thus the same concept can be reformulated as follows: if the electrostatic interaction is repulsive, the potential energy is of "barrier" type, i.e. in the region in which the repulsive force is acting a more positive potential is present w.r.t. the rest of the space. And the particle should overcome a potential barrier to prevail on the repulsive force. Otherwise if the force is attractive the potential is of "well" type, i.e. in the region in which the attractive force is acting a more negative potential is present w.r.t. the rest of the space; and the particle should overcome a potential barrier to leave that region of space and prevail on the attractive force.
Before proceeding with the examples, it is useful to highlight the general procedure and point out some useful remarks. The 1D Schrödinger's equation (17.53) can always be rewritten like:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x) + U(x)\psi(x) = E\psi(x)$$

$$\rightarrow \frac{d^2}{dx^2}\psi(x) + \frac{2m}{\hbar^2}\left[E - U(x)\right]\psi(x) = 0$$

and said:

$$k = \sqrt{\frac{2m}{\hbar^2}\left(E - U(x)\right)} \tag{17.55}$$

it becomes:

$$\frac{d^2}{dx^2}\psi(x) + k^2\psi(x) = 0 \tag{17.56}$$

Notice that k has exactly the meaning of wavenumber. Indeed if the particle is free (no force is acting on it) then $U(x) = 0$ and thus the total energy E of the particle is equal to its kinetic energy:

$$E = T = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \quad \rightarrow \quad k = +\sqrt{\frac{2m}{\hbar^2}E} = +\sqrt{\frac{2m}{\hbar^2}T}$$

(where the positive solution of the square root is considered since the wavenumber $k = 2\pi/\lambda$ is defined positive). In the case in which a force (attractive or repulsive) is acting on the particle, then it has both kinetic and potential energy, and the wavenumber is again given by the previous expression, in which the kinetic energy T is given by $T = E - U(x)$:

$$k = \sqrt{\frac{2m}{\hbar^2}\left(E - U(x)\right)}$$

The potential shapes that will be addressed in this section will be always piece-wise constant, thus it will be (for some x) $U(x) = U_0$, constant in space. It follows that equation (17.56) becomes a second order ordinary differential equation with constant coefficients, indeed k^2 depends on energy but not on space when $U(x) = U_0$ is constant since it becomes: $k = \sqrt{\frac{2m}{\hbar^2}(E - U_0)}$. Its solution can be found exploiting the characteristic polynomial (to this purpose please refer also to appendix **??**):

$$\lambda^2 + k^2 = 0 \quad \rightarrow \quad \lambda_{1,2} = \pm\sqrt{-k^2} = \pm ik$$

where k was defined above. The solution of the equation is then:

$$\psi(x) = Ae^{+\lambda_1 x} + Be^{+\lambda_2 x} = Ae^{+ikx} + Be^{-ikx} \tag{17.57}$$

Notice that this holds in general if $k \in \mathbb{R}$, i.e. if $E > U_0$, $\forall x \in \mathbb{R}$. In the case in which $E < U_0$ for some value of x then k is purely imaginary. Anyway possible to define the function β_k such that it is positive real and get (see again also appendix **??**):

$$\beta_k = \sqrt{\frac{2m}{\hbar^2}\left(U(x) - E\right)} \quad \rightarrow \quad k = i\beta_k \quad , \quad \beta_k \in \mathbb{R} \quad , \quad k \in \mathbb{C}$$

$$\psi(x) = Ae^{+ikx} + Be^{-ikx} = Ae^{-\beta_k x} + Be^{+\beta_k x} \tag{17.58}$$

The constants A and B will be determined by enforcing the boundary conditions. Moreover notice that in the last case in which $E < U_0$ the exponential terms in the solution (eq. (17.58)) have real exponents, thus they are non-propagating terms, and they correspond to exponentially attenuated wave-functions. Instead the solutions in the case of $E > U_0$ correspond to propagating waves. Indeed in equation (17.57) the two exponential terms correspond to two harmonic waves propagating respectively toward positive x values (forward wave) and toward negative x values (backward wave). In order to better visualize it let's consider the solution of the time-dependent Schrödinger's equation. From the previous treatment of the general solution of the Schrödinger's equation in section 17.3.5, it is known that the steady states, i.e. the factorized solutions of the time dependent Schrödinger's equation, are those given by equation (17.48). Thus the total wave-function in the 1D problem must be still multiplied by an exponential term that contains the time dependence as follows:

$$\psi(x;t) = \psi(x)e^{-i\frac{E}{\hbar}t} = \psi(x)e^{-i\omega t} \tag{17.59}$$

where the Planck's relation $E = \hbar\omega$ is used. Thus the total wave-function (in the case of $E > U_0$) is:

$$\psi(x;t) = Ae^{+ikx}e^{-i\omega t} + Be^{-ikx}e^{-i\omega t}$$

The first term (the forward wave) can be rewritten as:

$$\psi(x;t) = Ae^{+i(kx-\omega t)}$$

that corresponds to a monochromatic plane wave in 1D. Its phase velocity can be estimated by differentiating the phase and set it to zero:

$$\phi(x,t) = kx - \omega t \quad \rightarrow \quad d\phi = \frac{\partial\phi(x,t)}{\partial x}dx + \frac{\partial\phi(x,t)}{\partial t}dt = kdx - \omega dt = 0$$

$$\rightarrow \quad v_{phase} = \frac{dx}{dt} = \frac{\omega}{k} \quad (positive)$$

Since v_{phase} is positive it is an harmonic forward wave. Analogous procedure for the second term leads to a phase velocity $v_{phase} = -\omega/k$, thus negative. It means that it is a backward propagating monochromatic wave.

It must finally said that it has no much sense talking about a 1D plane wave. Indeed a

plane wave is a wave in which the wavefronts (i.e. the constant phase surfaces in the 3D space) are planes. The general 3D expression for a plane wave is the following:

$$\psi(\vec{r};t) = Ce^{i(\vec{k}\cdot\vec{r}-\omega t)} \tag{17.60}$$

The phase is:

$$\phi = \vec{k}\cdot\vec{r} - \omega t$$

Now the constant phase surfaces can be found by fixing the time instant at a given t_0 and set the phase ϕ constant, the result is:

$$\phi = constant \quad \Leftrightarrow \quad \vec{k}\cdot\vec{r} = k_x x + k_y y + k_z z = constant$$

The last expression correspond to a parametric equation of a plane, thus the wavefronts are planes. In 1D the harmonic plane wave of eq. (17.60) actually becomes :

$$\psi(x;t) = Ce^{i(k_x x-\omega t)} \tag{17.61}$$

from which the name used before. It is useful to notice that a plane wave has a purely parabolic energy dispersion relation. The energy dispersion relation is the relation that links the total energy E to the wavenumber k. For a plane wave of the kind of eq. (17.61) it was already pointed out that $k = k_x$ is linked to the total energy of the particle by means of the de Broglie relation:

$$E(k) = \frac{\hbar^2 k^2}{2m}$$

Thus the energy dispersion relation $E(k)$ is quadratic.

1D Free particle

A free particle, with no forces acting on it, is now considered as first practical example. The particle can be e.g. an electron. It was pointed out in the previous section that no forces acting on the particle means that the potential energy is constant in each point (indeed if $\vec{F_x} = 0$ in eq. (17.54) then $U(x)$ is constant). Since the potential energy is defined up to an additive constant then this constant can always be set such that the particle "feels" no potential energy but its energy E is only kinetic: $U(x) = 0$, $\forall x \in \mathbb{R}$. Thus the 1D steady state Schrödinger's equation (17.53) becomes:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x) + U(x)\psi(x) = E\psi(x) \quad , \quad U(x) = 0 \ \ \forall x \in \mathbb{R}$$

$$\rightarrow \frac{d^2}{dx^2}\psi(x) + k^2\psi(x) = 0 \quad , \quad k = \sqrt{\frac{2m}{\hbar^2}E} \ \in \mathbb{R}$$

The wavenumber is real since the total energy E, that equals in this case the kinetic energy, cannot be negative (indeed a negative energy E would lead to unacceptable solutions). By writing the characteristic polynomial as pointed out in the previous subsection, the solution of this equation turns out to be:

$$\psi(x) = Ae^{+ikx} + Be^{-ikx} \tag{17.62}$$

i.e. the superposition of a forward and a backward wave. Even if the wavenumber k is positive (by definition) there are two possible wave vectors (or propagation constants), that in this 1D case corresponds exactly to $\pm k$, and they appear in the two exponents. Indeed as already mentioned the first term Ae^{+ikx} is a forward wave, thus with positive propagation constant $+k$ and momentum $p = \hbar k$; while the second term Be^{-ikx} is a backward wave, with negative propagation constant $-k$ and momentum $p = -\hbar k$. To both the terms, i.e. to

both the wave-functions, the same energy value E is associated, indeed as already pointed out this is the case of an harmonic plane wave, with energy:

$$E(k) = \frac{\hbar^2 k^2}{2m}$$

Consequently to each energy value E (or eigenvalue in the notation of eigenvalue problems), two wave-functions are associated, with the two different propagation constant - see eq.(17.62). This is obvious if the dispersion relation $E(k)$ is considered. Indeed it is parabolic and two symmetric values of the wavenumber, namely $\pm k$, lead to the same energy value E. It is said that the energy level E is degenerate with degeneracy equal to two, indeed two states, or wave-functions (namely Ae^{+ikx} and Be^{-ikx}), are associated to the same energy eigenvalue E. Notice that there are no constraints on the possible range of values of E (apart that it must be non-negative), thus $E \in \mathbb{R}^+$ and the Hamiltonian spectrum is thus continuous.
In general if no interaction occurs between the particle and external forces or entities it will persevere in its state of motion, thus if it assumed to be an incoming particle from negative to positive x values then its state will be simply the forward wave ($B = 0$ since it cannot be reflected since nothing is present apart it):

$$\psi(x) = Ae^{+ikx} \tag{17.63}$$

In the case in which the kinetic energy is null, that is $E = 0$, then the steady state Schrödinger's equation simplifies in:

$$\frac{d^2}{dx^2}\psi(x) = 0$$

thus the solution is very simple, and it is recovered by integrating two times in dx:

$$\psi(x) = A + Bx$$

Nevertheless in section 17.3.2 it was discussed that the wave-function must be limited, while here $Bx \to \infty$ when $x \to \infty$, thus in order to ensure a limited wave-function everywhere B must be chosen equal to zero: $B = 0$.
An important remark is that a free particle is unlocalized. This is evident considering eq. (17.63). The magnitude squared of $\psi(x)$ is constant (recall that $|e^{\pm ikx}|^2 = 1$), and the same holds for the case of $E = 0$. Obviously the same again holds true for a single backward propagating wave. In general it is possible to rewrite the 1D wave-function of eq. (17.62) like:

$$\psi(x) = Ce^{ikx} \quad \rightarrow \quad |\psi(x)|^2 = |C|^2 \tag{17.64}$$

where depending on the sign of k (if it is accepted to embed the sign in k that becomes no more the wavenumber but the 1D propagation constant instead) the wave-function is a forward or backward wave. The point is that a free particle has the same probability of being in each point x of the space and it is unlocalized (indeed its squared magnitude is constant). This is even more clear considering that a wave-function $\psi(x) = e^{\pm ikx}$ describes a particle whose momentum $p = \hbar k$ is precisely known: that is $\Delta p = 0$. Indeed it is a monochromatic wave with a well defined k, and from the Heisenberg's uncertainty principle this requires $\Delta x \to \infty$ corresponding to unlocalization in space. Notice that unlocalization corresponds to a continue range of energy E, i.e. a continuous spectrum of the Hamiltonian operator $\widehat{H}$.
A final remark on the wave-function normalization is provided. In order to appreciate it let's consider again the last expression for the 1D plane wave of eq. (17.64). The modulus squared of the wave-function is constant everywhere. The problem is that the particle is supposed to be free in all the space, that corresponds to the fact that the potential energy is everywhere null. If in a region of space a force field is acting then this assumption is no more

true, since a potential energy variation should appear in that region. In order to normalize the wave-function the following integral must be considered (see the normalization condition eq. (17.19)):

$$\int_{all\ space} |\psi(x)|^2\, dx = \int_{-\infty}^{+\infty} |C|^2\, dx = 1 \quad (impossible!)$$

Nevertheless the problem is that that integral does not converge since it is unlimited (infinite). In other words the wave-function is not square-integrable: $\psi(x) \notin L^2(all\ space)$. The problem is conceptual, because if the wave-function is not square-integrable it cannot be normalized, and thus it cannot be associated to the physical meaning of position probability density. This issue can be overcome in three possible ways.

First, the solution is to say that the universe, thus all the space is limited and not infinite. In this way, said V the volume of the entire universe, the integral is finite thus the wave-function can be normalized similarly to what done in example 2.1 (section 17.3.2). For example assuming $V = [0, x_0] \subset \mathbb{R}$ the solution in 1D becomes:

$$\int_V |\psi(x)|^2\, dx = \int_0^{x_0} |C|^2\, dx = |C|^2\, x_0 = 1 \quad \Leftrightarrow \quad C = \frac{1}{\sqrt{x_0}}$$

This solution has very important and unsolved physical, metaphysical and even philosophic implications. Nevertheless the way of thinking is that the quantum mechanical systems are always confined at least in the laboratory in which the experiments are carried out. In this last optics it has sense. Thus a particle can be modeled as "free" even if it is not (since it is confined by external forces / materials). The continuous energy E would be no more continue at all (see later - discretization arises from confinement), but since the laboratory is macroscopic (thus extremely larger that the particle itself) the energy levels are so close that it would be impossible to distinguish between two successive levels and they are at all practical effects considered continuous, thus making true the model and its interpretation as practical tool.

The second solution starts from the consideration that an infinite physical entity cannot be directly measured (since it is infinite) and thus believing that it may have an intrinsic physical reality is a matter of personal philosophy. Consequently the solution is very practical: independently on the existence or not of an infinite space, the solution consists in saying that even if $\psi(x) \notin L^2(all\ space)$, it keeps anyway the meaning of a relative probability:

$$\frac{\int_{V_1} |\psi(x)|^2 dx}{\int_{V_2} |\psi(x)|^2 dx}$$

In the moment in which V_1 and V_2 tend to infinity their ratio is finite, leading to a relative probability that is finite and can be suitably normalized. Nevertheless monochromatic plane waves are mathematically very simple and useful. Indeed the set of all the monochromatic plane waves (in 1D) $\left\{\frac{1}{\sqrt{\Omega}} e^{ikx}\right\}_k$ can be shown to be orthonormal and complete, and thus it can be used as basis for the Fourier Transform (if k is continuous) or series (if k is discrete); and the usefulness of the Fourier analysis is of extreme both conceptual and practical importance in quantum mechanics, as it should be clear from the treatment of previous sections 17.2, 17.3.3, 17.3.5, 17.3.6. In this solution it is implicitly admitted their physical inconsistency.

The third solution is widely accepted. It is referred as "wide sense normalization". The mathematical procedure is quite similar to the one conventionally adopted for introducing (non formally) the concept of Fourier transform in signal processing courses [216], [217]. In few words it consists in limiting the domain to a finite domain (similarly to what done in first solution), but then to enforce periodic boundary conditions, such that the system is left open. This means the the wave-function does not tend to zero at the domain boundaries (that would mean confinement, indeed a null wave-function means no probability of finding the particle there), but instead it has the same phase condition at all the boundaries (i.e.

the frontier) of the domain. Thus since the domain is limited a discretization occurs (see next examples to better understand this point). Nevertheless, then, the boundaries of the domain are made tend to infinity, and from a discrete system (Fourier series) a continuous system is obtained (Fourier transform). So the continuous case is recovered but the gain is that a normalization condition is found for the continuous case, i.e. for the unlocalized free particles (monochromatic harmonic plane waves). The normalization condition involves the functional theory, and in particular it involves the delta Dirac function. For the 1D case it is:

$$\int_{-\infty}^{+\infty} |\psi_k(x)|^2 dx = (\psi_k, \psi_{k'}) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i(k-k')} dx = \delta(k-k')$$

Sometimes in electromagnetic fields courses similar expressions are found for the modal functions. With this new delta Dirac normalization condition, that holds for continuous systems, it is then possible to normalize the unlocalized wave-function in a wide sense, and thus to assign them the meaning of position probability density as usual.
A complete discussion with a focus on the important physical, metaphysical and philosophic implications of all the three solutions is provided in [210].

3D free particle

In this section is recovered the expression for the wave-function of a free particle in the 3D space. Starting from the points gained in the previous section for the 1D free particle wave-function, and considering the separation of variables presented at the beginning of the section, it is possible to directly write the final expression for the wave-function. Indeed from the separation of variables it is possible to solve separately three 1D problems in x, y and z like already done in the previous subsection, and then the full 3D wave-function is given by the product of the three 1D wave-functions. By doing so:

$$\psi(x,y,z) = A_x A_y A_z e^{ik_x x} e^{ik_y y} e^{ik_z z} = C e^{i\vec{k}\cdot\vec{r}} \quad , \quad C = A_x A_y A_z \ \ (\textit{generally complex})$$

If time is considered (see eq. (17.49) for the total solution of the time dependent Schrödinger's equation):

$$\psi(x,y,z;t) = C e^{i\vec{k}\cdot\vec{r}} e^{-i\omega t} = C e^{i(\vec{k}\cdot\vec{r}-\omega t)}$$

Notice that it is a plane wave in the 3D space (as pointed out previously in this section). Since it is a monochromatic harmonic wave (ω is fixed once E is fixed: $E = \hbar\omega$), it is an unlocalized wave-function (see also section 17.2) and the same remarks done in the previous subsection are valid here. In particular its squared magnitude is constant and thus there is the same probability of finding the particle in each point of the 3D space.

Potential step

In this subsection the solution of the steady state Schrödinger's equation for the case of a potential step is addressed. It is supposed to have zero potential energy for $x < 0$ and a constant positive value $U_0 > 0$ for $x > 0$:

$$U(x) = \begin{cases} 0 & if\ x < 0 \\ U_0 & if\ x \geq 0 \end{cases}$$

For example this shape of the potential can be the mathematical simplified representation of the potential experienced by free electron in a metal nearby the surface. Indeed electrons in a metal can be considered as a free gas around the nuclei, but they are confined in the material (in normal conditions the metal does not emit electrons). Thus they experience no potential deep in the metal, but they "feel" a potential step (of height of the order of the work function) close to the surface, such that they are confined inside the metal. Thus the potential step can be a mathematical (abrupt) simplification of the (smoother) surface

potential experienced by electrons in a metal [211].
In order to proceed it is better to solve separately the two cases for $E < U_0$ and $E > U_0$.
Case of $E < U_0$: In this case the classical mechanics predicts that the particle cannot be found in the region with $x > 0$. Thus it is a classical forbidden region. Think for example to a marble with kinetic energy E coming from left toward a potential barrier. The potential energy barrier can be a hillock, that can be overcome by the marble only if it has enough kinetic energy E to overcome the gravitational potential barrier U_0: i.e. if $E > U_0$. Otherwise the marble will go up a bit but then it will come back on the left side. Let's split the problem into two parts for negative (region I) and positive (region II) x values and then merge together the solutions by enforcing the wave-function properties presented in section 17.3.2. For $x < 0$ no potential is acting on the electrons thus the solution of the Schrödinger's equation is analogous to the one already found in the previous subsections for a free particle:

$$\psi_I(x) = Ae^{+ik_1x} + Be^{-ik_1x} \quad , \quad k_1 = \sqrt{\frac{2m}{\hbar^2}E} \in \mathbb{R}^+$$

For $x > 0$ instead the Schrödinger's equation becomes:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi_{II}(x) + U_0\psi_{II}(x) = E\psi_{II}(x) \quad \rightarrow \quad \frac{d^2}{dx^2}\psi_{II}(x) + k_2^2\psi_{II}(x) = 0$$

$$with: \quad k_2 = i\beta_k \quad , \quad \beta_k = \sqrt{\frac{2m}{\hbar^2}(U_0 - E)} \in \mathbb{R}^+$$

that has the general solution (see also previously the general procedure for 1D problems):

$$\psi_{II}(x) = Ce^{+ik_2x} + De^{-ik_2x} = Ce^{-\beta_k x} + De^{+\beta_k x}$$

Nevertheless $De^{+\beta_k x}$ is unacceptable since the wave-function must be everywhere limited and for $x \rightarrow +\infty$ the term $De^{+\beta_k x} \rightarrow +\infty$; thus D is set to zero: $D = 0$.
The total wave-function is thus:

$$\psi(x) = \begin{cases} \psi_I(x) = Ae^{+ik_1x} + Be^{-ik_1x} \\ \psi_{II}(x) = Ce^{-\beta_k x} \end{cases}$$

A very important remark that makes quantum mechanics different from classical mechanics is that with quantum mechanics there is a certain probability of finding the particle (electron) in the region II even in the case it has an energy $E < U_0$. Indeed the wave-function in the region II with $x > 0$ it is non-zero, but it is $\psi_{II}(x) = Ce^{-\beta_k x}$. Since β_k is real (ad positive) the probability of finding the particle on the other side of the potential step is exponentially decreasing with the distance x. Moreover the smaller is E w.r.t. U_0, the faster is the decay (see the expression for β_k above). In other words the exponential decay is faster if U_0 is increased, and for $U_0 \rightarrow +\infty$ the wave-function exactly zero at the boundary ($\psi_{II}(x) \rightarrow 0$, that means that if the potential step is infinitely high there is no probability of finding the particle in the classically forbidden region, and the classical result is recovered). In physical terms there is a given probability of finding the electrons beyond the metal surface, even this probability rapidly decays with the distance.
The constants A, B and C can be determined by enforcing the continuity on the wave-function and its first derivative in $x = 0$:

$$\begin{cases} \psi_I(x=0) = \psi_{II}(x=0) \\ \\ \frac{d\psi_I(x)}{dx}|_{x=0} = \frac{d\psi_{II}(x)}{dx}|_{x=0} \end{cases} \rightarrow \begin{cases} Ae^{+ik_1 0} + Be^{-ik_1 0} = Ce^{-\beta_k 0} \\ \\ ik_1Ae^{+ik_1 0} - ik_1Be^{-ik_1 0} = -\beta_k Ce^{-\beta_k 0} \end{cases}$$

These conditions yield $A + B = C$ and $ik_1(A - B) = -\beta_k C$, which in turn give:

$$B = \frac{(ik_1 + \beta_k)A}{ik_1 - \beta_k} \quad and \quad C = \frac{2ik_1A}{ik_1 - \beta_k}$$

Notice that one parameter is free (indeed there are three unknowns A, B and C and only two equations; thus from Rouché-Capelli's theorem there are ∞^1 solutions). A common choice is to assume A as known, since it represents the incoming (from left) wave amplitude, and normalized to 1. The incoming wave field intensity is thus $|A|^2$, while $|C|^2$ is the intensity of the transmitted wave field and $|B|^2$ the intensity of the reflected back wave field. Notice that:

$$|B|^2 = \left|\frac{ik_1+\beta_k}{ik_1-\beta_k}A\right|^2 = \frac{(ik_1+\beta_k)}{(ik_1-\beta_k)}\frac{(-ik_1+\beta_k)}{(-ik_1-\beta_k)}|A|^2 = |A|^2$$

Therefore the incident and the reflected wave fields have the same intensity. This result should be interpreted by saying that all incoming (from left) particles reaching the potential step with an energy $E < U_0$ bounce back, including also those that penetrate slightly into the region II. Thus at the end all the particles are reflected, and this is a consequence of the fact that the potential step extends infinitely at right side (for $x > 0$). If it were of finite width, and enough thin, there may happen that a particle can cross it even if it has an energy $E < U_0$ (since the exponential decay may be not enough to ensure zero probability of finding the particle at right side). This is exactly what happens for enough thin potential barriers, see later in this section the potential barrier 1D problem.
In figures 17.4 and 17.5 few examples of interest concerning a potential step are reported. The figures were created in *MatLab* environment implementing a finite difference method for solving the Schrödinger's equation, following what is presented in section 17.4 (the code is reported in appendix **??**). Notice the exponential tails and the larger penetration in region II when the energy of the incoming particle is increased.
Case of $E > U_0$: In this case the classical prediction, if the particle is assumed coming from left, is that it always overcome the step, although it has a smaller velocity in region II than in region I. Nevertheless the quantum mechanical picture is different even in this case. Indeed there is a given probability of reflection even if $E > U_0$.
The solution of steady state Schrödinger's equation in region I is the same of before:

$$\psi_I(x) = Ae^{+ik_1x} + Be^{-ik_1x} \quad , \quad k_1 = \sqrt{\frac{2m}{\hbar^2}E} \in \mathbb{R}^+$$

Instead for $x > 0$ the Schrödinger's equation becomes:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi_{II}(x) + U_0\psi_{II}(x) = E\psi_{II}(x) \quad \rightarrow \quad \frac{d^2}{dx^2}\psi_{II}(x) + k_2^2\psi_{II}(x) = 0$$

$$with: \quad k_2 = \sqrt{\frac{2m}{\hbar^2}(E-U_0)} \in \mathbb{R}^+$$

It has the general solution:

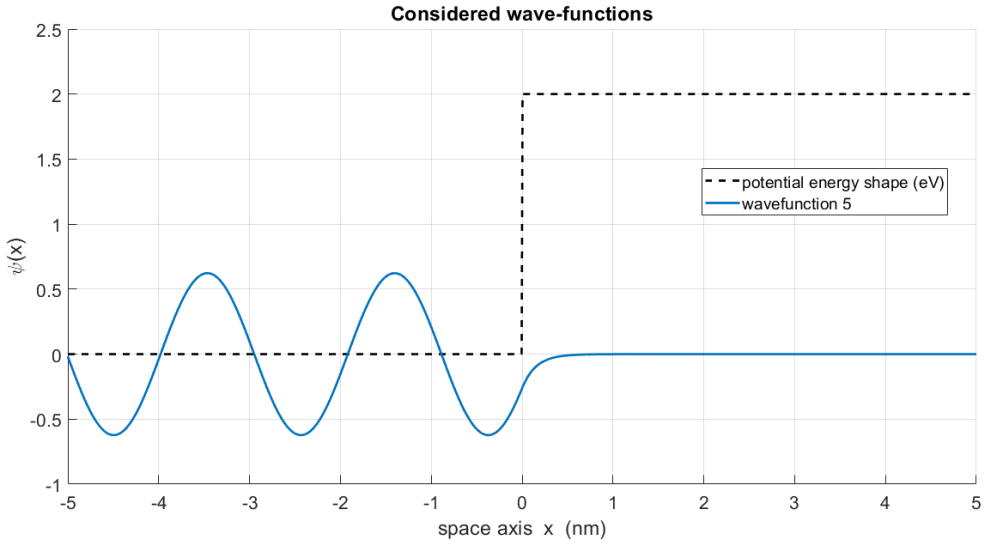
$$\psi_{II}(x) = Ce^{+ik_2x} + De^{-ik_2x}$$

Since now k_2 is real (positive) then both the forward and the backward waves are limited and mathematically acceptable.

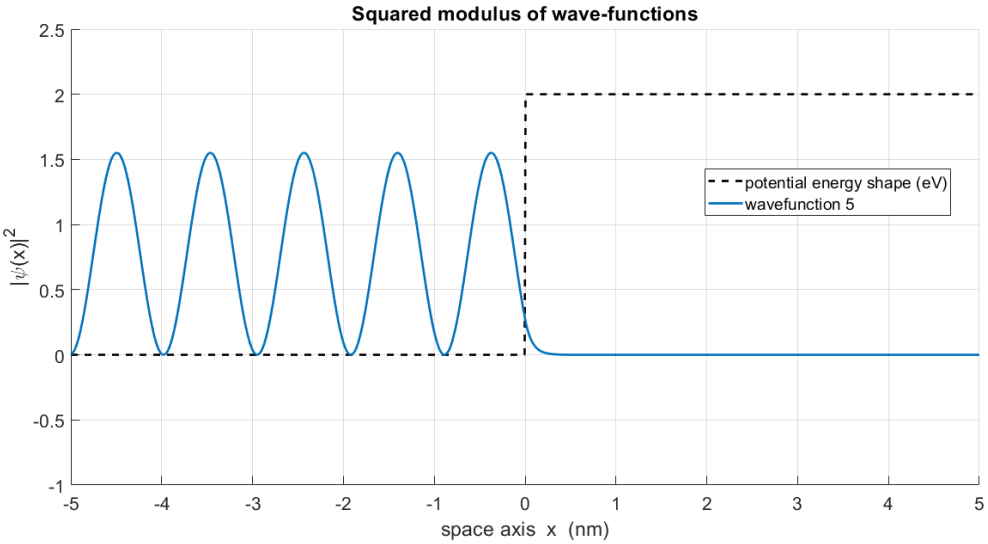
Nevertheless assuming that the particle is incoming from left, once it overcomes the potential step there is no more reason why it could be reflected back (no other variations of the potential occurs, and if $U(x)$ is constant the particle perseveres in its state of motion with no reflection or scattering phenomena).

For this reason $D = 0$ (no backward wave in the region II). Consequently:

$$\psi_{II}(x) = Ce^{+ik_2x} \quad and \quad \psi(x) = \begin{cases} \psi_I(x) = Ae^{+ik_1x} + Be^{-ik_1x} \\ \psi_{II}(x) = Ce^{+ik_2x} \end{cases}$$

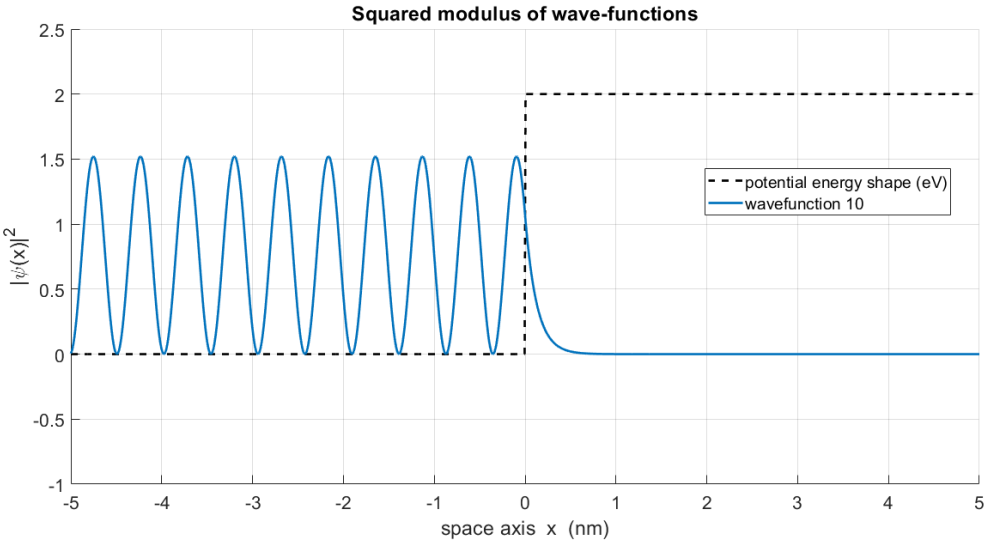


(a) wave-function

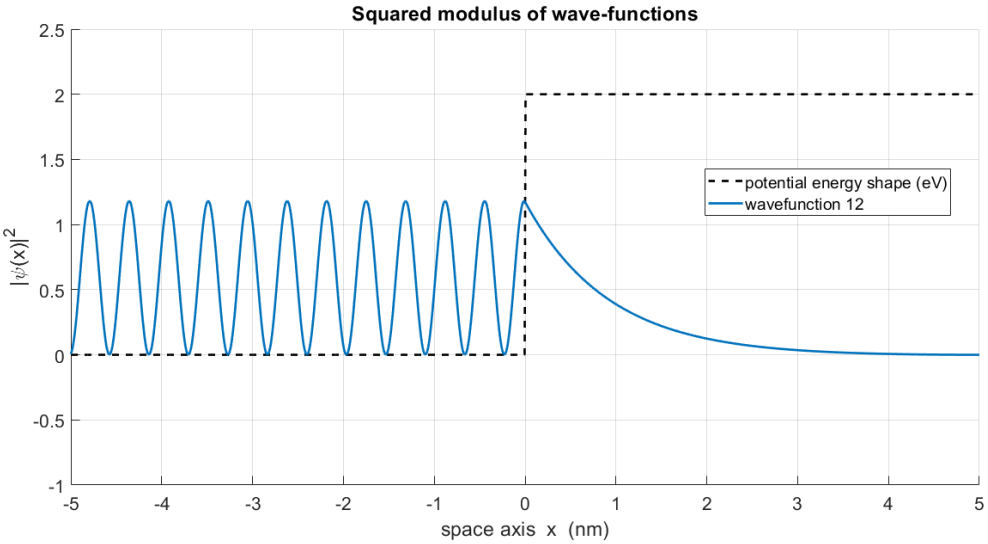


(b) wave-function squared modulus

Figure 17.4: Wave-function (a) and wave-function squared modulus (b) corresponding to an electron with energy $E = 0.354\,\text{eV}$ incoming from left toward a potential step of height $U_0 = 2\,\text{eV}$. The wave-function and its squared modulus are unnormalized.



(a) wave-function squared modulus for $E = 1.4\,\text{eV}$



(b) wave-function squared modulus for $E = 1.989\,\text{eV}$

Figure 17.5: Wave-function squared modulus corresponding to an incoming electron from left side, with energy $E = 1.4\,\text{eV}$ (a) and $E = 1.989\,\text{eV}$ (b) toward a potential step of height $U_0 = 2\,\text{eV}$. The wave-functions and their squared moduli are unnormalized.

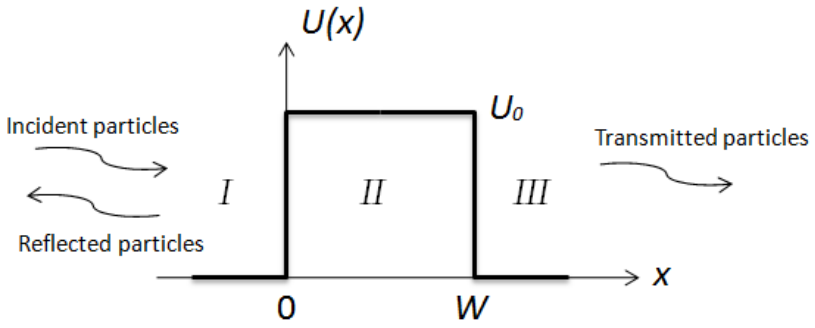


Figure 17.6: Rectangular potential energy barrier. The barrier height is U_0, its width is W. Incident (from left), reflected and transmitted paths are shown. The three regions are: I for $x < 0$; II for $0 \leq x \leq W$; and III for $x > W$.

Applying then the boundary conditions on the wave-function and its first derivative continuity:

$$\begin{cases} \psi_I(x=0) = Ae^{+ik_1 0} + Be^{-ik_1 0} = \psi_{II}(x=0) = Ce^{+ik_2 0} \\ \frac{d\psi_I(x)}{dx}|_{x=0} = ik_1 Ae^{+ik_1 0} - ik_1 Be^{-ik_1 0} = \frac{d\psi_{II}(x)}{dx}|_{x=0} = ik_2 Ce^{+ik_2 0} \end{cases}$$

$$\rightarrow \quad \begin{cases} A + B = C \\ k_1(A - B) = k_2 C \end{cases} \quad \rightarrow \quad B = \frac{k_1 - k_2}{k_1 + k_2} A \quad and \quad C = \frac{2k_1}{k_1 + k_2} A$$

The important fact, as already pointed out, is that in this case B is not zero, thus a particle has a non-null probability of being reflected back in $x = 0$ due to the potential step even in it has an energy E greater than the step height: $E > U_0$. This is again an effect that arises only in quantum mechanics. Notice that this reflection is a characteristic behavior of all the wave fields whenever a region of discontinuity of the physical properties of the medium is present (analogously to the well known behavior of elastic or electromagnetic waves).
A final remark is provided concerning negative energies. If the calculations are repeated with $E \leq 0$, enforcing the boundary conditions it turns out that the only possible solution is $\psi(x) = 0$ (i.e. the trivial one), that means no state is possible with negative or null energy (zero probability means no particle in that state). This holds is general, and as already pointed out for the free particle case, this is not surprising.

Potential barrier and tunnel effect

It is now considered a potential barrier of thickness W and height U_0, of the kind of the one in figure 17.6. The corresponding expression of the potential is:

$$U(x) = \begin{cases} 0 & if\ x < 0 \quad (region\ I) \\ U_0 & if\ \ 0 \leq x \leq W \quad (region\ II) \\ 0 & if\ x > W \quad (region\ III) \end{cases}$$

It is supposed to have an incoming particle (electron) of energy E from left side (as indicated in the picture). Analogously to the previously considered examples the case in which the incoming particle energy is negative or null ($E \leq 0$) is not interesting since it leads to null states of the kind $\psi(x) = 0$. Thus only the case $E > 0$ is taken into account.
Case of $E < U_0$: In this case from classical mechanics it is expected a full reflection of the incoming particles (as already described for the potential step). Nevertheless in quantum mechanics it turns out that there is a finite probability of finding the incoming particles

on the right side of the barrier, thus there is a certain probability of transmission. This is referred as "tunnel effect".
In regions I and III the potential energy $U(x)$ is zero, thus the steady state Schrödinger's equation is the one for the free particle already discussed.
The general solutions in the two regions are:

$$\psi_I(x) = Ae^{ik_1x} + Be^{-ik_1x} \quad and \quad \psi_{III}(x) = Ce^{ik_3x} + De^{-ik_3x}$$

$$with: \quad k_1 = k_3 = k = \sqrt{\frac{2m}{\hbar^2}E} \ \in \mathbb{R}^+$$

In region I both an incident and a reflected wave fields are expected. Instead by supposing no other potential energy variations after W in region III only a transmitted wave is expected, thus it is chosen $D = 0$ (no backward wave in region III).
In region II the Schrödinger's equation becomes:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi_{II}(x) + U_0\psi_{II}(x) = E\psi_{II}(x) \quad \rightarrow \quad \frac{d^2}{dx^2}\psi_{II}(x) + k_2^2\psi_{II}(x) = 0$$

$$with: \quad k_2 = i\beta_k \ , \ \beta_k = \sqrt{\frac{2m}{\hbar^2}(U_0 - E)} \ \in \mathbb{R}^+$$

that has the general solution:

$$\psi_{II}(x) = Fe^{+ik_2x} + Ge^{-ik_2x} = Fe^{-\beta_k x} + Ge^{+\beta_k x}$$

Since the barrier extends from $x = 0$ to $x = W$ no constraints are present on $\psi_{II}(x)$ since it results always limited. The fact that both a positive and a negative exponentials are presents has to be interpreted as follows. The incoming particle from left has a probability to be transmitted at $x > 0$, but this probability decays exponentially with space accordingly with $Fe^{-\beta_k x}$. Since in $x = 0$ there is a potential discontinuity there is a non-null probability of reflection that give rise to the backward term in region I. Analogously the potential discontinuity in $x = W$ gives rise to a non-null probability of reflection of the (exponentially decaying) transmitted wave again region II. This is a backward propagating wave, that is exponentially attenuated (since $E > U_0$), and corresponds to the term $Ge^{+\beta_k x}$ (since it is propagating in $-x$ direction it corresponds to an attenuated wave). Then if the barrier width W is finite (and enough thin) a non-negligible probability of finding the particle in the region III is present, described by the transmitted term: $\psi_{III}(x) = Ce^{ik_3x}$. Thus the total wave-function is:

$$\psi(x) = \begin{cases} \psi_I(x) = Ae^{ikx} + Be^{-ikx} & if \ x < 0 \qquad (region \ I) \\ \psi_{II}(x) = Fe^{-\beta_k x} + Ge^{+\beta_k x} & if \ \ 0 \le x \le W \quad (region \ II) \\ \psi_{III}(x) = Ce^{ikx} & if \ x > W \qquad (region \ III) \end{cases}$$

The quantum mechanical result for which it is possible to find the particle in region III even if it has an energy $E < U_0$ has no classical analogous and is called "tunnel effect". It is at the origin of the gate current in conventional MOSFETs and in FinFETs. Notice that (as mentioned in the introductory chapter -section **??**-) the tunneling is an exponential phenomenon with the barrier width W, thus a linear decrease in the barrier width leads to an exponential increase of the tunneling probability.
The values of the five constants A, B, C, F and G can be determined as done in the previous examples by enforcing the continuity on the wave-function and on its first space derivative in the discontinuity points $x = 0$ and $x = W$. The incoming wave field intensity $|A|^2$ can be assumed to be known (indeed it is obtained a system of four equations and five unknowns). In this case usually the explicit expression of the final wave-function is not of

interest. Instead the transmission and reflection coefficients T and R are usually considered interesting. It is possible to define them as:

$$T = \left|\frac{C}{A}\right|^2 \quad and \quad R = \left|\frac{B}{A}\right|^2$$

It can be verified that $T + R = 1$. The details on these calculation are not reported here, and are left to the interested reader. Good references can be [210] (in Italian) and [211].
An example of tunneling through a potential barrier is reported in figure 17.7, that is the output of a *MatLab* code (reported in appendix **??**) implementing a finite difference method based on the considerations of section 17.4.
Case of $E > U_0$: Similarly to what happens in the case of a potential step in this case it is classically expected that all the incoming particles reaching the barrier from left are transmitted. Instead with quantum mechanics a finite probability of reflection is present both at $x = 0$ and $x = W$, i.e. in the potential discontinuity points.
The same procedure already carried out for the previous case can be followed for obtaining the following result for the total wave-function in the three regions:

$$\psi(x) = \begin{cases} \psi_I(x) = Ae^{ikx} + Be^{-ikx} & if \;\; x < 0 \quad (region\ I) \\ \psi_{II}(x) = Ce^{ik'x} + De^{-ik'x} & if \;\; 0 \le x \le W \quad (region\ II) \\ \psi_{III}(x) = Fe^{ikx} & if \;\; x > W \quad (region\ III) \end{cases}$$

$$where: \quad k = \sqrt{\frac{2m}{\hbar^2}E} \in \mathbb{R}^+ \quad and \quad k' = \sqrt{\frac{2m}{\hbar^2}(E - U_0)} \in \mathbb{R}^+$$

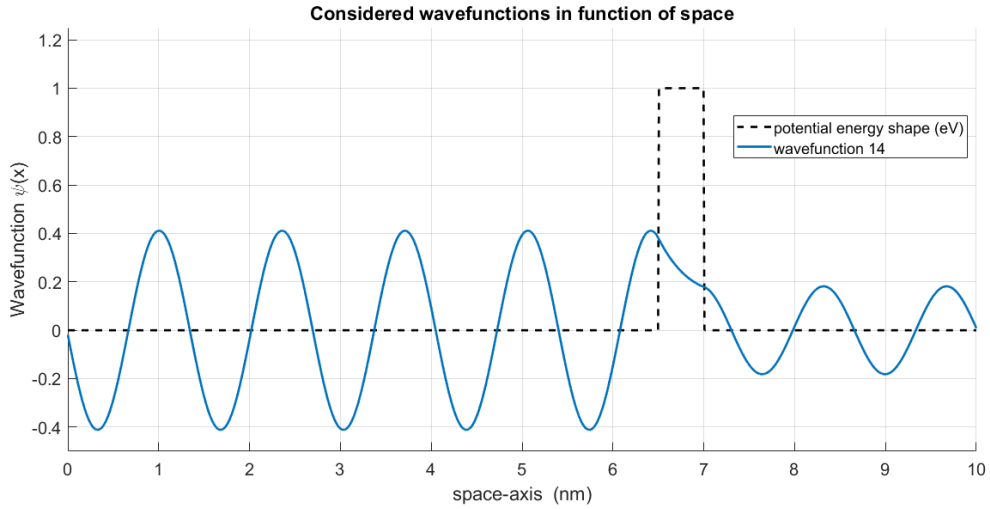
Notice that k corresponds to the wavenumber of a free electron and $\hbar k'$ is the momentum of the particle while crossing the barrier. The values of the constants B, C, D and F can be determined as functions of A by applying the boundary conditions on wave-function continuity and on its first derivative continuity in $x = 0$ and $x = W$, again four equations are recovered and thus one free parameter, namely A, appears in the solution. The term Be^{-ikx} corresponds to a back reflected wave that has no classical analogous. It is possible to define a reflection coefficient analogously to what done previously in the case of $E > U_0$. Instead of defining the transmission coefficient, in this case, it is possible to define the so called "transparency" T of the barrier: $T = |F|^2/|A|^2$. It is possible to show (see e.g. [211] or [210]) that there are values of E/U_0 for which there is a perfect transmission, i.e. $T = 1$. These values of input energy E correspond to a particle de Broglie wavelength (inside the barrier) $\lambda' = \frac{2\pi}{k'}$ that is an integer multiple of $2W$:

$$resonance\ condition: \quad \lambda' = \frac{2\pi}{k'} = \sqrt{\frac{2\pi^2\hbar^2}{m(E - U_0)}} = 2Wn \quad , \quad n \in \mathbb{N} \setminus \{0\}$$

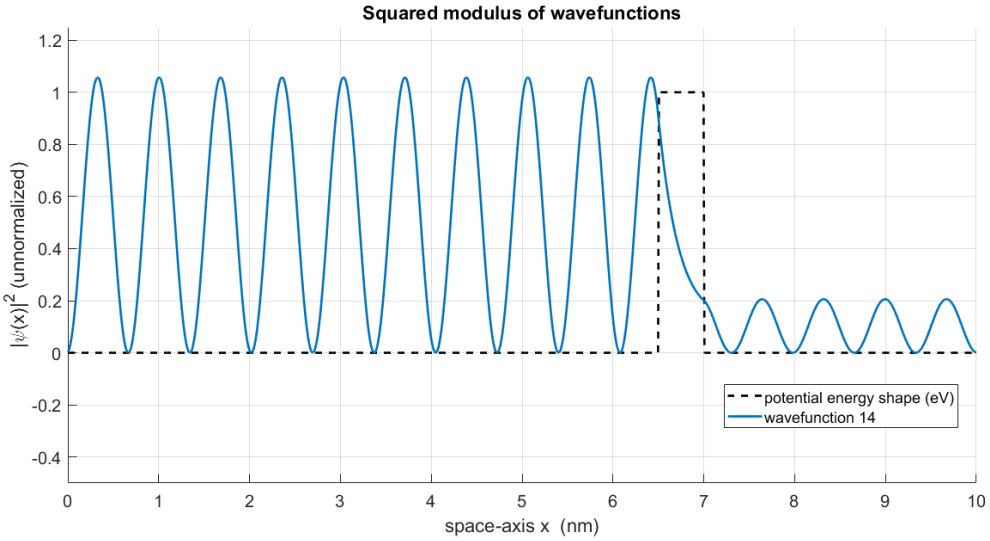
This phenomenon is called "resonance effect". In figure 17.8 a pair of examples of transparency T as a function of particle energy E are reported for rectangular potential barriers. Notice the resonance peaks of transmission.

1D potential well

The potential energy shape considered in this section has an opposite shape w.r.t. the potential barrier considered previously. An example is reported in figure 17.9. As mentioned at the beginning of the present section, a positive potential (e.g. of barrier type) corresponds to a repulsive interaction, instead a negative potential (e.g. of well type) corresponds to an attractive potential. In figure 17.9 (a) it is evident that the particle experiences a negative potential in the region $-\frac{L}{2} \le x \le +\frac{L}{2}$, while it experiences a null potential in $x < -\frac{L}{2}$ and $x > +\frac{L}{2}$. This means that there is an electrostatic attractive force acting on the particle in the central region, within the well, while around it the particle does not feel it and behaves like a free particle (since no potential is acting on it). In particular, also in this case it is



(a) wave-function



(b) wave-function squared modulus

Figure 17.7: Wave-function (a) and wave-function squared modulus (b) corresponding to an electron with energy $E = 0.82\,\text{eV}$ incoming from left toward a potential barrier of width $0.5\,\text{nm}$ and height $U_0 = 1\,\text{eV}$. The wave-function and its squared modulus are unnormalized.

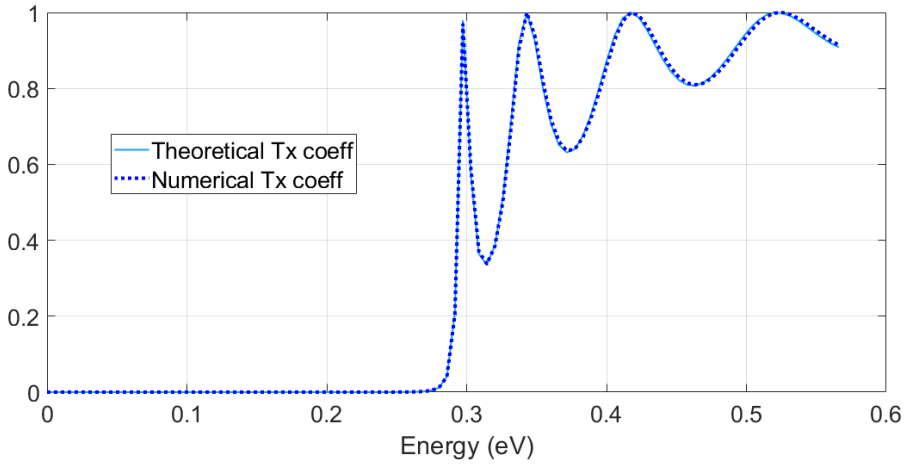


Figure 17.8: Rectangular potential energy barrier transparency example. The barrier height is of about $U_0 = 4.5\,\text{eV}$ while its thickness is 20 nm. The transmission coefficient is calculated both starting from analytical theoretical formulae (following the approach of [218]), and with a numerical method. The numerical method is an implementation of a Finite Element Method for the solution of the Schrödinger's equation, conceptually similar to what will be presented in section 17.4.1. The two methods provide very similar results indeed the two curves are well overlapped. Notice the resonance peaks and the non-null transmission also for $E < U_0$.

convenient to separate the various cases depending on what is the total energy E of the considered particle (electron). If $E < -U_0$ no state is possible (see previous subsections, here it is analogous) and this case is not of interest. If $-U_0 \leq E \leq 0$ it is said that the particle is in a "bound state". Classically it cannot overcome the potential barrier (exactly as already seen for the potential step or the barrier) and go out of the well, thus the particle is classically confined within the well. Thus bound states are conventionally the ones with negative energy. Instead if $E > 0$ then the particle is classically free to move everywhere in the space (analogously to the barrier case), indeed it has enough energy to jump out of the well and move in the zero potential region. The states with positive energy are called "free states" or "continuum states". Indeed the solution of the Schrödinger's equation in this case is analogous to the one for the free particle, and the energy E is continuous: $E \in \mathbb{R}^+$. The sign convention on energy E is thus:

- bound states: with negative energy $-U_0 \leq E \leq 0$. They represent classically confined states. Analogously to the cases of the potential step and barrier in quantum mechanics there is a given probability of finding the particle outside the well (if it has a finite height). It will be shown in a while that these states have a discrete spectrum (of Hamiltonian operator), i.e. discrete energy levels.
- continuum states: with positive energy $E > 0$. They represent free particles moving with no particular constraints. These states have continuous range of eigenvalues ($E \in \mathbb{R}^+$), thus the Hamiltonian spectrum is said to be continuous.

This convention is widely spread and often used, especially in chemistry. In which a negative energy level corresponds to a bound state, that means that the electron is confined in that chemical structure (e.g. atoms, molecules, crystal, etc...). The binding energy (or dissociation energy) is the amount of energy to make the electron free, thus it corresponds to the energy difference between the considered energy level and the zero-energy level (vacuum level). If that amount of energy is provided to the electron (e.g. in a molecule) it undergoes to a state transition from a bound state (in which e.g. it is kept in the molecule) to a free state. Thus the binding energy is exactly the amount of energy to be provided to an electron in a bound state in order to break its chemical bond and make it free.

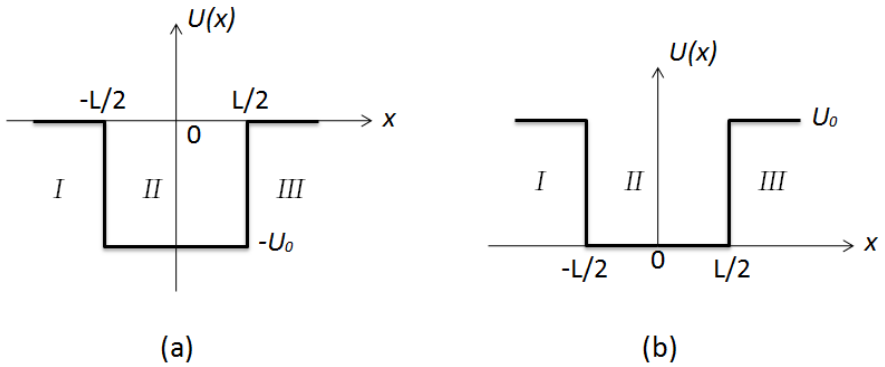


Figure 17.9: 1D potential energy well. The well height is U_0, its width is L, the potential has even symmetry. Left side (a) shows a potential well with the sign in accordance with the standard sign convention for bound and free states. The binding potential is negative $-U_0$, the bound states are those with energy $-U_0 \leq E \leq 0$, the free states are those with energy $E > 0$. Right side (b) shows the potential well considered here with positive U_0, and with the following sign convention. The bound states are those with $0 \leq E \leq +U_0$, the free states those with $E > U_0$.

Even if this convention is widely spread, in the practical solution of the steady state Schrödinger's equation addressed in this section it will be assumed to have zero potential inside the well, thus $U(x) = 0$ for $-\frac{L}{2} \leq x \leq +\frac{L}{2}$, and a positive potential value U_0 (constant) outside the well. Since the potential energy is defined up to an additive constant the final result does not change and exactly the same results are recovered. The only thing that changes is the sign convention on states, that is now different from the one above. In particular the bound states will be the ones with energy $0 \leq E \leq U_0$ and the free ones will be those with $E > U_0$ (see figure 17.9 (b)).

With this convention, the potential shape is:

$$U(x) = \begin{cases} U_0 & if\ x < -\frac{L}{2} \\ 0 \quad if\ -\frac{L}{2} \leq x \leq +\frac{L}{2} \\ U_0 & if\ x > +\frac{L}{2} \end{cases}$$

Notice that this potential shape can be an oversimplification of the real the case of a free electron gas within a metal. If the positive ions potential shape is neglected, the potential inside the metal is flat and chosen equal to zero. Nevertheless the metal does not emit electrons, that are confined in it, and this corresponds in having an high potential barrier U_0 at the metal interface. The height U_0 of the potential barrier will be of the order of the metal work function, that corresponds to the energy to be provided in order to extract an electron from the metal.
As already said, the case with negative energy $E < 0$ is not of interest. Instead let's start from the free states with energy $E > U_0$. This case is similar to the already considered one of potential barrier with energy E greater than the potential energy barrier one. The procedure of proceeding is exactly analogous and contrarily to classical mechanics it turns out that there is a finite probability of reflection in the two discontinuity points $x = -\frac{L}{2}$ and $x = +\frac{L}{2}$. It is possible to get explicit expressions for the transmission and the reflection coefficients, and also in this case it is possible to verify an energy dependence of these coefficients, with both transmission peaks and reflection peaks. A full treatment is provided in [210]. Notice that no constraints on the range of energy values is recovered thus they

can be continuous.
The attention is now focused on the most interesting case: the one of bound states, i.e. with energy $0 \leq E \leq U_0$.
In region I the potential is $U(x) = U_0 > E$ and the steady state Schrödinger's equation becomes:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi_I(x) + U_0\psi_I(x) = E\psi_I(x) \quad \rightarrow \quad \frac{d^2}{dx^2}\psi_I(x) - \frac{2m}{\hbar^2}(U_0 - E)\psi_I(x) = 0$$

$$\rightarrow \quad \frac{d^2}{dx^2}\psi_I(x) - \beta_k^2\psi_I(x) = 0 \quad \rightarrow \quad \frac{d^2}{dx^2}\psi_I(x) + k_1^2\psi_I(x) = 0$$

$$with: \quad k_1 = i\beta_k \quad , \quad \beta_k = \sqrt{\frac{2m}{\hbar^2}(U_0 - E)} \in \mathbb{R}^+$$

The characteristic polynomial is: $\lambda^2 - \beta_k^2 = 0$ from which the two roots are: $\lambda_{1,2} = \pm\beta_k = = \pm\sqrt{\frac{2m}{\hbar^2}(U_0 - E)}$. The general solution is thus:

$$\psi_I(x) = Ae^{+\beta_k x} + Be^{-\beta_k x} \quad , \quad with: \ A, B \in \mathbb{R}$$

In order to ensure an everywhere limited wave-function $B = 0$, because otherwise the limit for $x \rightarrow -\infty$ diverges. Thus the solution becomes:

$$\psi_I(x) = Ae^{+\beta_k x} \quad , \quad A \in \mathbb{R}$$

In region II the potential is null: $U(x) = 0$. Thus the solution is the one for free electrons:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi_{II}(x) = E\psi_{II}(x) \qquad \rightarrow \quad \frac{d^2}{dx^2}\psi_{II}(x) + k_2^2\psi_{II}(x) = 0$$

$$with: \quad k_2 = k = \sqrt{\frac{2m}{\hbar^2}E} \in \mathbb{R}^+$$

Since no confusion is possible k_2 is simply called k. The characteristic polynomial is: $\lambda^2 + k^2 = 0$; from which the two roots are: $\lambda_{1,2} = \pm ik = \pm i\sqrt{\frac{2m}{\hbar^2}E}$. The general solution is thus:

$$\psi_{II}(x) = Ce^{+ikx} + De^{-ikx} \quad , \quad with: \ C, D \in \mathbb{C}$$

No constraints are present on C and D since the exponentials are complex and the $\psi_{II}(x)$ is limited everywhere.
In region III the potential is again $U(x) = U_0 > E$, thus the solution is formally the same of region I, with the same β_k (since U_0 is supposed to be the same also β_k is the same). The solution can be written as:

$$\psi_{III}(x) = Fe^{+\beta_k x} + Ge^{-\beta_k x} \quad , \quad with: \ F, G \in \mathbb{R}$$

In order to ensure a limited wave-function everywhere $F = 0$, because otherwise $\psi_{III}(x)$ would diverge for $x \rightarrow +\infty$. Thus the solution in region III becomes:

$$\psi_{III}(x) = Ge^{-\beta_k x} \quad , \quad G \in \mathbb{R}$$

In summary the total wave-function is:

$$\psi(x) = \begin{cases} \psi_I(x) = Ae^{+\beta_k x} & if \ x < -\frac{L}{2} \qquad (region \ I) \\ \psi_{II}(x) = Ce^{+ikx} + De^{-ikx} & if \ -\frac{L}{2} \leq x \leq +\frac{L}{2} \quad (region \ II) \\ \psi_{III}(x) = Ge^{-\beta_k x} & if \ x > +\frac{L}{2} \qquad (region \ III) \end{cases}$$

By enforcing the boundary conditions on continuity of $\psi(x)$ and its first derivative in $x = -\frac{L}{2}$ and in $x = +\frac{L}{2}$ it is possible to find the values for the constants A, C, D and G. In order to do that it is simpler to rewrite the wave-function as follows:

$$\psi_{II}(x) = Ce^{+ikx} + De^{-ikx} = C\left[cos(kx) + isin(kx)\right] + D\left[cos(kx) - isin(kx)\right] =$$
$$= (C + D)\, cos(kx) + (iC - iD)\, sin(kx)$$

and said: $K_1 = A$, $K_2 = C + D$, $K_3 = iC - iD$ and $K_4 = G$ the wave-function becomes:

$$\psi(x) = \begin{cases} \psi_I(x) = K_1 e^{+\beta_k x} & if\ x < -\frac{L}{2} \quad (region\ I) \\ \psi_{II}(x) = K_2 cos(kx) + K_3 sin(kx) & if\ -\frac{L}{2} \leq x \leq +\frac{L}{2} \quad (region\ II) \\ \psi_{III}(x) = K_4 e^{-\beta_k x} & if\ x > +\frac{L}{2} \quad (region\ III) \end{cases}$$

The continuity conditions are:

$$\begin{cases} \psi_I(-\frac{L}{2}) = \psi_{II}(-\frac{L}{2}) & (cond.\ 1) \\ \psi_{II}(+\frac{L}{2}) = \psi_{III}(+\frac{L}{2}) & (cond.\ 2) \\ \frac{d}{dx}\psi_I(x)|_{x=-\frac{L}{2}} = \frac{d}{dx}\psi_{II}(x)|_{x=-\frac{L}{2}} & (cond.\ 3) \\ \frac{d}{dx}\psi_{II}(x)|_{x=+\frac{L}{2}} = \frac{d}{dx}\psi_{III}(x)|_{x=+\frac{L}{2}} & (cond.\ 4) \end{cases}$$

For example from (*cond.* 1) one gets:

$$K_1 e^{-\beta_k \frac{L}{2}} = K_2 cos(k\frac{L}{2}) - K_3 sin(k\frac{L}{2})$$

$$\rightarrow \quad K_1 e^{-\beta_k \frac{L}{2}} - K_2 cos(k\frac{L}{2}) + K_3 sin(k\frac{L}{2}) + K_4 \cdot 0 = 0$$

Proceeding analogously for the other three conditions an homogeneous linear system of four equations in four unknowns (that are K_1, K_2, K_3 and K_4) is recovered. Then it is possible to select only non-trivial solutions by enforcing its determinant (in matrix form) to be null (a trivial solution corresponds to all four constants K_1, K_2, K_3 and K_4 null, that means no state or null state $\psi(x) = 0$). The result is a mathematical condition to be satisfied by both k and β_k that leads thus to a condition on the allowed energy values (since both k and β_k are function of energy E). It turns out that the only possible energy values for steady states in a quantum well are discrete. Nevertheless a full analytical solution is not possible in the case of finite height quantum well. Notice that the quantization here comes out naturally, by enforcing the boundary conditions on $\psi(x)$. This is the main difference between a quantum theory based on the Schrödinger's solution and the first attempts of Bohr, in which the quantization was introduced *ad hoc*. Thus bound states have quantized energy levels, that is: the Hamiltonian operator (in a bound system) has a discrete spectrum for $0 < E < U_0$ (bound states), while has a continuous spectrum for $E > U_0$ (free states). The same result can be obtained in an another way by noting that the potential $U(x)$ has an even symmetry, and thus exploiting symmetry conditions of the wave-functions. Nevertheless, also in this case a full analytical solution is not possible and at a certain point a numerical or graphical solution of a system is required. This procedure is highlighted in appendix **??**.

Of course whatever is the method used for the solution of this problem the final result does not change: a finite number of discrete energy levels are present (or allowed) in the quantum well. Thus discretization is the main feature of the bound states. It is possible to show that the number of quantized energy levels increases with the height U_0 of the well and with the square of its length L. The quantity $U_0 \left(\frac{L}{2}\right)^2$ is sometimes called the "binding power" of the well since it represents the strength of the bond. If the binding power is increased then the number of discrete energy levels is increased, i.e. the number

of permitted bound states increases, thus meaning that the intensity of the attractive force is greater.
It is possible to solve numerically the steady state Schrödinger's equation for the case of a quantum well of height U_0. Examples of numerical results, obtained in *MatLab* environment by means of an implementation of the finite difference method presented in section 17.4 are reported in figures 17.10, 17.11 and 17.12. The *MatLab* code used to creates these figures is reported in appendix **??**. Notice that when the well is increased or enlarged the number of bound eigenstates is increased, as already pointed out.
In conclusion, the Hamiltonian operator, in the case of a finite height quantum well, presents a mixed spectrum: the bound states have discrete energy eigenvalues, corresponding to the discrete part of the spectrum, while the free states have a continuous range of energies E, thus corresponding to the continuous part of the spectrum.

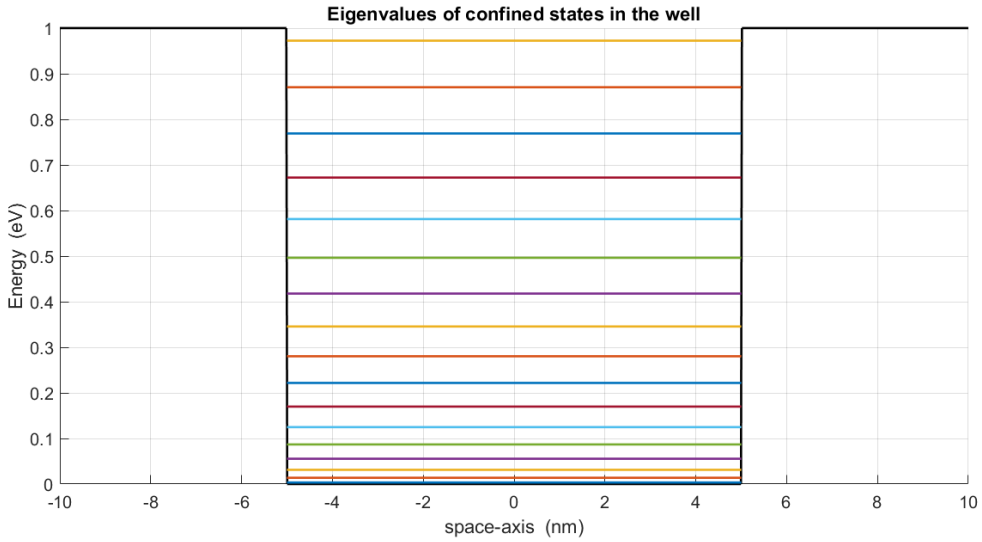


Figure 17.10: Permitted energy eigenvalues in a 1D potential energy well with $L = 10\,\text{nm}$ and $U_0 = 1\,\text{eV}$. Notice that they are discrete, and only 17 bound states are possible for this specific quantum well. In black the quantum well shape (in eV).

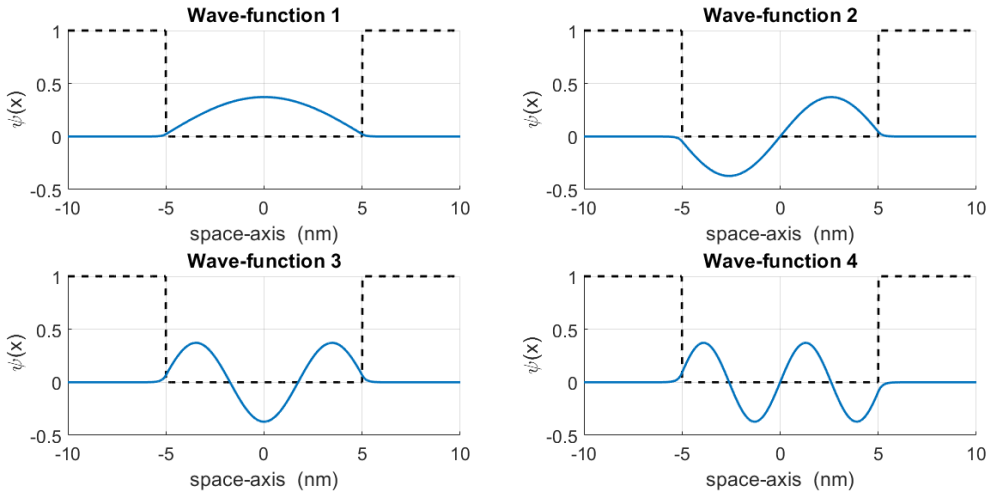


Figure 17.11: First four wave-functions corresponding to the first four energy levels of figure 17.10. The same quantum well is considered: 1D potential energy well with $L = 10\,\text{nm}$ and $U_0 = 1\,\text{eV}$. Notice that the exponential tails in the classical forbidden regions. The wave-functions are unnormalized; in black the quantum well shape (in eV).

1D infinite height potential well

In this section an infinite height quantum well is considered. The difference w.r.t. previous section is that $U_0 \to +\infty$ at the well boundaries. In this case there are no continuum states, since there cannot exist particles with energy E greater than infinity: $E > U_0 \to +\infty$. Thus only bound states will be found. In this case it is possible to get an analytical solution, as it will be shown in a while. The procedure is always the same of the previously investigated 1D problems. Before summarizing the main steps notice that if in the finite height well it is made the limit for $U_0 \to +\infty$, the real decaying exponentials in the two classically forbidden regions I and III tend to zero. Thus it is expected to find perfectly confined wave-functions, that go to zero at the well boundaries.
The potential shape is assumed to be analogous to the one of figure 17.9 but with $U_0 = +\infty$. The calculations are a little bit simplified with a shift in the x-coordinate axis such that the well is in between $x = 0$ and L. Figure 17.13 represents the new potential shape. The fact that the potential energy blows up to infinity at the well boundaries represents that very strong forces are acting in the two boundary points $x = 0$ and $x = L$, resulting in a full reflection of the particle. Indeed $\psi(x = 0) = 0 = \psi(x = L)$ means no probability of finding the particle in $x = 0$ and $c = L$, that means that it is completely reflected (conservation of matter and total probability holds).
The Schrödinger's equation in region I and III is analogous to the one already solved in the previous section, but with $U_0 \to +\infty$. It it possible to write the exponential functions and then make them tend to zero. The steps are summarized in the following equations:

$$\frac{d^2}{dx^2}\psi_I(x) + k_1^2\psi_I(x) = 0 \quad with: \quad k_1 = k_3 = i\beta_k \quad , \quad \beta_k = \sqrt{\frac{2m}{\hbar^2}(U_0 - E)}$$

The general solutions thus are:

$$\psi_I(x) = Ae^{+\beta_k x} + Be^{-\beta_k x} \quad , \quad \psi_{III}(x) = Ce^{+\beta_k x} + De^{-\beta_k x}$$

In order to ensure an everywhere limited wave-function it must be set: $B = 0$ and $C = 0$. And since $\beta_k = \sqrt{\frac{2m}{\hbar^2}(U_0 - E)} \to +\infty$ when $U_0 \to +\infty$ the two wave-functions tend to zero (for $\psi_I(x)$ remember that x is negative $x < 0$, thus making the exponent negative since β_k is positive):

$$\psi_I(x) = Ae^{+\beta_k x} \to 0 \quad , \quad \psi_{III}(x) = De^{-\beta_k x} \to 0$$

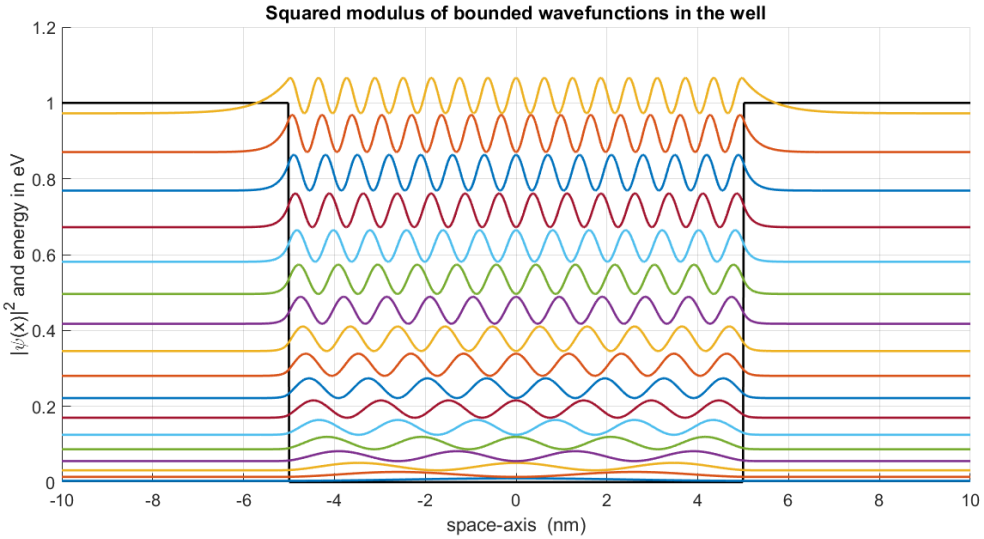


Figure 17.12: Squared moduli of all the 17 wave-functions corresponding to the 17 energy eigenvalues of figure 17.10. The same quantum well is considered: 1D potential energy well with $L = 10\,\text{nm}$ and $U_0 = 1\,\text{eV}$. The squared moduli are positioned at a height corresponding to their relative eigenvalue for aesthetic reasons. They are unnormalized; in black the quantum well shape (in eV).

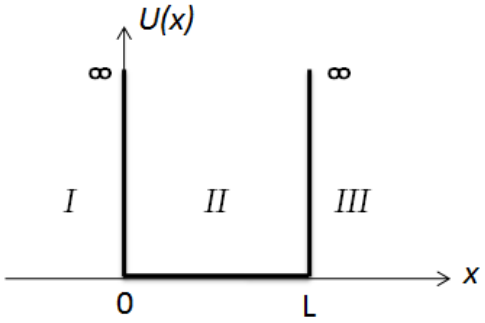


Figure 17.13: Infinite height 1D potential energy well. The well height is ∞ at $x = 0$ and $x = L$; the well width is L, the potential has even symmetry.

In region II the Schrödinger's equation is the one for the free particle:

$$\frac{d^2}{dx^2}\psi_{II}(x) + k_2^2\psi_{II}(x) = 0 \quad with: \quad k_2 = k = \sqrt{\frac{2m}{\hbar^2}E} \in \mathbb{R}^+$$

whose solution is:

$$\psi_{II}(x) = \psi(x) = Fe^{+ikx} + Ge^{-ikx}$$

By enforcing the continuity in $x = 0$ one gets $\psi(x = 0) = 0 = F + G$ from which $G = -F$, and so:

$$\psi(x) = Fe^{+ikx} - Fe^{-ikx} = F(e^{+ikx} - e^{-ikx}) = 2iFsin(kx) = Ksin(kx)$$

where $K = 2iF$ is constant. The boundary condition in $x = L$ leads $\psi(x = L) = 0 = Ksin(kx)$, but since the constant K cannot be null (otherwise $\psi(x) = 0$ for each x value, but this means no state - trivial solution), it can be rewritten as:

$$sin(kL) = 0 \quad \Leftrightarrow \quad kL = 0 + n\pi \quad , \quad n \in \mathbb{N} \setminus \{0\}$$

Notice that $n = 0$ would lead to zero k value that is again the trivial solution. Moreover in general it would be mathematically acceptable an integer $n \in \mathbb{Z} \setminus \{0\}$, but for positive

and negative symmetric n values the same wave-function would be recovered due to the *sin* function anti-symmetry, thus it has sense to choose $n \in \mathbb{N} \setminus \{0\}$. Solving for the momentum gives:

$$k = \frac{n\pi}{L} \quad , \quad p = \hbar k = \frac{n\pi\hbar}{L} \quad , \quad n \in \mathbb{N} \setminus \{0\} \tag{17.65}$$

The energy eigenvalues of the particles are then given by:

$$E = \frac{\hbar^2 k^2}{2m} = \frac{n^2\pi^2\hbar^2}{2mL^2} \quad , \quad n \in \mathbb{N} \setminus \{0\} \tag{17.66}$$

In conclusion, the electrons (or particles) in an infinite height quantum well cannot have arbitrary energy values but only discrete values given by eq. (17.66), i.e. the energy is quantized. This is a general result that holds whenever quantum confinement occurs, due to a potential shape that confines the particles in a certain region of space. Notice again that quantization comes out naturally by enforcing the boundary conditions on the general solutions of the steady state Schrödinger's equation, and it is not introduced *ad hoc*. An acceptable wave-function that satisfies these boundary conditions is possible only for certain values of energy. In the case of a step or of a barrier the particles were not confined and indeed continuum states were present, with in general a continuous energy E.
Moreover notice that the minimum energy for a particle in a quantum well is $E_1 = \frac{\pi^2\hbar^2}{2mL^2} > 0$ and not zero. This minimum energy is related to the uncertainty principle, indeed the uncertainty on the particle position is of the order of L: $\Delta x \sim L$. The particle is moving back and forth in the 1D well with a momentum p, leading to an uncertainty on the momentum of around $\Delta p \sim 2p$. The uncertainty principle (eq. (17.13)) requires $\Delta x \Delta p \geq h$, thus $2Lp \geq h$ that is $p \geq \pi\hbar/L$, that gives: $E \geq E_1$. The existence of a "zero point energy", like E_1, is typical of all the quantum confinement problems.
Finally the wave-functions $\psi(x) = Ksin(kx)$ can be normalized. From the normalization condition (eq. (17.19)):

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 \, dx = |K|^2 \int_0^L sin^2\left(\frac{n\pi}{L}x\right) dx = 1$$

The value of the integral is $L/2$, thus: $|K|^2 \frac{L}{2} = 1$, from which $K = \sqrt{\frac{2}{L}}$. Therefore the normalized wave-functions are:

$$\psi_{normalized}(x) = \sqrt{\frac{2}{L}} sin\left(\frac{n\pi}{L}x\right) \quad , \quad n \in \mathbb{N} \setminus \{0\} \tag{17.67}$$

3D potential well

In this section a 3D quantum well is considered. It means that the potential is of the kind of the one of figure 17.13 in all the three directions x, y and z. Since quantum confinement occurs in all directions, this structure is sometimes also called "quantum dot", or better: it can be an oversimplified representation of a real quantum dot. In general the lengths of the sides of potential "box" can be different, they are assumed to be L_x, L_y and L_z. Assuming that the potential can be expressed as the sum of the three different wells (as happens for a potential box just described), it is possible to speed up the solution by exploiting the separation of variables presented at the beginning of this section. In this way the total energy of a particle (electron) inside the 3D potential box is simply the sum of the energy eigenvalues in the three directions: $E = E_x + E_y + E_z$. From equations (17.65) and (17.66) one gets:

$$k_x = \frac{n_x\pi}{L_x} \quad , \quad k_y = \frac{n_y\pi}{L_y} \quad , \quad k_z = \frac{n_z\pi}{L_z} \quad , \quad n_x, n_y, n_z \in \mathbb{N} \setminus \{0\} \tag{17.68}$$

and said: $k^2 = k_x^2 + k_y^2 + k_z^2$, the energy eigenvalues are then given by:

$$E = E_x + E_y + E_z = \frac{\hbar^2 k_x^2}{2m} + \frac{\hbar^2 k_y^2}{2m} + \frac{\hbar^2 k_z^2}{2m} = \frac{\hbar^2 k^2}{2m}$$

$$E = \frac{\pi^2\hbar^2}{2m}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_y^2}{L_y^2}\right) \quad , \quad n_x, n_y, n_z \in \mathbb{N} \setminus \{0\} \tag{17.69}$$

Then the 3D wave-functions are given by the product of the 1D ones, thus:

$$\psi(x,y,z) = \psi(x)\psi(y)\psi(z) = K_x sin\left(k_x x\right) K_y sin\left(k_y y\right) K_z sin\left(k_z z\right)$$

$$\rightarrow \quad \psi(x,y,z) = C sin\left(\frac{n_x\pi}{L_x}x\right) sin\left(\frac{n_y\pi}{L_y}y\right) sin\left(\frac{n_z\pi}{L_z}z\right) \tag{17.70}$$

where $C = K_x K_y K_z = (2/L)^{3/2}$ (see eq. (17.67)). An interesting case is the one of $L_x = L_y = L_z = L$. In this case the above expressions become:

$$E = \frac{\hbar^2 k^2}{2m} = \frac{\pi^2\hbar^2}{2mL^2}\left(n_x^2 + n_y^2 + n_z^2\right) \quad , \quad n_x, n_y, n_z \in \mathbb{N} \setminus \{0\} \tag{17.71}$$

$$\psi(x,y,z) = C sin\left(\frac{n_x\pi}{L}x\right) sin\left(\frac{n_y\pi}{L}y\right) sin\left(\frac{n_z\pi}{L}z\right) \tag{17.72}$$

Notice that the energy eigenvalues depend only on $(n_x^2 + n_y^2 + n_z^2)$, this means that all the set of n_x, n_y and n_z that provide the same value of $(n_x^2 + n_y^2 + n_z^2)$, provide also the same energy level. Nevertheless when n_x, n_y and n_z are changed, even without changing E, they make changing the wave-function $\psi(x,y,z)$, thus the same energy level may correspond to different wave-functions or states and degeneracy occurs.

General potential shapes

So far various cases were treated, and the main result is that the quantum confinement gives rise to energy quantization. This holds in general and not only for energy eigenvalues. Indeed in section 17.3.6 it was pointed out that the steady state Schrödinger's equation is the eigenvalue problem for the Hamiltonian operator $\widehat{H}$, but in general it is possible to write an eigenvalue problem for each operator $\widehat{F}$ representing whatever physical observable F. By doing so an equation of the kind of:

$$\widehat{F}\psi_i = f_i\psi_i \tag{17.73}$$

is obtained. Its solutions corresponds to the physical observable eigenvalues $\{f_i\}_i$ and eigenstates (or eigenfunctions) $\{\psi_i\}_i$. The meaning (that comes out from the way in which eq. (17.73) is derived - see section 17.3.6 for details), is the following: by performing a measurement on F it is possible to find as a result one of the possible $\{f_i\}_i$ values, let's say f_n, that means that the system was in the corresponding state ψ_n. It is possible to verify that quantum confinement (of the kind of the quantum well investigated so far) does imply quantization, not only of energy levels but also of the other physical observables such as angular momentum and so on... This can be verified by solving the relative eigenvalue problems of the kind of eq. (17.73).

In general an operator can have a discrete spectrum (i.e. its eigenvalues are quantized), a continuous spectrum (i.e. its eigenvalues can assume each value in a real range of values with continuity) or a mixed spectrum (a mix of the previous cases is possible). For example in the general case of a potential energy of the kind of the one in figure 17.14, the Hamiltonian operator presents a mixed spectrum. Indeed, as indicated in the figure, for negative energies bound states are present (due to confinement) and its spectrum is discrete (E_1, E_2, E_3, etc... in the figure); while for $E > 0$ no confinement occurs and the energy is continuous (continuum states or unbound states).

It may be thus useful to gain an intuition about the physical insights of system under study, starting from the facts so far known. In figures 17.17, 17.18 17.15, 17.16 are reported some examples. In order to understand the physical results that can be obtained by a simulator it may be useful to intuitively draw the wave-functions specifying if they are propagating waves or attenuating in each region (before solving the equations). To this purpose a good reference may be [211].

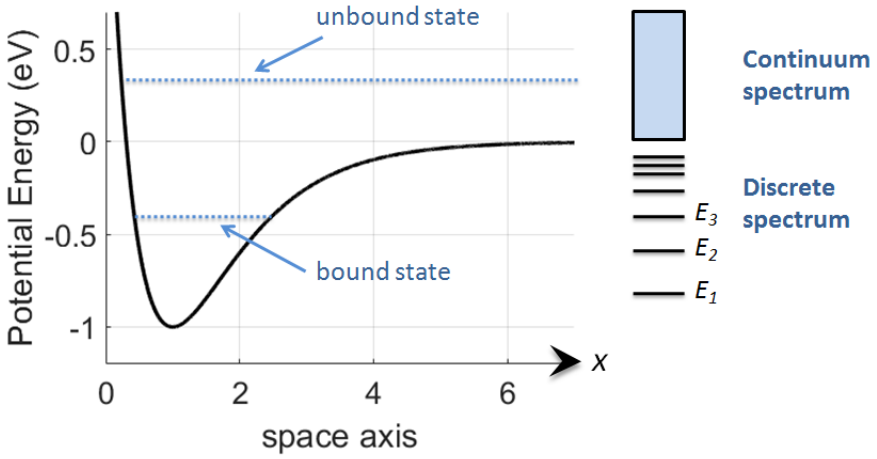


Figure 17.14: General potential energy shape example, typical of a central force problem. At large x value the interaction vanishes thus $U(x) \to 0$. Instead for $x = 0$ the potential energy tends to infinity $U(x) \to \infty$ because of the repulsion. For small x values an attractive interaction occurs and $U(x)$ is negative. On right side of the potential shape is reported the Hamiltonian operator spectrum (energy eigenvalues), discrete for negative energies (bound states) and continuous for positive energies (free states).

17.4 Finite difference method and matrix representation

The purpose of this section is to illustrate how a differential operator, such as the Hamiltonian one $\widehat{H}$, can be represented in matrix form. This will be useful to understand next theoretical topics in this thesis work, especially related to the Non-Equilibrium Green's Function formalism (see chapter **??**). As already mentioned in the introductory sections of this chapter, this representation is also the foundation of the Heisenberg's formalism of quantum mechanics, the so called "matrix mechanics", in which operators are indeed matrices. The task will be addressed in two steps, following the approach of [44].
Firstly, a finite difference discretization of the steady state Schrödinger's equation is presented. It is interesting to notice that Heisenberg initially conceived his matrix mechanics, starting from finite difference discretization of differential operators. In particular he developed a theory that involved only physical quantities directly observable. In doing so, he exploited the Fourier analysis to rewrite the classical dynamic (differential) equations in finite difference equations for the state transition probabilities, from one state to another state of the quantum system. Notice that this formulation was initially conceived by Heisenberg, but in the form known today, it is the result of the work of many other scientists, among which Dirac produced a great contribution.

Then, the general method for representing an operator in matrix form and the Dirac's notation will be briefly addressed.

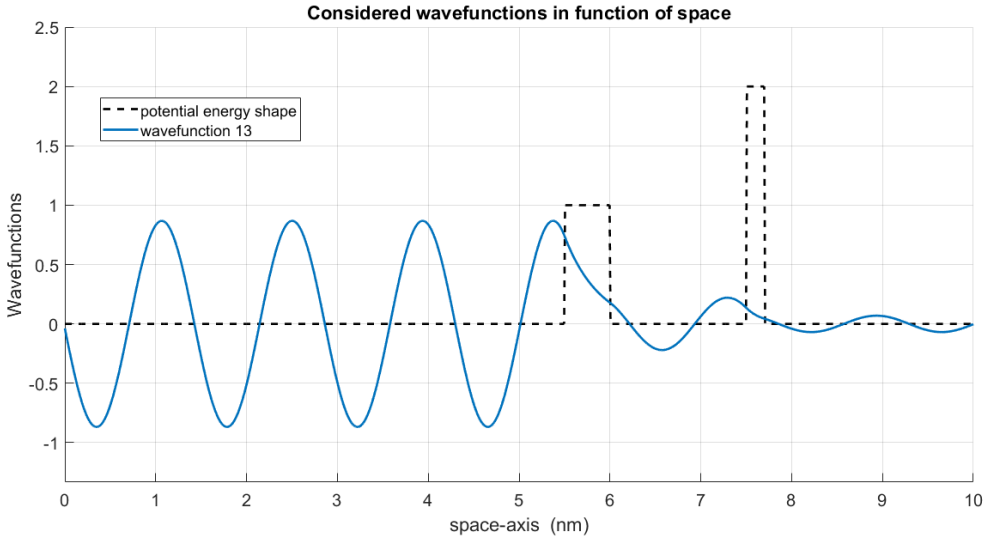


Figure 17.15: Wave-function corresponding to an electron tunneling through two consecutive potential barriers. The first barrier is 0.5 nm thick and 1 eV of height, the second is 0.25 nm thick and 2 eV of height. The electron energy is $E = 0.73$ eV. Notice the exponential decay within the barriers; the wave-function is unnormalized. The graph was generated with the *MatLab* code reported in appendix **??**.

17.4.1 Finite difference method for the solution of the steady state Schrödinger's equation

The 1D Hamiltonian operator is:

$$\widehat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + U(x)$$

It is possible to think of solving numerically the steady state Schrödinger's equation, and to do that it is necessary to discretize the domain x and the Hamiltonian operator.
The space axis can be discretized in N of nodes: $x = (x_1, x_2, x_3, ..., x_n, ..., x_{N-1}, x_N)$, in this way it becomes an array. Let's suppose a uniform spacing discretization, that means the the step between two nodes in the space array is constant and fixed to $a = x_{n+1} - x_n$ (for each considered n). Since the potential energy shape $U(x)$ is supposed known it can be also discretized in a set of N equidistant points: $U = (U_1, U_2, U_3, ..., U_n, ..., U_{N-1}, U_N)$, where the generic U_n is intended as: $U_n = U(x_n)$. The constant $-\frac{\hbar^2}{2m}$ is unchanged. The second derivative operator can be instead discretized by means of a finite difference formula, that is intuitively derived from the geometrical meaning of derivative. In numerical methods courses the finite difference method is usually treated, nevertheless what is needed here is just the finite difference formula for the second order derivative. Thinking that the derivative is defined by means of a limit of the incremental ratio, it is possible to approximate the first order derivative with one of the two formulae (forward or backward finite difference formula):

$$\frac{d\psi}{dx} \sim \frac{\psi_{n+1} - \psi_n}{a} \quad , \quad or: \quad \frac{d\psi}{dx} \sim \frac{\psi_n - \psi_{n-1}}{a}$$

where $\psi_n = \psi(x_n)$ and $a = x_{n+1} - x_n = x_n - x_{n-1}$. The second order derivative can be obtained by deriving two times w.r.t. space, thus:

$$\frac{d^2\psi}{dx^2} \sim \frac{\frac{\psi_{n+1}-\psi_n}{a} - \frac{\psi_n - \psi_{n-1}}{a}}{a} = \frac{\psi_{n+1} - 2\psi_n + \psi_{n-1}}{a^2}$$

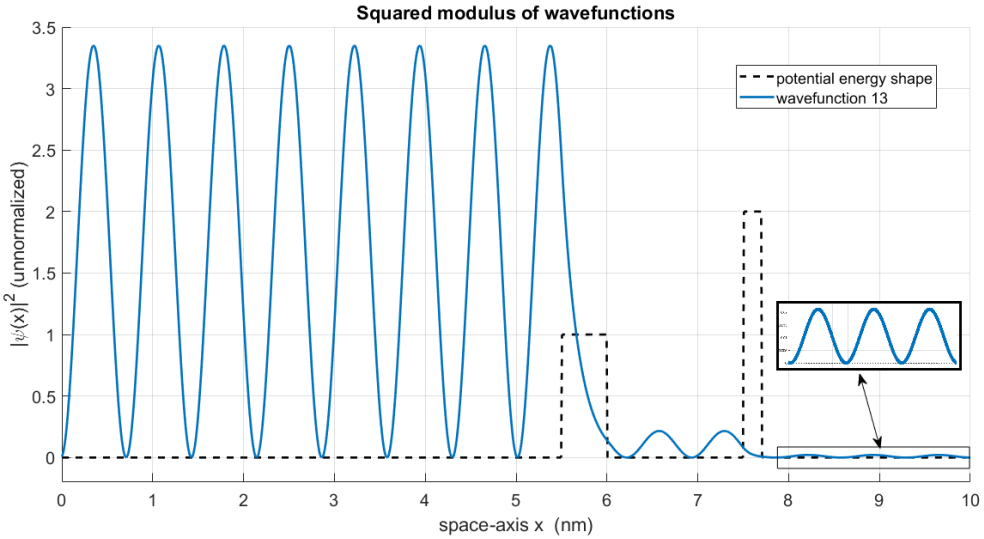


Figure 17.16: Wave-function squared modulus corresponding to an electron tunneling through two consecutive potential barriers. The potential shape and the structure is the same of figure 17.15. Notice the exponential decay within the barriers; the wave-function is unnormalized. The graph was generated with the *MatLab* code reported in appendix **??**.

It follows that the Hamiltonian operator, when it is applied to the wave-function $\psi(x)$ in the node x_n of the discretized domain, becomes:

$$\widehat{H}\psi(x_n) = \widehat{H}\psi_n \sim -\frac{\hbar^2}{2m}\frac{\psi_{n+1} - 2\psi_n + \psi_{n-1}}{a^2} + U_n\psi_n$$

Notice that it depends on the potential only in that node $U_n = U(x_n)$, and on the wave-function at nodes $n-1$, n and $n+1$. It is now defined the quantity t_0 (accordingly with [44]):

$$t_0 = +\frac{\hbar^2}{2ma^2}$$

and the product of the Hamiltonian operator applied to ψ_n becomes:

$$\widehat{H}\psi_n \sim -t_0\left(\psi_{n+1} - 2\psi_n + \psi_{n-1}\right) + U_n\psi_n = -t_0\psi_{n-1} + \left(2t_0 + U_n\right)\psi_n - t_0\psi_{n+1} \quad (17.74)$$

At this point the steady state Schrödinger's equation can be discretized exploiting the finite difference method, and it can be translated in a matrix equation:

$$\widehat{H}\psi = E\psi \quad \rightarrow \quad \left[H\right]\left\{\psi\right\} = E\left\{\psi\right\}$$

where the symbol $\left[H\right]$ indicates a matrix of dimension NxN and the symbol $\left\{\psi\right\}$ a column vector of dimension N. The product $\left[H\right]\left\{\psi\right\}$ is a matrix product (rows by columns). Since the application of $\widehat{H}$ to ψ_n involves the nodes $n-1$, n and $n+1$ the Hamiltonian matrix $\left[H\right]$ will be tridiagonal, with $(2t_0 + U_n)$ on the main diagonal, and $-t_0$ on the upper and lower diagonal (see eq. (17.74) for a reference). The final result is thus:

$$\begin{bmatrix} 2t_0+U_n & -t_0 & 0 & 0 & 0 & 0 & \dots \\ -t_0 & 2t_0+U_n & -t_0 & 0 & 0 & 0 & \dots \\ 0 & -t_0 & 2t_0+U_n & -t_0 & 0 & 0 & \dots \\ 0 & 0 & -t_0 & 2t_0+U_n & -t_0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ & & & & & & \end{bmatrix} \begin{Bmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \psi_n \\ \dots \\ \psi_N \end{Bmatrix} = E \begin{Bmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \psi_n \\ \dots \\ \psi_N \end{Bmatrix}$$

Notice that by performing the matrix product between $\left[H\right]$ and $\left\{\psi\right\}$ equation (17.74) is recovered.

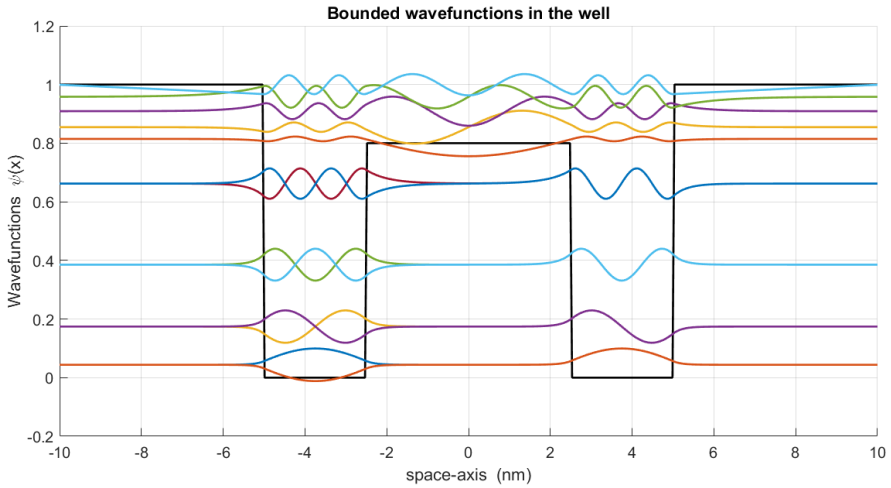


Figure 17.17: Confinement in a quantum well. The well is 10 nm long and 1 eV tall. Inside it a barrier of 5 nm of width and height 0.8 eV is present. The potential shape is the curve in black. In total 13 bound states are present in this structure. Their wave-functions are reported in this graph; they are unnormalized and they are placed at an height corresponding to their relative energy eigenvalues. Notice the degeneracy and the spacing between the enrgy levels. The graph was generated with the *MatLab* code reported in appendix **??**.

17.4.2 Matrix representation of differential operators

The purpose of this section is to address the matrix representation of differential operators. In order to do that, the important concept of basis set will be also introduced. These topics will be further discussed in the next section, where the Dirac's notation will be addressed. In this section I will follow the approach of [44], while in the next one the one of [210].
In section 17.3.5 it was mentioned that the general solution of the time-dependent Schrödinger's equation can be always be expressed as superposition of the steady state factorized solutions, that are solution of the steady state Schrödinger's equation, i.e. of the Hamiltonian operator eigenvalue problem. This result was summarized in eq. (17.49), in which the general case of a mixed spectrum Hamiltonian was considered. Equation (17.49) is reported here for convenience:

$$\psi(\vec{r};t) = \sum\!\!\!\!\!\!\int dE\ C(E)\ \Psi_E(\vec{r})\ e^{-\frac{i}{\hbar}Et}$$

The important fact, that enables to write the most general wave-function $\psi(\vec{r};t)$ as a superposition of the steady state (time-dependent) wave-functions $\Psi_E(\vec{r})\ e^{-\frac{i}{\hbar}Et}$, by means of the coefficients $C(E)$, is that the steady state wave-functions are an orthogonal (or if normalized an orthonormal) and complete set of functions. Exactly like an orthonormal set of N vectors can be used as a basis for a representation of each vector in the N-dimensional vector space, the orthonormal complete set of wave-functions can be used as a basis for the representation of whatever wave-function $\psi(\vec{r};t)$. Notice that "complete" means exactly that this procedure, namely the representation of a whatever function as superposition of the functions belonging to the orthonormal complete set, can be performed. In general, as pointed out in section 17.3.3, this procedure is analogous at all to perform a Fourier series expansion (for discrete systems) or transform (for continuous systems). Indeed as known from signal theory courses, the meaning of the Fourier analysis is to change the basis of representation of the function (signal) from time domain to frequency domain, or better: to express the signal as a superposition of purely harmonic (monochromatic) signals. A complete set of functions is thus a set of functions that can be used as basis for representing whatever function. The general function $\psi(\vec{r};t)$ can be thus expressed as a linear combination (i.e. a superposition) of the basis functions, each suitably weighted

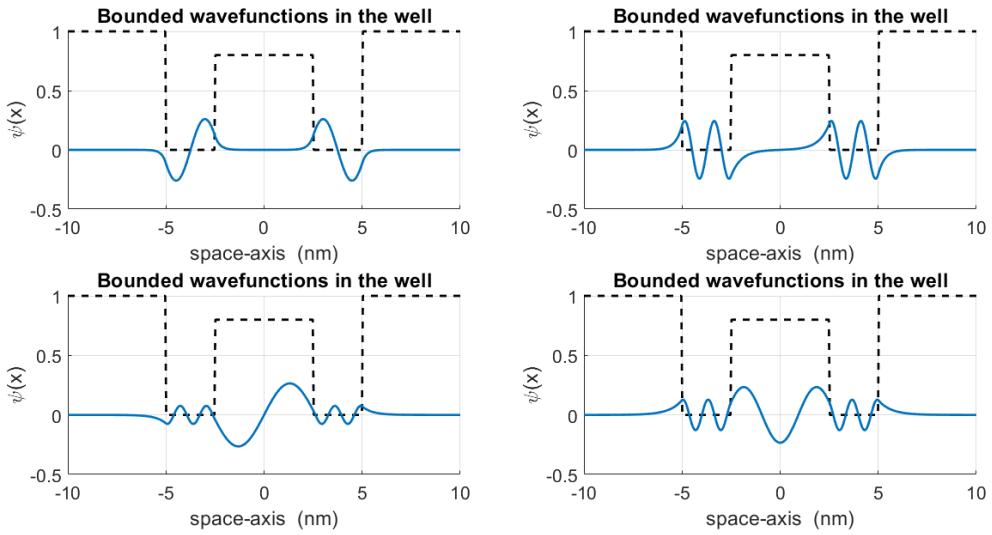


Figure 17.18: Confinement in a quantum well. The well is the same of picture 17.17. Here four wave-functions are reported, they correspond to energy levels: $E_3 = 0.174\,\text{eV}$ (top left), $E_8 = 0.662\,\text{eV}$ (top right), $E_{10} = 0.815\,\text{eV}$ (bottom left) and $E_{11} = 0.855\,\text{eV}$ (bottom right). Notice the exponential decay in the central region (of height $0.8\,\text{eV}$) of the top ones, while the bottom ones are free in that region (oscillating). The wave-functions are unnormalized. The graph was generated with the *MatLab* code reported in appendix **??**.

by means of the coefficient C. These coefficient can be interpreted exactly like Fourier coefficients (see e.g. [217] and [216]), since the procedure is essentially a Fourier expansion. Moreover these (Fourier) coefficients have an analogous meaning to the coefficients of a vector w.r.t. a new vector basis for the vector space: i.e. they are the projection of the general function $\psi(\vec{r};t)$ on the representation basis $\Psi_E(\vec{r})\ e^{-\frac{i}{\hbar}Et}$. For conventional vectors this corresponds to a scalar product, for functions it is the same but the scalar functional product must be considered. It is formally defined by equation (17.32). In the case of the projection of the general wave-function $\psi(\vec{r};t)$ onto the basis set made by the steady states it becomes:

$$C(E) = \left(\Psi_E(\vec{r})e^{-\frac{i}{\hbar}Et}, \psi(\vec{r};t)\right) = \int \Psi_E^*(\vec{r})e^{+\frac{i}{\hbar}Et}\psi(\vec{r};t)d\vec{r}$$

Let's focus for the moment only in the case of a discrete spectrum of the Hamiltonian operator. In this case equation (17.49) becomes a summation:

$$\psi(\vec{r};t) = \sum_i C(E_i)\ \Psi_{Ei}(\vec{r})\ e^{-\frac{i}{\hbar}E_i t}$$

i.e. it is a Fourier series (discrete case). The general 1D Fourier series expansion of the function $f(x)$ in the basis of harmonic exponential functions $\{e^{i\frac{2\pi n}{P}x}\}_n$ (where P is a constant often assumed equal to the period of $f(x)$ or equal to 2π) is usually written as (see signal processing courses or [217] and [216]):

$$f(x) = \sum_n c_n e^{i\frac{2\pi n}{P}x} \quad , \quad c_n = \frac{1}{P}\int f(x)e^{-i\frac{2\pi n}{P}x}dx = \left(e^{i\frac{2\pi n}{P}x}, f(x)\right)$$

From which it is evident that the coefficients $C(E_i)$ are essentially Fourier coefficient, with the meaning discussed above, for the basis set $\Psi_{Ei}(\vec{r})\ e^{-\frac{i}{\hbar}E_i t}$. The difference w.r.t. to the usual Fourier series is only the choice of the basis set, that usually is the set of harmonic waves (complex exponentials). At this point it should be very clear what discussed above, and the general theory can now be addressed.

All these concepts hold in general and not only for the steady state wave-functions. In particular the only requirement to express whatever function (wave-function) as a superposition (summation) of basis functions is that the set of basis functions is orthogonal (if normalized orthonormal) and complete. This is always true for the eigenfunctions of an Hermitian operator. Thus if the quantum mechanical operator is Hermitian, it is possible to exploit its eigenfunctions as basis set. It is possible to show that all the quantum mechanical operators that represent physical observables (such as the Hamiltonian one, the momentum and the angular momentum ones, and so on...) are Hermitian.
In general it is thus possible to express whatever wave-function as a linear combination of suitable functions, that constitute the so called basis set. If the basis functions are indicated with $\{u_m(\vec{r})\}_m$ then the wave-function can be written as:

$$\psi(\vec{r}) = \sum_m c_m u_m(\vec{r}) \quad , \quad c_m = (u_m(\vec{r}), \psi(\vec{r})) = \int u_m^*(\vec{r})\psi(\vec{r})d\vec{r} \tag{17.75}$$

where a time-independent wave-function was considered for simplicity (the general results do not change - see previously eq (17.49) and the discussion above), and the integral is generally a three dimensional integral in space ($d\vec{r} = dxdydz$).
From the geometrical meaning pointed out previously it is possible to represent the wave-function ψ simply by means of a column vector consisting of the expansion coefficients (exactly like is often done in linear algebra):

$$\psi(\vec{r}) \quad \rightarrow \quad \{c_1 \; c_2 \; ... \; c_n \; ...\}^T$$

where T indicates the transpose. Notice that in general the Fourier series involves an infinite number of basis functions $\{u_m(\vec{r})\}_m$, nevertheless also a finite number of basis functions can be used. The latter case is always the case to be used in a numerical implementations on computers. In the first case the dimension of the coefficients vector will be infinite, in the latter finite. A vector space with a infinite number of dimensions is called Hilbert space (that is indeed a generalization of the concept of vector space for infinite dimensions), thus in the first case the basis set is a basis (thus a coordinate system) for a representation in an Hilbert space of the wave-function. In the latter case if the (finite) dimension of the basis set is called M then the representation becomes:

$$\psi(\vec{r}) = \sum_{m=1}^{M} c_m u_m(\vec{r}) \quad thus: \quad \psi(\vec{r}) \quad \rightarrow \quad \{c_1 \; c_2 \; ... \; c_M\}^T \tag{17.76}$$

Notice that this last case is not so different from the presented finite difference discretization of the steady state Schrödinger's equation in the previous section 17.4.1. In that case indeed it was:

$$\psi(\vec{r}) \quad \rightarrow \quad \{\psi(\vec{r}_1) \; \psi(\vec{r}_2) \; ... \; \psi(\vec{r}_M)\}^T$$

An important point now is that the basis set $\{u_m(\vec{r})\}_m$ can be chosen quite arbitrary, and if a good choice is made, i.e. the basis functions $\{u_m(\vec{r})\}_m$ are similar to the wave-functions that appear in the considered problem, then it is possible to accurately represent the wave-function $\psi(\vec{r})$ with just few terms (to the extreme if $u_1 = \psi$ then only one term is necessary, with $c_1 = 1$ - even if it has no sense).
By substituting the wave-function expansion into the steady state Schrödinger's equation one gets:

$$\widehat{H}\psi = E\psi \quad , \quad with: \quad \psi(\vec{r}) = \sum_m c_m u_m(\vec{r})$$

$$\rightarrow \quad \sum_m c_m \widehat{H} u_m(\vec{r}) = E \sum_m c_m u_m(\vec{r})$$

since the Hamiltonian operator has to be applied to a function u_m cannot be taken out on the left, instead since $\widehat{H}$ is linear and c_m are constant in $\vec{r}$ (they are the result of an

integral in $d\vec{r}$ thus the space dependency is lost) they can be taken out on the left. Then by multiplying both sides by $u_n^*(\vec{r})$ and integrating over all $\vec{r}$ yields to:

$$\sum_m c_m \int \widehat{H} u_m(\vec{r}) u_n^*(\vec{r}) d\vec{r} = E \sum_m c_m \int u_m(\vec{r}) u_n^*(\vec{r}) d\vec{r}$$

From which:

$$\sum_m H_{nm} c_m = E \sum_m S_{nm} c_m \tag{17.77}$$

where:

$$H_{nm} = \int u_n^*(\vec{r}) \widehat{H} u_m(\vec{r}) d\vec{r} \tag{17.78}$$

$$S_{nm} = \int u_n^*(\vec{r}) u_m(\vec{r}) d\vec{r} \tag{17.79}$$

Equation (17.77) can be rewritten in matrix form as:

$$\left[H\right] \left\{c\right\} = E \left[S\right] \left\{c\right\} \tag{17.80}$$

where square brackets indicate a matrix and the curly ones a column vector (E is a scalar).

Example 2.4: To definitely get the point let's consider this example in which only two basis functions u_1 and u_2 are present. The summation over m is to be considered with n fixed (indeed eq. (17.77) was derived by multiplying the steady state Schrödinger's equation by a fixed function $u_n^*(\vec{r})$), thus the left hand side term is:

$$\sum_{m=1}^{2} H_{nm} c_m = H_{n1} c_1 + H_{n2} c_2 = \begin{bmatrix} H_{n1} & H_{n2} \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \end{Bmatrix}$$

where the product is a matrix product (rows by columns). This must hold for each n, that corresponds exactly to the matrix product between the matrix of elements H_{nm} (n is the row index and m the column one) and the column vector of the coefficients, that is:

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \end{Bmatrix} = \begin{Bmatrix} H_{11} c_1 + H_{12} c_2 \\ H_{21} c_1 + H_{22} c_2 \end{Bmatrix}$$

Indeed eq. (17.77) was written by multiplying the steady state Schrödinger's equation by $u_n^*(\vec{r})$, and thus in principle a system of equations of the kind of eq. (17.77) should be written for each n possible value. These equations are obtained by multiplying the steady state Schrödinger's equation by all the possible $u_n^*(\vec{r})$, one at time. And a system of equations of that kind can be though in matrix form, in which each row of the matrix corresponds to an equation of the kind of eq. (17.77) with a different n value, exactly as it happens in this simple example.
This argument should have convinced of the validity of eq. (17.80). The final column vector (last equation right hand side) consist of the coefficient column vector of the wave-function that is the result of the application of the Hamiltonian operator to the state ψ. □

Evaluating the integrals of eq. (17.78) and (17.79) is the most time-consuming step in this process. Nevertheless once they are known the conventional linear algebra can be used (much simpler than differential equations), and an almost direct numerical implementation is moreover possible.
Notice that $\left[S\right]$ is called overlap matrix, and in the case of an orthonormal basis set it coincides with the identity matrix $\left[I\right]$. Indeed in that case the basis functions are supposed normalized such that the integral of eq. (17.79) is always null except when $m = n$, that corresponds to diagonal elements, for which it is normalized to 1. In the case of an orthonormal basis set the matrix form of the Schrödinger's equation becomes:

$$\left[H\right] \left\{c\right\} = E \left[I\right] \left\{c\right\} \quad \rightarrow \quad \left[H\right] \left\{c\right\} = E \left\{c\right\} \tag{17.81}$$

indeed the matrix product between $E\left[I\right]$ and $\left\{c\right\}$ and the scalar product between E and $\left\{c\right\}$ are the same. Notice that now, with the matrix notation just introduced, the steady state Schrödinger's definitively resembles a standard eigenvalue problem, see especially the orthonormal basis set case of equation (17.81).
In general it is possible to have a complete set, that thus can be used as basis set, even if it is not orthonormal but simply a set of linear independent functions, from which the need of defining properly the overlap matrix. This point will be further addressed in the next section 17.4.3.
Supposing of exploiting the linear algebra instruments to solve the eigenvalue problem, one gets the energy eigenvalues E_i and the corresponding eigenvectors, or better their representation in the chosen basis. Thus the problem can be to turn back to real space and get $\psi(\vec{r})$, starting from the knowledge on the coefficient eigenvectors $\left\{c\right\}_i$. The solution is immediate when eq. (17.75) is considered. Indeed:

$$\psi_i(\vec{r}) = \frac{1}{\sqrt{Z_i}} \sum_m c_{mi} u_m(\vec{r}) \quad , \quad \psi_i^*(\vec{r}) = \frac{1}{\sqrt{Z_i}} \sum_n c_{ni}^* u_n^*(\vec{r})$$

where Z_i is a normalization constant for the $i\text{-}th$ eigenfunction, that is chosen in order to satisfy the normalization condition eq. (17.19):

$$\int \psi_i^*(\vec{r})\psi_i(\vec{r})d\vec{r} = \frac{1}{Z_i} \int \sum_n \sum_m c_{ni}^* u_n^*(\vec{r}) c_{mi} u_m(\vec{r}) d\vec{r} = 1 \quad \rightarrow \quad Z_i = \sum_n \sum_m c_{ni}^* c_{mi} S_{nm}$$

A set of very simple, clear and useful examples for becoming familiar with this notation and related topics are reported in [44].

17.4.3 The Dirac's notation and Hilbert spaces

In this section the same topics of the previous one will be discussed, but with the gain of the Dirac's notation. Moreover the mathematical formalism behind the addressed topics will be also briefly discussed. In this section I will mainly follow the approach of [210], and also of [44].
As already mentioned in the previous section, eq. (17.75) corresponds a representation of the wave-functions $\psi(\vec{r})$ in terms of the basis functions $u_m(\vec{r})$ and this is not so different from the representation in a given basis of a vector in a given vector space. In this optics the wave-function $\psi(\vec{r})$ can be seen as a state vector in the infinite dimensional function space called the Hilbert space. If the number of basis functions is finite (see eq. (17.76)) then the wave-functions $\psi(\vec{r})$ corresponds to a state vector in the M-dimensional Hilbert space. The coefficients c_m in equations (17.75) and (17.76) are like components of the state vector $\psi(\vec{r})$ along the basis functions $u_m(\vec{r})$, i.e. the coordinate system chosen as basis. Choosing a different basis set is exactly like choosing a different coordinate system, and the component c_m along the different axes change while the state vector $\psi(\vec{r})$ is the same. In the Dirac's notation the state vector associated with a given wave-function $\psi(\vec{r})$ is indicated with the so called "ket": $|\psi\rangle$, while the basis functions $u_m(\vec{r})$ are indicated with the kets: $|m\rangle$ (sometimes also with: $|u_m\rangle$). In this notation equation (17.75) becomes:

$$|\psi\rangle = \sum_m c_m |m\rangle \tag{17.82}$$

Before going on let's recall the definition of vector space. A linear vector space V is a set of elements called vectors, in which a sum operation and a product by a scalar operation are defined. The sum operation has commutative and associative properties, and associates to each pair of vectors in V a third vector again in V (V is closed w.r.t. sum) in a unique way. The scalar product has distributive and property w.r.t. vectors and distributive and associative properties w.r.t. scalars, and it has for result a vector that is again element in V (V is closed w.r.t. scalar product). Notice that the scalar can be in general a complex

number.
The concept of Hilbert space is a generalization of the concept of vector space for the case in which the basis vectors are indeed basis functions. Its elements are often indicated with the "ket" symbol introduced above. Analogously to conventional vector spaces, it is possible to define the scalar product between two functions in the Hilbert space. This concept was already introduced by means of eq. (17.32), and it is now rewritten pointing out the Dirac's notation:

$$\langle f|g\rangle = \int f^*(\vec{r})g(\vec{r})d\vec{r} \tag{17.83}$$

where f and g are any two functions. The similarity of this integral with a scalar product is seen very well considering that the integral has meaning of sum over a continuum $d\vec{r}$. Thus it sums all the products f^*g, that is effectively what a standard scalar product does: it sums the component by component products (if one works with complex numbers the complex conjugates of the first vector components are used, i.e. f^*). Notice that the complex conjugate of the function f is indicated with the so called "bra" $\langle f|$. The scalar product is thus represented by the juxtaposition of $\langle f|$ and $|g\rangle$, that is: $\langle f|g\rangle$.
In order to represent functions in an Hilbert space W a basis set of functions is needed. Analogously to conventional vector spaces, a basis is constituted by a linearly independent set of generators. With the term "generators" it is intended that each function of the Hilbert can be expressed as linear combination of the basis functions. Thus said $\{u_m\}_m$ the basis set for the space W:

$$|f\rangle = \sum_m c_m |m\rangle \quad , \quad \forall f \in W$$

while with linearly independent it is meant:

$$\sum_m c_m |m\rangle = 0 \quad \Leftrightarrow \quad c_m = 0 \ , \ \forall\, m$$

In general it is thus possible to have non-orthogonal basis set, from which the importance of the overlap matrix $\big[S\big]$ of elements $S_{nm} = \int u_n^*(\vec{r})u_m(\vec{r})d\vec{r}$; as already mentioned. The basis functions are orthogonal if their scalar products are null except for the case $n = m$. If they are also normalized to 1 then they said to be orthonormal. For an orthonormal basis set the following relation is satisfied:

$$\langle n|m\rangle = \int u_n^*(\vec{r})u_m(\vec{r})d\vec{r} = \delta_{nm} \quad , \quad \delta_{nm} = \begin{cases} 0 & if\ n \neq m \\ 1 & if\ n = m \end{cases}$$

where δ_{nm} is the Kronecker's delta. If the basis set is non-orthogonal (see eq. (17.79) for the definition of S_{nm}):

$$\langle n|m\rangle = \int u_n^*(\vec{r})u_m(\vec{r})d\vec{r} = S_{nm} \tag{17.84}$$

Again notice that for orthonormal basis sets the overlap matrix becomes the identity one: $\big[S\big] = \big[I\big]$ (which is a diagonal matrix with ones on its diagonal).

Operators and matrix representation

From linear algebra it is known that a linear application (e.g. $\widehat{H}$) can be represented in matrix form, thus it is intuitive that is possible to write a differential (i.e. linear) operator in a matrix form. Moreover it was already shown the procedure to turn the differential operator $\widehat{H}$ into a matrix representation $\big[H\big]$ with the discussion about the steady state Schrödinger's equation. To this purpose see equations (17.77) and (17.78); where the latter is a definition for the elements H_{nm} of the matrix $\big[H\big]$. Equation (17.78) is now rewritten exploiting the Dirac's notation:

$$H_{nm} = \int u_n^*(\vec{r}) \left(\widehat{H}u_m(\vec{r})\right) d\vec{r} = \langle n|\widehat{H}|m\rangle \tag{17.85}$$

where $\widehat{H}\,|m\rangle$ indicates the application of the Hamiltonian differential operator to the basis function u_m (i.e. the state $|m\rangle$), and the juxtaposition of the bra $\langle n|$ with $\widehat{H}\,|m\rangle$ indicates the scalar product. This holds in general for any operator $\widehat{F}$, that in matrix form becomes $\big[F\big]$ with elements F_{nm}:

$$F_{nm} = \int u_n^*(\vec{r}) \left(\widehat{F} u_m(\vec{r})\right) d\vec{r} = \langle n|\widehat{F}|m\rangle \tag{17.86}$$

Example 2.5: In order to prove the last statement let's start from the consideration that in general a differential operator $\widehat{F}$ acting on a state vector $|\psi_1\rangle$ changes it into a different state vector $|\psi_2\rangle$. Usually only transformations such that $|\psi_1\rangle$, $|\psi_2\rangle \in W$ (with W considered Hilbert space), are of interest. From the above discussion it is known that (see eq. (17.75) where Dirac's notation is now used):

$$|\psi_1\rangle = \sum_m c_m\,|m\rangle \quad , \quad c_m = \langle m|\psi_1\rangle = \int u_m^*(\vec{r})\psi_1(\vec{r})d\vec{r}$$

As said above, the application of $\widehat{F}$ on the state $|\psi_1\rangle$ produces a different state vector $|\psi_2\rangle$:

$$|\psi_2\rangle = \widehat{F}\,|\psi_1\rangle$$

The point now is how to get the matrix representation for $\widehat{F}$ such that the above relation can be written in matrix form as:

$$|\psi_2\rangle = \big[F\big]\,|\psi_1\rangle$$

In order to do that let's start from the first relation:

$$|\psi_2\rangle = \widehat{F}\,|\psi_1\rangle = \widehat{F}\sum_m c_m\,|m\rangle = \sum_m c_m\widehat{F}\,|m\rangle$$

then it is possible to represent the new state vector $\widehat{F}\,|m\rangle$, that represents the differential operator $\widehat{F}$ acting on the basis state $|m\rangle$, as a linear combination of the basis functions like:

$$\widehat{F}\,|m\rangle = \sum_n c_n\,|n\rangle \quad , \quad c_n = \langle n|\widehat{F}|m\rangle = \int u_n^*(\vec{r}) \left(\widehat{F} u_m(\vec{r})\right) d\vec{r} = F_{nm}$$

and thus, by substituting in the previous relation:

$$|\psi_2\rangle = \sum_m c_m\widehat{F}\,|m\rangle = \sum_m c_m \sum_n c_n\,|n\rangle = \sum_m\sum_n c_m\,\langle n|\widehat{F}|m\rangle\,|n\rangle = \sum_m\sum_n c_m F_{nm}\,|n\rangle$$

from which an explicit expression for the coefficient vector that represent the state vector ψ_2 in the basis set is recovered:

$$|\psi_2\rangle = \sum_n \left(\sum_m c_m F_{nm}\right)|n\rangle = \sum_n K_n\,|n\rangle \quad , \quad K_n = \sum_m F_{nm}c_m$$

Indeed the last relation corresponds to a representation of the vector state ψ_2 in the basis set $|n\rangle$, given by the set of basis functions $\{u_n(\vec{r})\}_n$ (now the index is called n but in principle it can be called m as done previously for ψ_1), by means of the coefficients K_n, that all together constitute the coefficient vector of ψ_2 in the basis $\{u_n\}_n$. The coefficient vector definition $K_n = \sum_m F_{nm}c_m$ can be interpreted in matrix form as (remember that the matrix product is row by column):

$$\big\{K\big\} = \big[F\big]\,\big\{c\big\} \quad \leftrightarrow \quad \begin{Bmatrix} K_1 \\ K_2 \\ ... \\ K_n \\ ... \end{Bmatrix} = \begin{bmatrix} F_{11} & F_{12} & ... \\ F_{21} & F_{22} & ... \\ ... & ... & ... \\ ... & ... & ... \\ ... & ... & ... \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \\ ... \\ c_m \\ ... \end{Bmatrix}$$

Where $\left[F\right]$ is exactly the matrix of elements F_{nm} as defined above. In conclusion it follows that (again the summation over n is interpreted as matrix product - i.e. the scalar product of the coefficient column vector with the basis set vector):

$$|\psi_2\rangle = \sum_n K_n |n\rangle = \begin{Bmatrix} K_1 \\ K_2 \\ ... \\ K_n \\ ... \end{Bmatrix} \{u_1 \quad u_2 \quad ... \quad u_n \quad ...\} =$$

$$= \begin{bmatrix} F_{11} & F_{12} & ... \\ F_{21} & F_{22} & ... \\ ... & ... & ... \\ ... & ... & ... \\ ... & ... & ... \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \\ ... \\ c_n \\ ... \end{Bmatrix} \{u_1 \quad u_2 \quad ... \quad u_n \quad ...\} = \left[F\right] |\psi_1\rangle$$

where indeed:

$$|\psi_1\rangle = \sum_m c_m |m\rangle = \begin{Bmatrix} c_1 \\ c_2 \\ ... \\ c_n \\ ... \end{Bmatrix} \{u_1 \quad u_2 \quad ... \quad u_n \quad ...\} = \{c\} \{u\}^T$$

In the last equation the summation over m is interpreted as a scalar product between the coefficient column vector and the vector that has the basis functions for components ($u_m \rightarrow |m\rangle$); then it is interpreted as matrix product (rows by columns). □

The overlap matrix elements S_{nm} were already written with the new notation in equation (17.84). The steady state Schrödinger's equation in matrix form was already written in equation (17.80). Notice again that if the basis set is orthonormal then $\left[S\right] = \left[I\right]$ and the Schrödinger's equation becomes the one of eq. (17.81). From the discussion carried on in the example 2.5 above, it follows that is also possible to write the steady state Schrödinger's equation directly like:

$$\left[H\right] |\psi\rangle = E |\psi\rangle$$

Indeed:

$$\left[H\right] \sum_m c_m |m\rangle = E \sum_m c_m |m\rangle \quad \rightarrow \quad \left[H\right] \{c\} \{u\}^T = E \{c\} \{u\}^T$$

That corresponds exactly to eq. (17.81) with a multiplication (to the right) of both members by $\{u\}^T$ (notice that previously $\{u\}$ was assumed to be a row vector).
Notice that the time-dependent Schrödinger's equation with the Dirac's notation becomes:

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = \widehat{H} |\psi\rangle$$

where the Hamiltonian operator can be intended in its differential or matrix form.

Hermitian operators and some properties

Good references for all the properties of linear operators can be [210] (in Italian) and [212] (a practical introduction is also provided in [44]). In the following of this work it will not needed to handle operators in matrix operation (the only concepts needed are the ones already introduced), consequently all the properties of operators in Hilbert spaces are not introduced.
For example it is possible to perform changes of basis (in a similar manner to what is done in

the linear algebra courses), to define the inverse operator (quite similarly to inverse matrix definition), to define operations on operators such as their product, the commutation and anti-commutation operations and so on...
In this section I will only mention a pair of properties that will be important in the next chapters of this work.

An operator is said unitary if its inverse F^{-1} equals its complex conjugate transpose $F^{\dagger}$:

$$F^{-1} = F^{\dagger} \rightarrow F^{\dagger}F = 1$$

where "1" is the identity operator: $1|m\rangle = |m\rangle$.

A linear operator $\widehat{F}$ is said to be Hermitian when its matrix representation equals its complex conjugate and transpose:

$$\big[F\big] = \big[F\big]^{\dagger} \quad , \; i.e. \;\; F_{ij} = F^{*}_{ji}$$

It is possible to show that all the operators representing real physical observables are Hermitian and that an Hermitian operator has real eigenvalues. This has sense if one thinks to the physical meaning of eigenvalues: they are the only possible results of a measurement on the physical observable F represented by $\widehat{F}$. Moreover it is possible to show that a matrix that is Hermitian in a given basis remains Hermitian also in any other representation. Another useful property of Hermitian operators is that their eigenfunctions are orthogonal (after normalization orthonormal) and thus linear independent and they can be used as basis set. Notice that a representation of an Hermitian operator (e.g. $\widehat{H}$) in the basis of its eigenfunctions is diagonal (thus in energy domain $\big[H\big]$ is diagonal), and the diagonal elements are its eigenvalues (thus the energy eigenvalues in the case of $\big[H\big]$).

The continuum case

In this section extra-topics that are not strictly required for understanding the next of this thesis about the modeling of molecular devices are reported for completeness, since they were beforehand mentioned in some sections. They will be introduced not rigorously, but with a focus on the meaning.
So far it was implicitly assumed that the basis set $\{u_m\}_m$ constitutes a discrete set of functions, even if infinite. This is the case of the Fourier series theory. Nevertheless in equation (17.49) the symbol $\sum\!\!\!\!\!\!\int$ was used to indicate the general case of a mixed spectrum Hamiltonian operator, meaning that a continuum basis set can also be chosen. Now it is moment of definitively clarify this point by introducing the correct nomenclature.
It is well known that the maximum number of linearly independent vectors that can be extracted by a vector space V equals its dimensionality (e.g. the 3D space with coordinate axes x, y and z has dimensionality 3 - isomorphous to $\mathbb{R}^3$). The same concept holds also for an Hilbert space W. In particular the Hilbert spaces can have a finite dimensionality (think to eq. (17.76) in which only M basis functions are used to represent the general state ψ), or an infinite one (think to eq. (17.75) or (17.82) in which an infinite number of basis functions is used to represent the general state ψ). Moreover in the latter case the dimensionality can be infinite and countable (discrete case) or alternatively finite and uncountable (thus with at least a part that is continuous). To understand better, think that the cardinality (i.e. the number of elements of a set) of the integer numbers sets such as $\mathbb{N}$ or $\mathbb{Z}$ is infinite countable, while the one of the real numbers set $\mathbb{R}$ is infinite uncountable. For the Hilbert spaces the situation is similar. Hilbert spaces with finite dimensionality or with an infinite countable dimensionality are called "separable" Hilbert spaces. Instead Hilbert spaces with an infinite uncountable dimensionality are called "non-separable" Hilbert spaces.
To get the point, let's assume that the chosen basis set is the set of the eigenfunctions of the Hamiltonian operator (that is Hermitian and thus with a complete set of linearly

independent eigenfunctions). In the case of quantum confinement (bound states) $\widehat{H}$ has a discrete spectrum, i.e. the discrete eigenvalues E_i cannot assume each real value but only few discrete values. The corresponding eigenfunctions are thus a countable set of generators functions. More precisely if an infinite height quantum well is considered, the number of bound states is infinite, thus the eigenfunctions are an infinite countable set of functions. In this case a general state ψ can be represented in the Hamiltonian operator basis $\{u_m\}_m$ by means of eq. (17.75), or with the Dirac's notation eq. (17.82). Otherwise if a finite height quantum well is considered, the number of bound states is finite, thus the eigenfunctions are finite countable set of functions. In this case a general state ψ can be represented in the Hamiltonian operator basis $\{u_m\}_m$ by means of eq. (17.76). Nevertheless in this last case the finite height quantum well admits also a part of the Hamiltonian spectrum that is continuous: indeed for energies above the well height there is no constraint on the energy values, that can be continuous (real). Thus if both the bound and the unbound states of a finite quantum well are chose as basis for an Hilbert space representation, then its dimensionality will be infinite and uncountable, because of the continuum states (with energy eigenvalues that are an infinite range of real numbers).
In the continuum case of infinite uncountable dimensionality of the Hilbert space the previously relations must be reinterpreted as follows. The representation of the general state ψ as linear combination of the basis vector is no more a Fourier series expansion, but it becomes a Fourier anti-transform (think to the -not formal!- introduction of the Fourier transform as limit case of the Fourier series - see [217], [216]). Assuming that the chosen basis is the one composed by the Hamiltonian operator eigenfunctions (such that the notation is the one already used in section 17.3.5), and considering only the continuum part of $\widehat{H}$ spectrum:

$$\psi(\vec{r};t) = \int dE\ C(E)\ \Psi_E(\vec{r})\ e^{-\frac{i}{\hbar}Et} = \int dE\ C(E)\ u_E$$

where the basis eigenfunctions are simply called $u_E = \Psi_E(\vec{r})\ e^{-\frac{i}{\hbar}Et}$. The previous equation corresponds to a Fourier anti-transform. The coefficients $C(E)$ correspond instead to a Fourier transform w.r.t. to the basis u_E:

$$C(E) = (u_E, \psi(\vec{r};t)) = \int dE u_E^* \psi(\vec{r};t) = \int dE \Psi_E^*(\vec{r})\ e^{i\frac{i}{\hbar}Et}\psi(\vec{r};t)$$

Even in this case everything is analogous to what discussed in the previous sections. The main difference is in the normalization condition, that in the continuum case must involve the delta Dirac function instead of the Kronecker delta:

$$\langle n|m\rangle = \delta(n-m)$$

where $\delta(n-m)$ indicates a delta Dirac. A delta Dirac is indeed a generalization of the Kronecker delta for the case of continuum systems, and indeed it appears also in the normalization condition of unbounded wave-functions (see section 17.3.7).

17.5 Few Other useful topics in quantum mechanics

The purpose of this section is to briefly summarize other important results of quantum mechanics. These are "extra" topics in the sense that they are not necessary to understand the rest of this work. Nevertheless they are important achievements and in my opinion a brief mention is due. A full treatment is present e.g. in [210].

Correspondence principle

Bohr in 1923 stated the correspondence principle: a correct quantum theory must "contain" also the well known classical laws, in the sense that classical laws must be recovered from

quantum predictions each time they are applied to a macroscopic (classical) system.

Ehrenfest's theorem

Ehrenfest's theorem states that the time evolution of expected (average) values of physical observables, follow the same laws of the corresponding classical physics quantities.
The main consequence of this theorem is that in solid state physics, e.g. in conventional electronic devices, the classical physics can be used, if the physical quantities are intended to be average values. This is the main theoretical foundation for using classical physics in modeling conventional electronic devices, in which indeed average quantities (such as the mobility, the average velocity, etc...) are used. This allows to use the drift-diffusion set of equations for modeling electronic devices.
Notice that there is no a theoretical foundation for the widely used quantum corrections of such semi-classical models. Indeed there is no reason why a "bit" of quantum mechanics should be used in a semi-classical model (intended in the sense of Ehrenfest's theorem). Instead a purely quantum approach has all the reasons to be used in such devices, since indeed they are devices in which quantum mechanical effects are non-negligible and the classical approximation of physics (see above the Bohr's correspondence principle) does not hold. The non-equilibrium Green's function formalism, introduced later in this work, is a purely quantum mechanical approach for modeling transport in nano-electronic devices, thus it should be used.

Coefficient vector physical meaning

Notice that for normalized wave-functions expressed in an orthonormal basis, the following relation holds (it is called the Parseval's relation):

$$\begin{aligned}\langle\psi|\psi\rangle = \int \psi^*\psi d\vec{r} = \int |\psi|^2 d\vec{r} = 1 = \int \left(\sum_m c_m |m\rangle\right)^* \sum_m c_m |m\rangle \, d\vec{r} = \\ = \int \left(\sum_m c_m^* \langle m|\right) \sum_m c_m |m\rangle \, d\vec{r} = \sum_m c_m^* c_m \int \langle m|m\rangle \, d\vec{r} = \\ = \sum_m |c_m|^2 = 1 \quad (if\ normalized).\end{aligned}$$

The coefficients c_m are the projections of the wave-function ψ on the basis functions u_m:

$$c_m = \langle m|\psi_1\rangle = \int u_m^*(\vec{r})\psi_1(\vec{r})d\vec{r}$$

if they are normalized such that the above summation of their squared moduli is 1, then their moduli squared represent exactly the percentage of the basis function u_m that is present in ψ.
This concept is important in the eigenvalues problems, because it provides a direct physical interpretation of the eigenvalues. Let's consider the generic eigenvalue equation:

$$\widehat{F}\psi_n = f_n\psi_n$$

where ψ_n are the eigenfunctions and f_n the eigenvalues. It was already mentioned (see section 17.3.6) that because of the way this general equation is derived, the physical meaning is that the eigenvalues f_n are the only possible results of a measurement on the physical observable F represented by the operator $\widehat{F}$. If the system is in the state ψ_n then the result of the measurement will be f_n. Moreover since F is supposed to be a real physical observable for the system, then $\widehat{F}$ will be Hermitian, and thus the eigenfunctions ψ_n are linearly independent and orthogonal and can be used as basis set (see section 17.4.3). In this case, supposing also to have suitably normalized the ψ_n, one gets:

$$\psi = \sum_n c_n\psi_n \quad , \quad c_n = \langle\psi_n|\psi\rangle = \int \psi_n^*\psi d\vec{r} = (\psi_n, \psi)$$

At this point if the system is in the state ψ, that is a linear combination of the eigenstates ψ_n, then a measurement process will make the wave-function collapse into one of the eigenstates ψ_n and the corresponding obtained value will be the corresponding eigenvalue f_n. Let's now consider the average value of the observable F, that accordingly with eq. (17.31) is (the Dirac's notation is also pointed out here, together with the previously adopted one):

$$\langle F\rangle = \langle\psi|\widehat{F}|\psi\rangle = \int \psi^*\widehat{F}\psi d\vec{r} = \left(\psi, \widehat{F}\psi\right) = \left(\psi, \widehat{F}\sum_n c_n\psi_n\right) = \left(\psi, \sum_n c_n\widehat{F}\psi_n\right) =$$
$$= \left(\psi, \sum_n c_n f_n\psi_n\right) = \int \psi^* \sum_n c_n f_n\psi_n d\vec{r} = \sum_n c_n f_n \int \psi^*\psi_n d\vec{r} =$$
$$= \sum_n f_n c_n \left(\psi, \psi_n\right) = \sum_n f_n c_n c_n^* = \sum_n f_n |c_n|^2$$

where the following relations are used:

$$\widehat{F}\psi_n = f_n\psi_n \quad and \quad c_n^* = \left(\int \psi_n^*\psi d\vec{r}\right)^* = \int \psi_n^*\psi d\vec{r} = (\psi, \psi_n)$$

In conclusion:

$$\langle F\rangle = \sum_n f_n |c_n|^2$$

from which it is clear that the quantity $|c_n|^2$ assumes exactly the meaning of probability of finding the eigenvalue f_n when a measurement is performed on F (if the coefficient are normalized to 1 as indicated by the Parseval's relation reported above).
In the case of a continuum spectrum it is possible to show an analogous relation, that has an analogous meaning (where $c(f)$ is the Fourier transform):

$$\langle F\rangle = \int f|c(f)|^2 df$$

Simultaneous measurements of physical observables

It was highlighted many times that in the moment in which a measurement is performed on a system, because of the measurement process, the system state changes from the generic state ψ to one of its possible eigenstates ψ_n (collapse of the wave-function) such that the result of the measurement is one of the possible eigenvalues f_n. The only case in which the system does not change state, when a measurement is performed on it, is when it is already (before the measurement) in one of its eigenstates ψ_n such that the result of the measurement is f_n. This is an intrinsic property of quantum systems.
Now a question may be when it is possible to perform simultaneous measurements on two different physical quantities F and G. The answer is that a simultaneous measurement of two distinct physical quantities F and G is possible if and only if the two operators $\widehat{F}$ and $\widehat{G}$ admit common eigenstates. Indeed if the measurement process cause the wave-function collapse, the state of the system must collapse in an eigenstate of both $\widehat{F}$ and $\widehat{G}$ such that it is possible to get a result for the measurement (that will be f_n and g_n). It is possible to provide a rigorous demonstration for this statement, and to represent it by means of mathematical relations. Notice that this is also the origin of the Heisenberg's uncertainty principle. Indeed as mention in section 17.2 it is possible to provide a demonstration of the uncertainty principle (in spite of its name). The formal demonstration of the Heisenberg's indetermination principle leads to the following inequality:

$$\Delta r_i \Delta p_i \geq \frac{\hbar}{2}$$

where r_i and p_i are the *i-th* component of the position vector $\vec{r}$ and of the momentum vector $\vec{p}$ respectively (with $i = x, y, z$). In the 1D case:

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

Notice that Δf indicates the standard deviation of the quantity f, as defined in eq. (17.26). The uncertainty principle holds for each pair of canonically conjugate variables (linked by a time derivative) f and g:

$$\Delta f \Delta g \geq \frac{\hbar}{2}$$

And moreover it is possible to demonstrate the following energy-time uncertainty relation (that holds formally with the physical meaning introduced in section 17.2):

$$\Delta E \Delta t \geq \frac{\hbar}{2}$$

Bohr's complementary principle

In classical physics, for a full description of a physical system with n degrees of freedom (e.g. in 3D, $n = 3$) it is necessary the simultaneous knowledge of n pairs of canonically conjugated quantities (like the position $\vec{r} = (r_x, r_y, r_z)$ and the momentum $\vec{p} = (p_x, p_y, p_z)$ such that the concept of trajectory is defined). Nevertheless because of the Heisenberg's uncertainty principle this is not possible in quantum mechanics. It follows that the full characterization of a physical system is determined by a complete set of simultaneously measurable quantities. The maximum number of simultaneously measurable quantities in a system with n degrees of freedom is n (think again to the Heisenberg's uncertainty principle). This is essentially the statement of the Bohr's complementary principle, that holds true for all the physical quantities that have a classical analogous. It can be stated as: the complete characterization of a quantum system with n degrees of freedom is possible by means of the choice of a complete set of at most n compatible physical quantities (i.e. for which the simultaneous measurement is possible). This characterization is not unique, but more than one complete set can be chosen, and the different choices correspond to equivalent descriptions of the same system, that are thus complementary.
Notice that when quantization occurs a given physical quantity can be expressed in function of an integer number (think for example to the energy in a quantum well), thus choosing n quantities corresponds in choosing n "quantum numbers" (i.e. integers) by means of which the state of the system can be characterized. Indeed the wave-function solution of the Schrödinger's equation will depend on n integer numbers, that are related to different physical quantities (e.g. the energy, the angular momentum, and so on...). This is the origin of the quantum numbers introduced in the basic courses of chemistry.
Finally notice that in a quantum system, beside to classical degrees of freedom for which hold the previous discussion, there can be also purely quantum quantities (such as the spin) that have no classical analogous. The n quantities and the Bohr's complementary principle are referred to physical quantities that have classical analogous only. The knowledge on such purely quantum quantities can be considered separately and joined to the others. For example in the 3D space there are $n = 3$ degrees of freedom. From the Bohr's complementary principle only $n = 3$ physical quantities, that have classical analogous, can be used in quantum mechanics to provide its characterization. The quantum systems description in the 3D space is usually provided in terms of the quantum numbers n, l, m and s. The (principal) quantum number n is associated to the energy. As discussed in section 17.3.6 this is the most important physical quantity that provides information also about the history of the system. The angular momentum quantum number l is associated to the magnitude of the angular momentum (that is quantized and thus characterized by means of the integer l). The magnetic quantum number m is associated to the third component of the angular momentum (i.e. to angular momentum orientation, that is quantized and thus characterized by m). These three quantities have classical analogous and thus they provide a complete set (their simultaneous measurement is possible accordingly with Heisenberg's uncertainty principle). In addition the quantum number of spin s that provide information concerning the spin can be also used in the system description. Indeed it is associated to a

quantity (the spin) that has no classical analogous. Other choices are in principle possible, in that case a complementary and equivalent descriptions of the system would be provided.

Systems of identical particles

In quantum mechanics holds the so called "principle of indistinguishability". In classical mechanics it is always possible to distinguish between two particles that are identical. Indeed it is possible to distinguish them thanks to different space localization or thanks to the concept of trajectory. Then, by means of the knowledge on their trajectories it is possible to understand always which of the two particles is. Instead in quantum mechanics this is no more true, because of the Heisenberg's uncertainty principle that makes no more possible talking of trajectories. For example, if two hydrogen atoms are considered, it is possible to distinguish between the two electrons in the two atoms only when they are at large distance, because of the spatial localization of such electrons. If instead the two hydrogen atoms are close (think e.g. to the hydrogen molecule H_2) then the two wave-functions are overlapped and there is no way of understand which of the two electrons was initially belonging to what atom.
In a system of N identical particles (it can be also a molecule with N electrons) if two particles are exchanged a new wave-function is in principle obtained. For example in a molecule with N electrons the steady state Schrödinger's equation can be solved for a given configuration, and the eigenstate ψ_α is obtained. Then, for example, two identical particles are exchanged, that could be: an electron 1 that is in a given quantum state described by a given set of quantum numbers $\{n_1, l_1, m_1, s_1\}$ is exchanged with an electron 2 described by $\{n_2, l_2, m_2, s_2\}$; and thus after the exchange the electron 1 is in the state described by $\{n_2, l_2, m_2, s_2\}$ and the electron 2 is in the state described by $\{n_1, l_1, m_1, s_1\}$. Then, the steady state Schrödinger's equation can be solved again, and the wave-function ψ_β is obtained. Both ψ_α and ψ_β are solution of the steady state Schrödinger's equation for that system, and it turns out that the eigenvalues E_α and E_β associated to the two wave-functions are equal: $E_\alpha = E_\beta$, that means that ψ_α and ψ_β are degenerate (exchange degeneration). Thus in principle in a system of N identical particles there is a degeneracy equal to $N!$ (indeed the same procedure can be iterated for each possible permutation of the states - i.e. each possible exchange of electron states). Nevertheless, because of the indistinguishability principle, since it is not possible to distinguish between the state ψ_α and the state ψ_β (since the two exchanged electrons are indistinguishable), it means that the two states must be one the multiple of the other, such that they differ only for a multiplicative constant K (parallel vectors in an Hilbert space) and they are essentially the same state (see also superposition principle - section 17.3.2). Thus: $\psi_\beta = K\psi_\alpha$. If the two electrons are again exchanged the same reasoning leads to: $\psi_\beta = K\psi_\alpha = K^2\psi_\beta$. From which must be $K^2 = 1$ and thus only two cases are possible: $K = \pm 1$. The first case $K = +1$ leads to symmetric wave-functions, the second one $K = -1$ to anti-symmetric wave-functions. This means that: $\psi_\beta = \pm\psi_\alpha$ and instead of getting an exchange degeneracy of $N!$ it is obtained an exchange degeneracy equal to 2. Moreover it is an experimental evidence that for bosons (i.e. integer spin particles - e.g. photons) the only possibility is the symmetric one, with $K = +1$, while for fermions (i.e. fractionary spin particles - e.g. electrons) the only possibility is the anti-symmetric one, with $K = -1$. It means that if two (identical) bosons are exchanged of state then the same wave-function is obtained, while if two (identical) fermions are exchanged of state the anti-symmetric wave-function is obtained.
This is important because it is intimately linked with the Pauli exclusion principle. As known from basic chemistry courses, it states that in a given quantum system it is not possible to have two fermions (e.g. electrons) in the same quantum state (i.e. with all the quantum numbers -n, l, m, s- that are equal). Indeed it is possible to show (see e.g. [210]) that this would imply that their wave-function must be null (i.e. no state). And this

follows directly from the requirement of having an anti-symmetric wave-function when an exchange of state is performed. Instead notice that it is possible to have two bosons in the same quantum state.

Causality principle and time evolution operator

Also in quantum mechanics the causality principle must hold. In particular it states that once the initial state of the system is known, the physical laws that govern its time evolution must allow to predict the system state in each time instant. This is possible thanks to the direct application of the Scrhödinger's equation. Indeed it was alredy mentioned that the (time-dependent) Schrödinger's equation is the fundamental equation in quantum mechanics that allows to know the dynamics of the system, i.e. its time evolution (see sections 17.3 and 17.3.5). Such time evolution is then determined by the Hamiltonian operator of the system, that indeed appears in the Schrödinger's equation. Starting from the time-dependent Schrödinger's equation (17.17) it is possible to recover an expression for a quantum mechanical operator that is able to provide the time evolution of the state of the system. Such an operator is called "time evolution operator" and it is defined such that:

$$\psi(\vec{r};t) = \widehat{S}(t,t_0)\psi(\vec{r};t_0) \tag{17.87}$$

The time evolution operator $\widehat{S}(t,t_0)$ is thus the operator that, when applied to an initial state $\psi(\vec{r};t_0)$ supposed known at the time instant t_0, is able to provide the state of the quantum system at the instant t (the final time instant). This operator is implicitly defined in the Schrödinger's equation. It is possible to show that [210]:

$$\widehat{S}(t,t_0) = e^{-\frac{i}{\hbar}\widehat{H}(t-t_0)} \tag{17.88}$$

Its main properties are:

$$\left[\widehat{S}(t,t_0)\right]^{-1} = \left[\widehat{S}(t,t_0)\right]^{\dagger} \qquad i.e.\ it\ is\ unitary.$$

$$and: \quad \widehat{S}(t_0,t_0) = 1$$

Different pictures in quantum mechanics

In section 17.4 it was derived a matrix representation for quantum mechanical operators. This representation is at the basis of the so called Heisenberg's matrix mechanics formalism (even if it was developed also thanks to many other scientists among which Paul Dirac). The purpose here is to briefly highlight the differences between the Schrödinger's and the Heisenberg's approaches.

In the Schrödinger's formulation (also called wave mechanics) the focus is on the wave-function of a quantum system, and the time evolution (the dynamic evolution) of the system is expressed in terms of the wave-function itself. The state of a system is represented by means of the wave-function, and it evolves in time following the law of the Schrödinger's equation, that as mentioned in the previous section provides the time evolution of the state of the system. The mathematical formalism is the one of differential equations and indeed operators are differential (linear) operators. Notice that here the states depend on time and indeed they evolve in time, while the operators are fixed in time and they do not change.

Vice versa in the Heisenberg's approach the fundamental role is played directly by the observable quantities (the physical observables), that are represented by means of operators in matrix form. The matrix elements of the operators are quantum mechanical representations of the physical observables. In this approach the dynamics, i.e. the time evolution, is related to the time evolution of operators, while the states are unchanged in time. Thus,

as time goes by, the state does not evolve in time, but the system simply changes its state, under the action of time-dependent operators that make the system passing through different states. The time evolution of the operators is derived from the time evolution laws of the physical quantities they represent. The mathematical formalism is essentially the one of linear algebra, thus algebraic (matrix) equations are considered instead of differential ones.

The two formalisms were derived in different ways at almost the same time. Even if slightly different, they lead to same predictions and they have essentially the same content (as already noticed by Schrödinger in 1926). In the following a brief deeper discussion is provided in order clarify the link between the two approaches.

Assuming that exists a linear (Hilbert) space W, that includes all the possible states of all the possible physical systems, then it will be always possible to represent the generic state $\psi(\vec{r};t)$ as a vector $|\psi\rangle$ in this linear space. The Dirac's notation introduced in section 17.4.3 is used to indicate that vector. The vector $|\psi\rangle$ is thus the well known wave-function or state of a given physical system. The time evolution of a system can be evaluated in different ways in quantum mechanics. As already pointed out the two extremes are the Schrödinger's picture and the Heisenberg's one. The point now is how to switch between these two pictures.
In the Schrödinger's picture the states $|\psi\rangle$ are seen as elements in the space W that evolve or change with time: $|\psi\rangle = |\psi(t)\rangle \ \in W$. The vector $|\psi\rangle$ will thus change in time: $|\psi(t)\rangle$. Instead all the operators that do not involve an explicit time dependence (e.g. if they do not have a time derive like it happens with $\widehat{H}$), and they can be assumed to be unchanged as time goes by. The time evolution of the state will be given by the Schrödinger's equation:

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \widehat{H}\,|\psi(t)\rangle$$

or alternatively, as mentioned in the previous subsection, by the application of the time evolution operator:

$$|\psi(t)\rangle = \widehat{S}(t,t_0)\,|\psi(t_0)\rangle \quad , \quad \widehat{S}(t,t_0) = e^{-\frac{i}{\hbar}\widehat{H}(t-t_0)}$$

Instead in the the Heisenberg's picture the states $|\psi\rangle$ are time independent and thus unchanged in time, while operators are time dependent. The space W is seen as a static space, whose vectors $|\psi\rangle$ do not change in time. While the operators evolve in time, under the laws that will be presented now. The states (that do not evolve in time) and the operators (that evolve in time) in the Heisenberg's picture are defined as (subscript H is used to indicate that they are in the Heisenberg's picture):

$$|\psi_H(t)\rangle = \widehat{S}\,|\psi_H(t_0)\rangle = e^{-\frac{i}{\hbar}\widehat{H}(t-t_0)}\,|\psi_H(t_0)\rangle = |\psi_H(t_0)\rangle \quad \textit{constant in time.}$$

$$\widehat{F}_H = \widehat{F}_H(t) = \widehat{S^\dagger}\widehat{F}_S\widehat{S} = e^{+\frac{i}{\hbar}\widehat{H}(t-t_0)}\widehat{F}_S e^{-\frac{i}{\hbar}\widehat{H}(t-t_0)} \quad \textit{changes in time.}$$

where $\widehat{F}_S$ indicates that the operator $\widehat{F}$ is represented like done so far in this work (section 17.4), i.e. it is the operator in the Schrödinger's picture.
The last two relations, and especially the one that defines the operators $\widehat{F}_H$ in the Heisenberg's picture, constitute the link between the two picture.

Example 2.6: Let's verify the correspondence between the two pictures, and thus that the two pictures are equivalent. In the Schrödinger's picture the operators $\widehat{F}_S$ are time independent. Nevertheless they matrix elements F_{mn} (in their matrix representation) are time dependent: $F_{mn}(t)$. This because the matrix elements are scalar products between states, and the states depend on time: $u_m(t)$. Indeed:

$$F_{mn}(t_0) = \langle u_m(t_0)|\widehat{F}_S|u_n(t_0)\rangle \quad \neq \quad F_{mn}(t) = \langle u_m(t)|\widehat{F}_S|u_n(t)\rangle$$

In particular since the time evolution of states in the Schrödinger's picture can be evaluated by applying the time evolution operator:

$$|u_m(t)\rangle = \widehat{S}\,|u_m(t_0)\rangle = e^{-\frac{i}{\hbar}\widehat{H}(t-t_0)}\,|u_m(t_0)\rangle$$

Then the matrix element at time instant t is:

$$F_{mn}(t) = \langle u_m(t)|\widehat{F}_S|u_n(t)\rangle = \left(u_m(t), \widehat{F}_S u_n(t)\right) = \left(\widehat{S}u_m(t_0), \widehat{F}_S\widehat{S}u_n(t_0)\right)$$
$$= \int \left(\widehat{S}u_m(t_0)\right)^* \widehat{F}_S\left(\widehat{S}u_n(t_0)\right) d\vec{r} = \int \widehat{S}^\dagger u_m^*(t_0)\widehat{F}_S\widehat{S}u_n(t_0)d\vec{r} =$$
$$= \int \widehat{S}^\dagger\left(u_m(t_0)\right)^* \widehat{F}_S\widehat{S}\left(u_n(t_0)\right) d\vec{r}$$

Since the operator $\widehat{F}_S\widehat{S}$ is acting on $u_n(t_0)$, that in Dirac notation is the ket $|u_n(t_0)\rangle$, and the operator $\widehat{S}^\dagger = e^{+\frac{i}{\hbar}\widehat{H}(t-t_0)}$ is acting on $u_m^*(t_0)$ that in Dirac notation is the bra $\langle u_m(t_0)|$, and noticing that the application of an operator on a bra is indicated with: $\langle u_m|\,\widehat{F}$, that in this case implies $\widehat{S}^\dagger\left(u_m(t_0)\right)^* = \langle u_m(t_0)|\,\widehat{S}^\dagger$, the last relation can be rewritten as:

$$F_{mn}(t) = \langle u_m(t)|\widehat{F}_S|u_n(t)\rangle = \int \widehat{S}^\dagger u_m^*(t_0)\widehat{F}_S\widehat{S}u_n(t_0)d\vec{r} = \langle u_m(t_0)|\widehat{S}^\dagger\widehat{F}_S\widehat{S}|u_n(t_0)\rangle$$

Since in the Heisenberg's picture the states are time independent: $|u_{mH}(t)\rangle = |u_{mH}(t_0)\rangle$ (where subscript H indicates that this holds in the Heisenberg's picture), the correspondence between Schrödinger's picture operators and Heisenberg's picture operators is recovered:

$$F_{mn}(t) = \langle u_m(t)|\widehat{F}_S|u_n(t)\rangle = \langle u_m(t_0)|\widehat{S}^\dagger\widehat{F}_S\widehat{S}|u_n(t_0)\rangle = \langle u_{mH}(t)|\widehat{S}^\dagger\widehat{F}_S\widehat{S}|u_{nH}(t)\rangle =$$

$$= \langle u_{mH}(t)|\widehat{F}_H|u_{nH}(t)\rangle \qquad , \qquad \widehat{F}_H = \widehat{S}^\dagger\widehat{F}_S\widehat{S}$$

that correspond to what stated previously. Notice that the matrix elements $F_{mn}(t)$ in the two pictures are the same, indeed they are intimately linked with the expected values of physical observables, and thus they must be unchanged (since directly linked to measurements). Moreover notice that the states in the Schrödinger's picture in this example were simply indicated with u_m with no subscript. □

In addition to these two pictures a third one is also widely used in quantum mechanics. It is the so called "interaction picture". In this representation both the states and the operators are time dependent. It is particularly useful if the Hamiltonian operator (that appears also in the time evolution operator) can be written as the sum of an interacting part that accounts for the interactions among particles and a non-interacting one (whose eigenvalues and eigenfunctions can be calculated exactly). It is possible to show that the time evolution of operators and states in the interaction picture depends only on the non-interacting Hamiltonian, and thus it is simple to evaluate. This picture is often used in the second quantization formalism (see next section), that constitutes the background in which the non-equilibrium Green's function formalism finds place in its traditional formal derivation. In this work I will follow the approach of [44] and [91] in which such a formalism is introduced without the need of the second quantization formalism, and thus is made accessible also to electronic engineers in a fast way without too much effort.

A brief mention about the second quantization

So far the presented quantum mechanical theory was the so called "first quantization" one. As seen, in this theory the concept of wave-particle duality and wave field allow to associated to each microscopic quantity both undulatory and particle behaviors. An important feature is that the forces acting on systems are still considered classical forces. Indeed

the potential energy by means of which it is represented the interactions of the quantum systems under study and the rest of the world, can be still derived in a classical manner (e.g. it can be an electrostatic Coulomb potential, etc...). The force field is thus said to be non-quantized. The quantized quantities are the results of the double behavior of matter particles, to which a wave field is associated. This theory succeeded in explaining a wide range of phenomena, and in particular all the non-relativistic quantum phenomena (from which it is sometimes referred as non-relativistic quantum mechanics).
Nevertheless this theory cannot still definitively explain some issues, especially related to the interaction between the electromagnetic radiation and matter. In order to get a quantum theory able to precisely describe and predict also such phenomena the so called "second quantization" and the so called "quantum field theory" were introduced.
As mentioned, in the first quantization the force fields are still classical, while the physical observables are associated to operators. Instead in the second quantization the force fields, and thus all the interactions between particles, are quantized. In order to do that the force fields are associated to the so called "field operators", in a similar way of what done in the first quantization with the physical observables. These field operators are able to create or annihilate a particle at a given time instant and in a given point, because of the action of the force field. The field operators are defined in an Hilbert space that is the so called the Fock's space. It was mentioned previously (see the subsection on systems of identical particles) that by exchanging two identical particles in a quantum system a new state is obtained, whose wave-function must be symmetric (bosons) or anti-symmetric (fermions) w.r.t. to the initial one. Usually these properties must be enforced on the wave-functions to get proper results. The Fock's idea at the basis of his proposed linear space, is to "embed" these symmetry properties directly in the way in which the states are represented in a given Hilbert space: namely the Fock's space. Therefore in Fock's space the mathematical operations of creation and annihilation are suitably defined, such that the Pauli exclusion principle and the correct symmetry on wave-functions is always guaranteed automatically whenever a particle is added (created) to the system or eliminated (annihilated) from it. This is the starting point for the so called relativistic quantum mechanics, that is the quantum field theory. In quantum field theory the so called canonical quantization (i.e. second quantization or quantization of force fields) is used to build up the field operators in the Fock's space. Quantum field theory (QFT) is a relativistic theory that combines classical field theory, special relativity and quantum mechanics (but not general relativity). It is a theory that is able to definitely explain the interaction between electromagnetic radiation and matter (this specific branch of quantum field theory is called quantum electrodynamics), and many other phenomena. All the interactions (all kind of interactions, e.g. electromagnetic, gravitational, nuclear etc..) are quantized in QFT. For example the electromagnetic field is rethought in terms of photons (i.e. quanta of electromagnetic radiation), the gravitational field in terms of gravitons, and so on...
Notice that the existence of the spin comes out naturally in QFT by solving the relativistic Dirac's equation (the fundamental equation of QFT - for importance analogous to the Schrödinger's one in non-relativistic quantum mechanics). Instead in non-relativistic quantum mechanics (the one considered in this chapter and in this entire work) it is introduced from experimental considerations (usually presenting the Stern-Gerlach famous experiments) and with the phenomenological Pauli's correction of the Schrödinger's equation.
The second quantization formalism, and especially the Fock's space and field operators, are of extreme importance also in the so called "quantum many-body" systems analysis and "quantum many-body perturbation theory". These two branches of quantum mechanics focus on systems constituted by many body, like it could also a molecule for example. Indeed a molecule with tens of atoms can also contains hundreds of electrons, thus making clear that it is a many-body system. The situation is even worse in solid state physics in which even a nano-scale crystalline device can include hundreds or thousands of atoms.

CHAPTER 18

Molecular electronic structure

The first step in the study of the transport in whatever kind of device (also conventional 3D electronic devices) is to understand what are the possible electronic states within the conducting channel. Indeed the electronic states that will take part to the conduction (essential to understand the number of electrons involved in conduction and thus the electrical current) will be a subset of the allowed electronic states within the channel. In molecular electronic sensors, and in general in molecular electronic devices, the active part of the device is a single molecule (or at most a small packet of parallel molecules), and the electronic states that are permitted for a given molecular channel can be found by solving the steady state Schrödinger's equation $\widehat{H}\psi = E\psi$. They corresponds indeed to stable states within the molecule, that are stationary states. Consequently the first step in studying the conduction in a molecular device is to derive and handle the molecular Hamiltonian operator $\widehat{H}$. Even if molecules are small, they are already enough complicated, and often with enough atoms composing them, that this task can be very hard and challenging. This chapter focuses on how to deal with and to solve the steady state Schrödinger's equation for molecules, starting from the molecular Hamiltonian operator. Various methods for achieving the task of calculating the electronic structure of a molecular channel are briefly presented and reviewed. The purpose is to provide the needed information to be used in setting up atomistic simulations like the ones presented and discussed in the part II of this work. A good reference on these topics (also for beginners) can be for example [219] - many other references are present in literature.
In section 18.1 the Born-Oppenheimer approximation is addressed. It is the starting point for all the methods discussed in the rest of this chapter and it will be always assumed. Then in section 18.2 the general mean field method and its implications are discussed. This is the fundamental approximation around which all the other presented methods revolve. In appendix **??** it is further discussed with simple examples and application formulae to give the physical insights and sensibility to understand the approximations and hypotheses also used in the other methods. In section 18.3 the Hartree and Hartree-Fock methods are addressed. They assume the mean field approximation introduced in section 18.2. Next in section 18.4 the most used classes of methods for the evaluation of the electronic structure are presented. They all can be classified and understood starting from the Hartree-Fock method. Moreover the force field methods and the post-Hartree-Fock methods are briefly treated. Finally in section 18.5 and 18.6 two of the most widely spread methods are discussed, namely the Density Functional Theory (DFT) and the Extended Hückel Theory (EHT). They are also the only two methods employed in the practical part (part II) of this work. The chapter ends with a discussion of the intermolecular interactions and the methods for simulating them within the DFT environment in section 18.7.

18.1 Molecular Hamiltonian and the Born - Oppenheimer approximation

In section 17.3.4 it was reviewed that the Hamiltonian operator is the quantum mechanical operator associated to the total energy of the system, and in general it is given by the expression of equation (17.44), reported here for clarity:

$$\widehat{H} = \widehat{T} + \widehat{U} = -\frac{\hbar^2}{2m}\Delta + U(\vec{r})$$

A molecule is a set of atom nuclei around which electrons are displaced, some of which (usually the so called valence electrons) are involved in chemical bonds between the atoms forming the molecule. Consequently the explicit expression of a molecular Hamiltonian corresponds to a many-body problem. Indeed even the simple benzene ring (C_6H_6) consists of six carbon atoms (with atomic number $Z_C = 6$) and six hydrogen atoms (atomic number $Z_H = 1$) meaning that there are 42 electrons. The number of electrons within a molecule increase rapidly with the complexity of the molecule itself, reaching e.g. the number of 360 electrons for the C_{60} fullerene molecule. It is intuitive thinking that some approximations must be used to deal with the molecular Hamiltonian and the solution of the Schrödinger's equation for a molecular channel.
It is possible to write the general expression of a molecular Hamiltonian considering that the kinetic energy operator can be written as the sum of the kinetic energy operators of the N_e electrons and the N_n nuclei (Δ is linear, $\widehat{T}$ is a linear operator and superposition of effects can be exploited):

$$\widehat{T} = \widehat{T}_e + \widehat{T}_n = -\sum_{i=1}^{N_e}\frac{\hbar^2}{2m_e}\Delta_{r_i} - \sum_{j=1}^{N_n}\frac{\hbar^2}{2m_n}\Delta_{R_j}$$

where $\Delta_{r_i} = \left(\frac{\partial^2}{\partial x_i^2}, \frac{\partial^2}{\partial y_i^2}, \frac{\partial^2}{\partial z_i^2}\right)$, $\Delta_{R_j} = \left(\frac{\partial^2}{\partial x_j^2}, \frac{\partial^2}{\partial y_j^2}, \frac{\partial^2}{\partial z_j^2}\right)$ and where the coordinate vectors of electrons and nuclei respectively $\vec{r_i} = (x_i, y_i, z_i)$ and $\vec{R_j} = (x_j, y_j, z_j)$.
Instead the potential energy operator (equal to the potential energy in real space) is constituted by three terms, that represent the three principal interactions among electrons and atomic nuclei:

- U_{nn}: contribution due to Coulombic nuclei-nuclei repulsion
- U_{ne}: contribution due to Coulombic electrons-nuclei attraction
- U_{ee}: contribution due to Coulombic electron-electron repulsion

Thus the general expression for the molecular Hamiltonian operator assumes the following form:

$$\widehat{H} = -\sum_{i=1}^{N_e}\frac{\hbar^2}{2m_e}\Delta_{r_i} - \sum_{j=1}^{N_n}\frac{\hbar^2}{2m_N}\Delta_{R_j} + U_{nn} + U_{ne} + U_{ee}$$

that becomes rapidly unmanageable as the number of atoms in the molecule increases. The steady state Schrödinger's equation becomes very hard to be solved also with the help of today's computers, if no approximations are considered.
In the so called Born-Oppenheimer approximation [220] the kinetic and potential energy contributions of the nuclei are neglected. Indeed, since the nuclei are several tens of thousands times heavier than electrons, their velocities are a much less than those of electrons, and thus the electrons can respond quickly to variations in the nuclei configuration, maintaining the system state essentially unaltered. Consequently the nuclei are approximated as fixed in the reference coordinate system, that means that their kinetic energy (and kinetic energy operator $\widehat{T}_n$) is null. Moreover also the potential energy contribution U_{nn} related to nuclei-nuclei repulsion is neglected: indeed since they are fixed, they are supposed to be

unaffected by repulsive forces among them, that are neglected. Therefore within the Born-Oppenheimer approximation, for a system with N_e electrons and N_n nuclei, the electronic Hamiltonian operator becomes:

$$\begin{aligned}\widehat{H} &= \widehat{T}_e + U_{ne}\left(\vec{r_i}, \vec{R_j}\right) + U_{ee}\left(\vec{r_i}, \vec{r_k}\right) = \\ &= -\sum_{i=1}^{N_e} \frac{\hbar^2}{2m_e}\Delta_{r_i} - \sum_{j=1}^{N_n}\sum_{i=1}^{N_e} \frac{Z_j q^2}{4\pi\varepsilon_0 \left|\vec{r_i} - \vec{R_j}\right|} + \sum_{1\leq\ i,k\ \leq N_e} \frac{q^2}{4\pi\varepsilon_0\ \left|\vec{r_i} - \vec{r_k}\right|}\end{aligned} \tag{18.1}$$

where q is the elementary charge, $-q$ is the electron charge, ε_0 is the vacuum permittivity, Z_j is the *j-th* nucleus atomic number (equal to the number of protons of charge $+q$ in the *j-th* atom). The previous expression is often called "many-body Hamiltonian", and it is often expressed in atomic units as [221]:

$$\widehat{H} = -\frac{1}{2}\sum_{i=1}^{N_e} \nabla_{r_i}^2 - \sum_{j=1}^{N_n}\sum_{i=1}^{N_e} \frac{Z_j}{|\vec{r_i} - \vec{R_j}|} + \sum_{1\leq i,k\leq N_e} \frac{1}{|\vec{r_i} - \vec{r_k}|} \tag{18.2}$$

Within this many-body framework the steady state many-body Schrödinger's equation assumes the following expression:

$$\widehat{H}\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_{Ne}) = E\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_{N_e}) \tag{18.3}$$

where $\psi(r_1, r_2, ..., r_{Ne})$ is the wave-function of the many-body system of N_e electrons composing the molecule, i.e. it is a steady state for the entire molecule (in other words a molecular orbital). As mentioned at the beginning of this section its exact solution is an impossible task for systems having more than a few electrons, even within the Born-Oppenheimer approximation. Indeed as N_e increases, the number of degrees of freedom in the many-body Schrödinger's equation, after it is discretized (e.g. with a finite difference method or another numerical method), increases exponentially making the problem computationally infeasible even on the most powerful computers available nowadays. Just to give an idea of the order of magnitude of the problem, if r_i is discretized with an $k \times k \times k$ grid (in the Fourier domain, remembering that $\vec{r} \leftrightarrow \vec{k}$), the dimension N_H of H is k^{3N_e}, that means that for example with $k = 32$ and $N_e = 5$, N_H is greater than 3.5×10^{22} [222]. Consequently several methods have been devised to find accurate approximated solutions. The choice of the approximated method to use, is a matter of suitability with the system under analysis and of numerical accuracy, and it will further discussed in this and successive chapters of the present work.

A final remark on the term "electronic structure" used in the introduction of this chapter: accordingly with [223] the electronic structure is the state of motion of electrons in an electrostatic field created by stationary nuclei. This corresponds in finding the eigenfunctions $\{\psi\}_i$ and the eigenvalues $\{E\}_i$ for a given structure. Notice that the eigenfunction moduli squared correspond to surfaces in the 3D space, that are high position probability surfaces, i.e. molecular orbitals in our case (in the coordinate space). Instead the eigenvalues $\{E\}_i$ are the molecular energy levels corresponding to each orbital. Therefore the electronic structure is obtained by solving the steady state Schrödinger's equation, usually under the aforementioned Born-Oppenheimer approximation, i.e. in the clamped-nuclei case. Notice finally that the electron states that will participate to conduction, will be part of the permitted (steady) states of the molecule, from which the importance of knowing the electronic structure of molecular channels.

18.2 The SCF procedure

The electron-electron interaction, i.e. the term U_{ee} in eq. (18.1), is likely the main issue in solving the many-body Schrödinger's equation. In this section the Self Consistent Field (SCF) procedure for approximating the electron-electron interaction is addressed. I will follow the introduction of [44]. The treatment here has no pretension of being a formal detailed introduction to these topics, but more a conceptual introduction to understand the basic reasoning behind these topics, even to electrical and electronic engineers who have never seen these topics before. The purpose is indeed to provide the means to correctly set up practical and useful simulations, in commercial or free available softwares without getting lost. In order to do that not all the theoretical details are necessary, but at least the fundamental concepts.

The basic idea is to transform the many-body problem of equation (18.3) into a N_e single-particle (one-body) problems, that are decoupled. Thus equation (18.1) is transformed in a system of N_e decoupled equations. In order to do that, the electron-electron potential U_{ee} is substituted with a suitable SCF potential $U_{SCF}(\vec{r})$ that arises from all but one electrons. A single electron is thus considered at a time, and the $U_{SCF}(\vec{r})$ potential to which it undergoes, arises from the remaining $N_e - 1$ electrons. This because an electron does not "feel" any potential due to itself [44]. In order to calculate the potential $U_{SCF}(\vec{r})$ a self consistent iterative procedure must be implemented. This because the considered (single) electron undergoes the sum of the attractive potentials of atomic nuclei, that are fixed under the Born-Oppenheimer approximation, and the aforementioned $U_{SCF}(\vec{r})$ (representing the repulsive interactions with all the other electrons), that is unfortunately a function of the final (many-body) wave-function. In practice $U_{SCF}(\vec{r})$ is derived from the electronic charge (think to a conventional electrostatic potential derived from a continuous charge density). The electronic charge in the 3D space depends on the wave-function (since this gives the high probability regions), and in turns the wave-function is a solution for the steady state Schrödinger' equation that contains $U_{SCF}(\vec{r})$. The self-consistent procedure can be summarized as follows [44]:

1. initial guess for the SCF potential $U_{SCF}(\vec{r})$
2. solve steady state Schrödinger's equation and get the eigenvalues $\{E\}_i$ and the eigenfunctions $\{\psi\}_i$
3. calculate the electron density $n(\vec{r})$
4. calculate the $U_{SCF}(\vec{r})$ from $n(\vec{r})$
5. if the new $U_{SCF}(\vec{r})$ is significantly different (i.e. more than the chosen maximum tolerance) from the last guess (or initial guess at first iteration) then: suitably update $U_{SCF}(\vec{r})$ and go back to step 2; if the new $U_{SCF}(\vec{r})$ is enough close to the last guess (i.e. less than the chosen maximum tolerance) the result is considered converged and the calculation is complete

The electron density $n(\vec{r})$ in step 3 can be calculated by summing up the probability distributions for all the occupied eigenstates:

$$n(\vec{r}) = \sum_{occupied\ i} |\psi_i(\vec{r})|^2 \tag{18.4}$$

The charge density is then given by the electron density multiplied the electron charge: $-q \cdot n(\vec{r})$. The SCF potential energy in step 4 is recovered from classical electrostatics by integrating the charge density $-q \cdot n(\vec{r})$ over the entire volume V of interest:

$$U_{SCF}(\vec{r}) = \frac{(N_e - 1)}{N_e} \frac{q^2}{4\pi\varepsilon_0} \int_V \frac{n(\vec{r'})}{|\vec{r} - \vec{r'}|} d\vec{r'} \tag{18.5}$$

where the multiplicative factor $(N_e - 1)/N_e$ finds the following explanation: the appropriate charge density for each eigenstate should exclude the electron eigenstate under consideration, since as mentioned no electron feels any repulsion due to itself. However it is more

convenient to simply take the total charge density and scale it by the factor $(N_e - 1)/N_e$ that "weights" it on all but one electrons. Notice that the total electronic charge in all the volume V is always equal to $-qN_e$. Further details on this procedure, and application examples can be found in [44].

Notice that the SCF approach allows to split the many-body problem into a system of N_e single-electron (one-body) problems:

$$\widehat{H}\psi = E\psi \quad \longrightarrow \quad \widehat{H}_i\left(n(\vec{r})\right)\ \psi_i = E_i\ \psi_i \quad , \quad i = 1, 2, ..., N_e \tag{18.6}$$

where the Hamiltonian operator $\widehat{H}_i$ of the single particle picture can be written from eq. (18.1) as follows (for the *i-th* electron of the system whose position vector is indicated simply with $\vec{r}$ - no confusion is possible since it is the only electron in the single-electron Schrödinger's equation):

$$\begin{aligned} \widehat{H}_i\left(n(\vec{r})\right) &= \widehat{T}_e + U_{ne}\left(\vec{r}, \vec{R}_j\right) + U_{SCF}\left(\vec{r}\right) = \\ &= -\frac{\hbar^2}{2m_e}\Delta - \sum_{j=1}^{N_n} \frac{Z_j q^2}{4\pi\varepsilon_0\ \left|\vec{r} - \vec{R}_j\right|}\ + \frac{(N_e - 1)}{N_e}\frac{q^2}{4\pi\varepsilon_0}\int_V \frac{n(\vec{r'})}{|\vec{r} - \vec{r'}|}d\vec{r'} \end{aligned} \tag{18.7}$$

notice that it is function of the electron density $n(\vec{r})$, given by equation (18.4).
The SCF approximation (also called the Hartree approximation or the mean field approximation) is further discussed in appendix **??**, where a more quantitative, even if oversimplified, treatment is provided.
The main problem of this approximation is related to the so called "electron correlation". Indeed if the many-body picture is considered, it is possible to show that the electron-electron interaction is less than the one predicted by the single-electron picture with the mean field just introduced. In particular it is possible to show that the real SCF potential that accurately accounts for all the other electrons in the single-electron picture is given by the difference between two contributions. The first one is an SCF potential of the kind of the one of eq. (18.5), that is called the Hartree potential (name that comes from the original Hartree method that was the first SCF-based method). The second contribution is a contribution that represent the aforementioned electron correlation. Indeed the actual interaction energy is less than the one of eq. (18.5) because electrons can correlate their motion so as to avoid each other. From the wave-function standpoint this corresponds to the fact that the probability of finding two electrons simultaneously in two points $\vec{r}_1$ and $\vec{r}_2$ of the space is not simply proportional to the electron density n in the two points: $n(\vec{r}_1)$ and $n(\vec{r}_2)$; but it is somehow reduced because electrons "try to avoid each other". In other words the presence of an electron in a region of space, makes decreasing the probability of finding another electron in that region w.r.t. the case in which the first electron is not present. This electron correlation comes out by doing all the calculations on the multi-electron picture. The simple semi-classical expression of equation (18.5) does not account for this correlation, and thus the actual SCF potential should be given by:

$$U_{SCF} = U_{Hartree} + U_{XC} \quad , \quad with: \quad U_{Hartree} = \frac{(N_e - 1)}{N_e}\frac{q^2}{4\pi\varepsilon_0}\int_V \frac{n(\vec{r'})}{|\vec{r} - \vec{r'}|}d\vec{r'}$$

where U_{XC} is a negative term accounting for the correlation, and it is called "exchange-correlation" potential. The accurate estimation of U_{XC} required great effort in research and it is still an unsolved problem for some specific branches of application in which high accuracy is required. Notice that the SCF approach is at the basis of many methods for the calculation of the electronic structure of molecules such as the DFT one that will presented in section 18.5. This topic is more quantitatively discussed in appendix **??**.
In conclusion to this section I would like to point out that, as also mentioned in [44], it is quite surprising that a single-electron picture with a suitable SCF can be often used as a reasonable accurate description of a multi-electron system. The fact that it works well has no convincing mathematical proof, but it is instead proved by many experiments in different conditions. Almost all the electronic structures of atoms, molecules and solids are somehow derived from this method, as will be discussed in section 18.4.

18.3 A brief mention on the Hartree-Fock method

The purpose of this section is to provide an insight in the Hartree and Hartree-Fock methods for the calculation of the electronic structure of molecules (or in general conductive channels or devices). These methods are important for understanding the taxonomy of the currently used classes of methods aimed in electronic structure calculations, that is indeed presented in the next section 18.4.

The Hartree method

Like in any SCF-based method the starting point for the Hartree method is the Born-Oppenheimer approximation. Moreover it is assumed that the electrons are non-interacting, and thus it is possible to show that the many-body steady state Schrödinger's equation admits factorized solutions of the kind:

$$\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_{Ne}) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2)...\phi_{Ne}(\vec{r}_{Ne})$$

Thus in the original Hartree method the total many-body wave-function $\psi(\vec{r})$ is initially approximated as the product (called Hartree product) of the single-electron wave-functions for the individual electrons of the system. Again notice that it holds only if each wave-function of each electron is independent from the others (non-interacting electrons). Under these hypotheses the many-body problem becomes a set N_e single electron problems, as already mentioned. If the *j-th* single electron problem is considered, i.e. the *j-th* electron is considered, then the SCF potential in which it is immersed, that accounts for the repulsive interactions of all the electrons but the *j-th* one can be recovered starting from the electron density given by all the other electrons (be careful to the notation):

$$n_j(\vec{r}) = \sum_{i \neq j} |\phi_i|^2$$

and by solving the following Poisson's equation:

$$\nabla^2 V_{Hj}(\vec{r}) = -q\frac{n_j(\vec{r})}{\varepsilon_0}$$

where the charge density $\rho_j(\vec{r}) = -qn_j(\vec{r})$ is now the charge density due to all the electrons except the considered one; it is supposed of having free space (thus ε_0) around the nuclei and the electrons (if not so ε of the medium should be considered), and $V_{Hj}(\vec{r})$ is the electrostatic potential (not potential energy) in the Hartree approximation (subscript "H") related to the *j-th* electron. The resulting SCF potential energy, that will appear in the *j-th* single-electron Schrödinger's equation, is thus:

$$U_{Hj}(\vec{r}) = -qV_{Hj}(\vec{r}) = -\frac{q}{4\pi\varepsilon_0}\int \frac{\rho_j(\vec{r})}{|\vec{r} - \vec{r'}|}d\vec{r'} = +\frac{q^2}{4\pi\varepsilon_0}\sum_{i \neq j}\int \frac{|\phi_i|^2}{|\vec{r} - \vec{r'}|}d\vec{r'} \tag{18.8}$$

This is the starting guess to be included in the Schrödinger's equation, that as mentioned previously must be solved self-consistently by means of an iterative procedure. The assumption that each electron feels the average potential $U_{Hj}(\vec{r})$ is the mean field approximation, also called Hartree approximation.

Notice that this expression is essentially analogous to the one of eq. (18.5), and to the aforementioned Hartree approximation of eq. (**??**). Indeed if the summation is extended to all the electron states, $n_j(\vec{r})$ becomes the total electron density $n(\vec{r})$ and the corrective weight $(N_e - 1)/N_e$ should be introduces, such that eq. (18.5) is recovered. This way of rewriting is possible thanks to the fact that the electrons are identical particles (see also section 17.5 - subsection about the systems of identical particles), and even if they exchanged the *j-th* electron is always subjected to the same SCF potential, that is:

$$U_{SCFj}(\vec{r_j}) = U_{SCF}(\vec{r}) = \frac{(N_e - 1)}{N_e}\frac{q^2}{4\pi\varepsilon_0}\int_V \frac{n(\vec{r'})}{|\vec{r} - \vec{r'}|}d\vec{r'}$$

In this way the many-body problem is divided into a system of N_e single electron problems, like done in eq. (18.6). These equations are in general non-linear and must be solved self-consistently.

The Hartree-Fock method

The original Hartree method briefly summarized above was conceived by Hartree in 1927, and was called self-consistent field method [219]. Nevertheless it did not account for the Pauli's exclusion principle, and for the symmetry constraints on the total system wave-function ψ due to the exchange of two identical particles (i.e. electrons on this case) of the system. As briefly discussed in section 17.5 (subsection on systems of identical particles), if two electrons are exchanged then the obtained wave-function for the total many-body system must be anti-symmetric. With the Hartree method this is not guaranteed and must be enforced *a posteriori*. Instead a way of directly embed this condition in the Hartree method is to consider the so called Hartree-Fock method [219].
In general it is possible to show that if the wave-function can be written as the product of the single-electron ones:

$$\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_{Ne}) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2)...\phi_{Ne}(\vec{r}_{Ne})$$

then if two electrons i and j are exchanged (i.e. the electron i that was in state ϕ_i is now in the state ϕ_j and vice versa) an anti-symmetric wave-function is obtained [210]:

$$\begin{aligned}\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_i..., \vec{r}_j..., \vec{r}_{Ne}) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2)...\phi_i(\vec{r}_i)...\phi_j(\vec{r}_j)...\phi_{Ne}(\vec{r}_{Ne}) = \\ = -\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_j..., \vec{r}_i..., \vec{r}_{Ne}) = -\phi_1(\vec{r}_1)\phi_2(\vec{r}_2)...\phi_j(\vec{r}_j)...\phi_i(\vec{r}_i)...\phi_{Ne}(\vec{r}_{Ne})\end{aligned}$$

Considering all the $N_e!$ possible permutations of the N_e single electron wave-functions ϕ, the previous relation can be written as [210]:

$$\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_i..., \vec{r}_j..., \vec{r}_{Ne}) = \frac{1}{\sqrt{N_e!}} \sum_P (-1)^P \ \widehat{P}[\phi_1(\vec{r}_1)\phi_2(\vec{r}_2)...\phi_{Ne}(\vec{r}_{Ne})]$$

where P indicates the *P-th* permutation and $\widehat{P}$ is the operator that performs that permutation. The last expression can be interpreted as a determinant, the so called "Slater determinant", defined as follows [210]:

$$\psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_i..., \vec{r}_j..., \vec{r}_{Ne}) = \frac{1}{\sqrt{N_e!}} \begin{vmatrix} \phi_1(\vec{r}_1) & \phi_1(\vec{r}_2) & ... & \phi_1(\vec{r}_{Ne}) \\ \phi_2(\vec{r}_1) & \phi_2(\vec{r}_2) & ... & \phi_2(\vec{r}_{Ne}) \\ ... & ... & ... & ... \\ \phi_{Ne}(\vec{r}_1) & \phi_{Ne}(\vec{r}_2) & ... & \phi_{Ne}(\vec{r}_{Ne}) \end{vmatrix}$$

The exchange of two electrons correspond to the exchange of the two corresponding columns in the Slater determinant.
In the Hartree-Fock method the initial wave-function, given by the product of the single-electron ones, is built starting from the Slater determinant, such that it automatically embeds the anti-symmetrization on the total wave-function. This is the main conceptual difference w.r.t. the Hartree method. Many variants are possible and many ameliorations were proposed during years. Notice that also the Hartree-Fock method includes a self-consistent solution [219].
Finally notice that in modern Hartree-Fock methods the initial single-electron wave-functions ϕ_i are already a linear combination of atomic orbitals, and generally these atomic orbitals can be Gaussian type wave-functions or of other types (e.g. Slater-type orbitals, more sharp around the nucleus than a Gaussian function). In practice many basis sets (exactly in the sense of the concept of basis set introduced in section 17.4) can be used [219].

18.4 Taxonomy of methods for electronic structure calculations

In this section the attempt is to classify the various methods for the calculation of the electronic structure of molecules in such a way that the correct terminology is associated to each method and, more important, the main features of a method can be immediately understood by the knowledge on where it find place in the methods taxonomy. More than one classification are possible, depending on what specific feature is desired to be highlighted.

First of all a fundamental distinction is the one between *ab initio* methods and semi-empirical ones. In *ab initio* methods the solution of the (many-body) steady state Schrödinger's equation is generated without reference to experimental data, and starting only from the first principles and physical constants. In contrast in semi-empirical models the solution is found by exploiting information from measurements or from *ab initio* methods. Moreover in semi-empirical methods, suitable fitting parameters are present, whose value is usually derived from a combination of theory (or *ab initio* calculations) and experimental observations.

Another possible classification is related to the kind of variables considered as the unknown of the Schrödinger's equation during the solution. So far these variables were assumed to be the wave-functions. Nevertheless in the mean field approximation the wave-functions are somehow derived from the electron density $n(\vec{r})$ or analogously from the electron charge density $\rho(\vec{r}) = -qn(\vec{r})$. Indeed this variable appears both in the SCF potential and in the Schrödinger's equation that, once solved, provides the eigenfunctions $\{\psi_i\}_i$. Thus since the wave-functions are intimately connected with the electron density, it is possible to re-formulate the eigenvalue problem such that the unknown to be found is precisely the electron density (or the charge density). In light of this, it is possible to classify the methods between wave-function-based methods and density-based methods. Nevertheless things are more elaborated than this and each of these classes is further subdivided into different approaches [224], [139]:

1. Wave-function-based methods: an explicit form for the wave-function is written down and physical observables are calculated using that.
   - [a.] Perturbational: Møller-Plesset, diagrammatic methods, etc...
   - [b.] Variational: Hartree-Fock, configuration interaction, etc...
2. Density-based methods: the focus is shifted from the wave-function to the electronic density. The wave-function is not written explicitly. Examples are Thomas-Fermi approximation and density-functional theory (DFT).

In order to further appreciate the aforementioned terminology some considerations must be pointed out, and they start from the already discussed Hartree-Fock (HF) method.

First of all notice that the Born-Oppenheimer approximation is always assumed to be true. Under this approximation the coupling between the nuclei and electronic motion is neglected, and this allows the electronic part to be solved with the nuclear positions as parameters. In general the Born-Oppenheimer approximation is good, and it does not afflict significantly the accuracy of the final solution [219], [139]. The dynamics of a many-electron system is very complex, and consequently requires elaborate computational methods. A significant simplification, both conceptually and computationally, can be obtained by introducing independent-particle models, where the motion of one electron is considered to be independent of the dynamics of all other electrons [219]. This coincides in what was done in the self-consistent mean field approximation and the Hartree-Fock method, that were already discussed in sections 18.2 and 18.3. Notice that an independent-particle model means that the interactions between the particles is approximated, either by neglecting all but the most important one, or by taking all interactions into account in an average fashion. Within electronic structure theory, only the latter has an acceptable accuracy,

and it is again exactly what was done with the SCF Hartree-Fock (HF) method. The key conceptual point in the understanding of the electronic structure methods is the HF theory. For this reason before going on in the treatment, its main features are summarized below (in the words of [219]):

a. In the HF model, each electron is described by an orbital, and the total wave-function is given as a product of orbitals.
b. Since electrons are indistinguishable fermions (particles with a spin of 1/2), the overall wave-function must be anti-symmetric (change sign upon interchanging any two electrons), which is conveniently achieved by arranging the orbitals in a Slater determinant.
c. The best set of orbitals is determined by the variational principle, i.e. the HF orbitals give the lowest energy within the restriction of the wave function being a single Slater determinant.
d. The shape of a given molecular orbital describes the probability of finding an electron, where the attraction to all the nuclei and the average repulsion to all the other electrons are included.
e. Since the other electrons are described by their respective orbitals, the HF equations depend on their own solutions, and must therefore be solved iteratively.
f. When the molecular orbitals are expanded in a basis set, the resulting equations can be written as a matrix eigenvalue problem.

The HF model is a kind of branching point, where either additional approximations can be invoked, leading to semi-empirical methods, or it can be improved by adding additional determinants, thereby generating models that can be made to converge towards the exact solution of the electronic Schrödinger's equation [219], [139].
Semi-empirical methods are derived from the HF model by further neglecting some integrals related to electron-electron interaction. Since the HF model by itself is only capable of limited accuracy, such approximations will by themselves lead to a poor model. Nevertheless a semi-empirical method can also provide very accurate results, depending on the specific case. Indeed the success of semi-empirical methods relies on turning the remaining integrals into parameters, and fitting these to experimental data, especially molecular energies and geometries. Such methods are computationally much more efficient than the *ab initio* methods, but they are limited to systems for which parameters exist. As mentioned, in such systems they can also reach significant good accuracy becoming competitors of *ab initio* methods, nevertheless the final result is strongly dependent on the specific considered case. A semi-empirical method is e.g. the Extended Hückel Theory (EHT), that will be addressed in section 18.6.
The HF theory only accounts for the average electron–electron interactions, and consequently neglects the correlation between electrons (see sections 18.2, 18.3 and appendix **??**). Methods that include electron correlation are typically computationally much more involved than the HF model, but can generate results that systematically approach the exact solution of the Schrödinger's equation. These methods are often called "electron correlation methods" or "post-Hartree-Fock methods". They will be mentioned again at the end of this section.
A different approach is instead the one of the Density Functional Theory (DFT), that in the original in the Kohn–Sham version can be considered as an improvement on HF theory, where the many-body effect of electron correlation is modelled by a function of the electron density. DFT is, analogously to HF, an independent-particle model, and is comparable to HF computationally, but provides significantly better results. The main disadvantage of DFT is that there is no systematic approach to improving the results towards the exact solution, like in the aforementioned electron correlation methods. The DFT method will be addressed in section 18.5.
Finally there exists a class of methods which is completely unrelated to all the others presented so far, and starts from a complete different approach. This includes the so called

"Force Field" methods, also referred as "Molecular Mechanics" methods. They are very computationally effective methods, but also very rough. A brief discussion is reported at the end of this section, while a complete presentation can be found e.g. in [219].

Electron correlation methods (the post-Hartree-Fock methods)

The main limit of the Hartree-Fock method (see section 18.3) is that the electron-electron interactions are completely neglected (this is indeed the hypothesis that allows for writing the many-body wave-function as the product of the single-electron ones). In general this is a limit of the SCF approach in which the electron-electron interaction is represented by means of a mean field. This is evident when the initial guess on the total many-body wave-function is considered. It is given by the product of single-electron ones in which the U_{ee} term is substituted my the mean field one U_{SCF}. The Hartree-Fock methods are for this reason called "independent-particle" models, indeed the motion of an electron is considered independent on the dynamics of all the others.
Successive generations of methods were derived from the original HF one, all aimed in better estimating the electron-electron interaction by means of the so called correlation potential (this holds true also for the DFT method - see section 18.5), to account for the electron-electron repulsion [139]. As mentioned in section 18.2 (and further discussed in appendix **??**) the point is that electrons can correlated their motion such that they avoid each other, and this correlation depends also on spin. As pointed out again in section 18.2 (and appendix **??**) the electron correlation results in an additional (negative) contribution of energy U_{XC} that must be added to the SCF potential within the Schrödinger's equation. In general the resulting HF many-body wave-function is able to account for the great majority of the total energy (let's say the 99%), but the small remaining part, that is due to the electron correlation, can be very important in describing chemical phenomena. Since the HF solution usually provides the $\sim 99\%$ of the actual correct answer, the electron correlation methods normally use the HF wave-function as a starting point for improvements. For this reason they are also called "post-Hartree-Fock" methods. The basic concept is to superimpose to the HF wave-function ψ_{HF} a set of other wave-functions ψ_i such that the total wave-function becomes [219], [139]:

$$\psi_{TOT} = \psi_{HF} + \sum_i c_i \psi_i$$

where c_i are suitable coefficients. Notice that this approach can be interpreted as a "multi-determinant" approach. Indeed in section 18.3 it was mentioned that the HF wave-function ψ_{HF} can be interpreted as the Slater determinant of a matrix whose elements are atomic wave-functions (or in general they can be a suitable basis set of functions). If the same is applied to each additional ψ_i then the ψ_{TOT} can be found by calculating several determinants. The electron correlation methods are essentially different in how they calculate the coefficients c_i. Indeed once the basis functions are selected they are always the same in each considered Slater determinant for building the ψ_i, while by changing c_i it is possible to get different total wave-functions. There are mainly three methods to calculate the electron correlation and thus the coefficients c_i: the configuration interaction method, the many-body perturbation theory and the coupled cluster.
Notice that the functions ψ_i are already total many-body functions, i.e. they can in turn be represented as superposition of the chosen basis functions (that in section 18.3 were indicated with ϕ_i). To this purpose see also section 17.4. Analogously it is possible to think to the total wave-function ψ_{TOT}, called multi-determinant wave-function, as a superposition of the ψ_i, that defines a sort of "coordinate system" for the Slater determinants. In this optics the basis set determines the size of the one-electron basis (and thus limits the description of the single ψ_i), while the number of included considered Slater determinants defines the size of the many-electron basis (and thus limits the description of electron correlation and of the final ψ_{TOT}).

At this point, in analogy with the Fourier series, notice that if the number of basis functions is increased then the accuracy generally increases, and for an infinite number of considered basis functions the reconstruction of the original function is obtained. This is the result of the already mentioned Fischer-Riesz theorem that states that (for a complete basis set) the original function can be exactly recovered as Fourier series expansion (unless a set of isolated points, that is empty if the initial function is continuous - like electron wave-functions must be, see section 17.3.2). Consequently if the number of basis set functions is increased, the representation (in that basis set) of the single-electron wave-function ψ_i is ameliorated. And again if the number of considered Slater determinants is increased, the representation of the total multi-electron wave-function ψ_{TOT} is ameliorated. This is more than an intuitive reasoning, indeed it is possible to show that the electron correlation methods can systematically converge to the actual solution of the Schrödinger's equation. This happens if both the basis set and the number of determinants is increased. In particular the exact Schrödinger's equation is recovered if a complete basis set and a complete set of Slater determinants are used. In real applications this is not possible since it would imply the usage of an infinite number of functions or determinants, nevertheless if the number is correctly increased then the solution is closer to the real one. This concept is summarized in figure 18.1. On abscissa axis the basis sets are considered (see also section 18.5 -subsection on basis set- for the basis set names and their meanings), increasing the accuracy means moving toward greater abscissa coordinates. On ordinate axis the different electron correlation methods are considered (increasing the accuracy means moving toward greater ordinate coordinates).

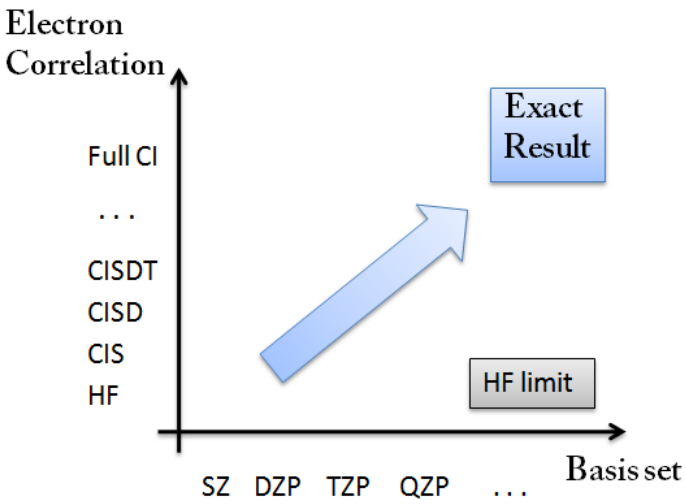


Figure 18.1: Schematic representation of the converge toward the exact solution of the electron correlation methods. On abscissa axis the basis sets are considered while on ordinate axis the different electron correlation methods are considered.

I will not use in the rest of this work any electron correlation method and thus I will not further discuss them. A good presentation is in my opinion provided in [219]. I conclude this section with a mention of some of these methods, just to give the idea of how ample is the current landscape (remember that three main approaches are possible for calculating the electron correlation):

- configuration interaction based correlation methods: the configuration interaction (CI) method, the multi-configuration self-consistent field, the multi-reference configuration interaction, etc...
- many-body perturbation theory based correlation methods: the Møller–Plesset (MP)

perturbation theory, unrestricted Møller–Plesset and projected Møller–Plesset, etc...
- coupled cluster based methods: truncated coupled cluster (CC) methods, etc...

Notice that also other methods and approaches are possible such as the so called Quantum Monte Carlo (QMC) method (that provides a statistical estimation of energy - intended as integral of the wave-function), etc... [219], [139].

Force field methods

This subsection has not the aim of being a complete introduction. It is essentially taken from [219], that instead provide a good introduction to the force field methods.
The "building blocks" in force field methods are atoms, i.e. electrons are not considered as individual particles. This means that bonding information must be provided explicitly, rather than being the result of solving the electronic Schrödinger's equation. In addition to bypassing the solution of the electronic Schrödinger's equation, the quantum aspects of the nuclear motion are also neglected. This means that the dynamics of the atoms is treated by classical mechanics, i.e. Newton's second law. For time-independent phenomena, the problem reduces to calculating the energy at a given geometry. Often the interest is in finding geometries of stable molecules or different conformations. The problem is then reduced to finding energy minima (in the principle that in nature a stable molecule is always in a configuration that minimizes its potential energy - ground state) on the potential energy surface. The electronic energy is thus written as a parametric function of the nuclear coordinates, and fitting the parameters to experimental or higher level computational data.
Moreover in force field methods the molecules are described by a sort of "ball and spring" model, with atoms having different sizes and "softness" and bonds having different lengths and "stiffness". The foundation of force field methods is the observation that molecules tend to be composed of units that are structurally similar in different molecules. For example carbons are essentially planar, or all C-H bond lengths are roughly constant in all molecules, being between 1.06 Å and 1.1 Å. If then the C-H bonds are further divided into groups, for example those attached to single-, double- or triple-bonded carbon, the variation within each of these groups becomes even smaller. This picture is not so different from the one widely used in organic chemistry, in which the organic molecules are seen as composed by the so called "functional groups". Force field methods are in a sense a generalization of these models, with the added feature that the atoms and bonds are not fixed at one size and length. As mentioned the main task in force field methods is almost always the one of minimizing the potential energy of the system, thus providing the stable configuration.
In a force field model the energy can be written as a sum of different contributions: the energy function for stretching a bond between two atoms, the energy required for bending an angle, the torsional energy for rotation around a bond, the non-bonded atom–atom interactions energy and the coupling energy between the first three terms. Given such an energy function of the nuclear coordinates, geometries and relative energies can be calculated by optimization, i.e. by minimizing it. Stable molecules correspond indeed to minima on the potential energy surface, and they can be located by minimizing the total potential energy as a function of the nuclear coordinates.

18.5 The DFT method

In the Density Functional Theory (DFT) the focus is shifted from the wave-function to the electronic density. The DFT method treats again the electron-electron interactions in a mean field manner (see section 18.2): each electron is considered moving inside a mean field created by all other electrons of the system under analysis. Hence, instead of solving an interacting N_e-particles problem, it solves N_e non-interacting one-particle problems self-consistently, that is pretty much more manageable. With this approach it is possible to

obtain the electronic structure of a system of a few tens to a few hundreds of atoms with reasonable accuracy and computational effort/time, if compared to traditional methods, such as Hartree-Fock theory and its descendants that include electron correlation.
The properties of a many-electron system, within the DFT framework, can be determined by using functionals, i.e. functions of another function, which in this case is the spatially dependent electron density $n(\vec{r})$. Before going on a note on semantics [219]: a function is a prescription for producing a number from a set of variables (coordinates); a functional is a prescription for producing a number from a function, which in turn depends on variables. A wave-function and the electron density are thus functions, while the energy depending on a wave-function or an electron density is a functional. Usually a function that depends on the variable x is indicated with $f(x)$, instead a functional F depending on a function $f(x)$ is conventionally indicated as $F[f]$ (with square brackets).
The basis for the DFT is the Hohenberg-Kohn's theorem that states that the ground state electronic energy is completely determined by the electron density. In other words, there exists a one-to-one correspondence between the electron density of a system and its energy. The importance of the Hohenberg–Kohn's theorem is well understood if a system with N_e electrons is considered. A steady state wave-function for such a system contains $4N_e$ variables (three spatial and one for spin - because of Pauli's exclusion principle the spin must be considered for discriminating among electron states). Instead the electron density $n(\vec{r})$ for the same system only depends on three spatial coordinates, independently on the number of electrons of the system. Therefore while the complexity of a wave-function increases exponentially with the number of electrons, the electron density has the same number of variables, independent of the system size. Moreover it is possible to demonstrate that each different electron density yields a different ground state energy, thus confirming the one-to-one correspondence. The problem is that the functional connecting the system energy with the electron density is not known. Indeed it is not possible to find an analytical explicit expression for it. Thus the goal of DFT methods is to design functionals that accurately connect the electron density with the energy [139].
Density functional theory is conceptually and computationally similar to Hartree–Fock theory [219], but provides much better results and has consequently become a very popular method. The main problem in DFT is the inability to systematically improve the results, like it happens instead for the electron correlation methods (see previous section 18.4). In addition, despite the many successes of DFT, there are some areas where the current functionals are known to perform poorly (especially concerning weak interacting systems):

- in describing certain types of weak interactions, e.g. van der Waals interactions, where instead the correlation methods have great success
- in describing systems with loosely bound electrons (the self-interaction error in these cases is larger than the actual binding energy, and thus lead erroneously to an unbound electron)
- in predicting the binding energy of some kinds of chemical bonds that are predicted to be too stable
- large errors in predicting excitation energies in some conditions
- poor result in some cases of strong electron-electron interaction (see also chapter 20)

For many of these issues some kind of correction exists and can be employed to improve the final accuracy. Few other critical issues are present in addition to the ones presented here, for reference see [219]. The only issue of interest in this work is the first one, namely the poor modeling of some kind of weak interactions such as the intermolecular van der Waals interactions, nevertheless corrections exist to improve the DFT accuracy and they are briefly presented in section 18.7.

18.5.1 Kohn-Sham Hamiltonian

Early attempts at designing DFT models (i.e. Thomas–Fermi and Thomas–Fermi–Dirac methods) tried to express all the energy components (the kinetic and all the potential contributions) as a functional of the electron density. In this optics it is possible to write directly an expression for the total energy of the system E_{tot} (energy eigenvalues), corresponding to the classical Hamiltonian (not the Hamiltonian operator) of the system (n is the electron density):

$$H_{tot} = E_{tot} = T_e[n] + U_{ne}[n] + U_{ee}[n]$$

Notice that this formulation embeds the exchange-correlation within the functional $U_{ee}[n]$ that represents the electron-electron interaction. Thus U_{XC} is not explicitly written. Unfortunately these methods had poor performance [219], mainly because it is very hard finding an enough accurate functional able to accurately express the kinetic energy term in function of the electron density. Although there have been some recent attempts at constructing such functionals, at the present this way is still unfeasible because of its poor accuracy [219]. Notice that this formulation of the many-body problem is completely orbital-free, indeed the wave-functions do not appear in this "pure" DFT approach. Notice also that if such functionals could be derived, the full potential of DFT in having only three variables independent of system size could be fully realized [139].
Nevertheless the current foundation for the use of DFT methods in computational chemistry is the introduction of some orbitals, as suggested by Kohn and Sham in 1965. These orbitals (i.e. wave-functions) are exploited to accurately represent the kinetic energy of the system. More precisely Kohn and Sham proposed that the electron kinetic energy should be calculated from an auxiliary set of orbitals, used for representing the electron density. As mentioned this assumption is the starting point of all the modern DFT methods. The gain in this approach is that it is possible to carry out very accurate calculations. Indeed the exchange–correlation energy U_{XC} (see section 18.2 and later in this section), which is a rather small fraction of the total energy, in this approach is the only unknown functional, and even relatively crude approximations for this term provide quite accurate computational models [219]. In the Kohn-Sham (KS) Hamiltonian the kinetic energy functional is split into two parts, one which can be calculated exactly, and a small correction term. Unfortunately price to be paid is that orbitals (i.e. wave-functions) are re-introduced and must be used to represent the kinetic energy term/operator which is not enough accurately represented in terms of electron density functional only. Thereby the complexity is increased from 3 (3 spatial variable of $n(\vec{r})$) to $3N_e$ variables, and the electron correlation re-emerges as a separate term, U_{XC} that must be summed up to the other potentials in the Hamiltonian. The KS model is closely related to the HF method, sharing identical formulas for the kinetic energy, electron–nuclear energy and Coulomb electron–electron energy (i.e. the Hartree potential of equations (18.5), (18.8) and (**??**)). In particular the many-body electronic Hamiltonian is re-conceived in terms of the one-electron Kohn-Sham Hamiltonian like [222]:

$$H_{KS}[n] = \sum_{i}^{N_e} \left[-\frac{\hbar^2}{2m_e}\Delta_i + U_{eff}(\vec{r_i}) \right] = \sum_{i}^{N_e} H_{KSi} \tag{18.9}$$

where the so called effective Kohn-Sham potential $U_{eff}(\vec{r_i})$ is defined as:

$$U_{eff}(\vec{r_i}) = U_H[n(\vec{r_i})] + U_{XC}[n] + U_{EXT}(\vec{r_i}) \tag{18.10}$$

In equation (18.9) the first term is the kinetic energy of the *i-th* electron, while the second term is the potential energy of the *i-th* electron moving in the mean field created by the other electrons as well as in any external potential field (e.g. the electrostatic potential of ions or any other external field). The latter is detailed in equation (18.10). The first term in eq. (18.10) is the Hartree potential due to the mean-field electrostatic interaction between the

electrons. It corresponds to the one of the already introduced equations (18.5), (18.8) and (**??**). The second one is the exchange-correlation potential, which arises from the quantum mechanical nature of the electrons and embeds all the many-particle interactions, thus it represents the correction term used to account for energy that the non-interacting reference system fails to capture. It corresponds exactly to the term introduced in section 18.2 (and in appendix **??**). Notice that these two terms are due to electron-electron interactions, which depend on the electron density $n(\vec{r})$, which depends on the wave-function ψ_i (that must be used and calculated for representing correctly the kinetic term), which in turn depend on U_{eff}. Therefore the problem of solving the Kohn-Sham equation has to be done in a self-consistent way, analogously to the SCF and HF methods. Usually in the KS method one starts with an initial guess for $n(\vec{r})$ (instead of a wave-function guess like in the HF method), then calculates the corresponding U_{eff} and solves the Kohn-Sham equations for the ψ_i. From these, one calculates a new electron density and starts again. This procedure is then repeated until convergence is reached. A non-iterative approximate formulation, called Harris functional DFT, is also possible [222] [219].
The third term in equation (18.10) represents any other electrostatic field in the system. It can be separated into two contributions: the electrostatic potential due to electron-nuclei interactions and an eventual external potential arising from applied electrostatic fields (given by one or more external sources).

Example 3.1: Very often the KS Hamiltonian is presented in atomic units, instead of SI units. In that case its various terms can assume slightly different aspects, for example the kinetic term becomes:

$$\widehat{T}_e = -\frac{\hbar^2}{2m_e}\Delta \quad \rightarrow \quad -\frac{1}{2}\Delta$$

A generic potential energy contribution instead can be:

$$U_{potential} = \frac{q^2}{4\pi\varepsilon_0 r} \quad \rightarrow \quad \frac{1}{r}$$

Consequently the Hartree potential of eqs. (18.5) and (18.8) becomes:

$$U_H = \frac{q^2}{4\pi\varepsilon_0}\int \frac{n(\vec{r'})}{|\vec{r}-\vec{r'}|}d\vec{r'} \quad \rightarrow \quad \int \frac{n(\vec{r'})}{|\vec{r}-\vec{r'}|}d\vec{r'}$$

that is often written as a convolution product (indicated with the symbol *):

$$U_H = n(\vec{r}) * \frac{1}{|\vec{r}|} = \int \frac{n(\vec{r'})}{|\vec{r}-\vec{r'}|}d\vec{r'}$$

with this notation the KS Hamiltonian becomes:

$$H_{KS}[n] = \sum_i^N [-\frac{\hbar^2}{2m_e}\Delta_i + U_{eff}(\vec{r_i})] \quad , \quad U_{eff}(\vec{r_i}) = n(\vec{r_i}) * \frac{1}{|\vec{r_i}|} + U_{XC}[n] + U_{EXT}(\vec{r_i})$$

as for example appear in some papers, like [222]. □

In summary, the many-body stationary Schrödinger's equation can be rewritten in N_e more manageable Kohn-Sham equations (since $\widehat{H} = \sum_i^{N_e} \widehat{H}_{KSi}[n]$):

$$\widehat{H}\psi = E\psi \quad \rightarrow \quad \widehat{H}_{KSi}[n]\ \psi_i = E_i\ \psi_i \quad , \quad i = 1, 2, .., N_e \tag{18.11}$$

As mentioned the main problem with the DFT method is that the exact functionals for exchange-correlation term are not known, except for the free electron gas. Therefore other approximations are needed. They are the subject of the next subsection.

18.5.2 Exchange-correlation functionals

The main issue in DFT is how to estimate the exchange-correlation functional $U_{XC}[n]$ that represents the electron-electron interaction in a precise way. The difference between various DFT methods is indeed the choice of the functional form for the exchange–correlation energy. It can be proven [219] that the exchange–correlation potential is a unique functional, valid for all systems, but an explicit functional form of this potential does not exist, except for special cases such as a uniform electron gas. Several functionals with different level of accuracy were developed [219], [139].
Notice that exchange–correlation functionals have a mathematical form containing parameters. There are two main philosophies for assigning values to these parameters, either by requiring the functional to fulfil theoretical criteria, or by fitting the parameters to experimental data, although in practice a combination of these approaches is often used [219]. The quality of exchange–correlation functionals will ultimately have to be settled by comparing the performance with experiments or high-level wave mechanics calculations. Such calibration studies, however, only evaluate the quality for the chosen selection of systems and properties. It has indeed been found [219] that the "best" functionals depend on the system and properties, some being good for molecular systems, others for delocalized (periodic) systems, and others again for properties such as excitation energies, etc... Since DFT is an active area of research, new and improved functionals are likely to emerge in future.
I will now introduce the main classes of functionals accordingly to what presented in [219]. For the next chapters of this work and for the purpose of setting up simulations it is just important to acquire the sensibility of the level of approximation of these functionals, such that it is possible to choose the most suitable for the specific application, considering if a more refined should be better depending on the structure to be simulated (e.g. a more refined functional could be needed to capture the essential of the interactions that are not well represented within a DFT framework - see previously in this section). The main functional classes are thus briefly considered in the following, while a comparison among them is provided in figure 18.2.

LDA and LSDA:

The simplest model is the Local Density Approximation (LDA), where the electron density is assumed to be slowly varying, such that the exchange–correlation energy can be calculated using formulae derived for a uniform electron density. In particular, within this approximation, U_{XC} in a given spatial point $\vec{r}$ is expressed only in terms of the electron density at that point $n(\vec{r})$. A simplified quantitative introduction is provided in [44], notice that the functional is usually taken proportional to the electron density to the power of 1/3 (in the same point): $U_{XC}[n(\vec{r})] \propto [n(\vec{r})]^{1/3}$. An straightforward improvement is obtained considering explicitly the electron spin, in the LSDA (Local Spin Density Approximation) functional.

GGA:

A significant improvement in the accuracy can be obtained by making the exchange– correlation functional dependent not only on the electron density but also on its derivatives. In the so called Generalized Gradient Approximation (GGA) methods, the first derivative of the electron density is included as a variable within the exchange-correlation functional. GGA methods are sometimes referred to as non-local methods, although this is somewhat misleading since the functionals only depend on the density (and derivative) at a given point, not on a space volume [219]. The following functionals belong to this class: B88 (from A. D. Becke), OPTX (OPTimized eXchange), LYP (Lee, Yang and Parr), PBE (Perdew–Burke–Ernzerhof), OLYP (combination of OPTX and LYP) and BLYP (combination of B88 and LYP), etc...

Name	Variables	Examples
Local density	n	LDA, LSDA, X_α
GGA	n , ∇n	BLYP, OPTX, OLYP, PW86, PW91, PBE, HCTH
Meta-GGA	n, ∇n, $\nabla^2 n$	BR, B95, VSXC, PKZB, TPSS, τ-HCTH
Hyper-GGA	n, ∇n, $\nabla^2 n$ *HF exchange*	H+H, ACM, B3LYP, B3PW91, O3LYP, PBE0, TPSSh, τ-HCTH-hybrid
Generalized RPA	n, ∇n, $\nabla^2 n$ *HF exchange* *Virtual orbitals*	OEP2

(a)

Functional	RMS (kJ/mol)	MAD (kJ/mol)
HF	649	885
LSDA	439	510
PW91	80	99
PBE	87	93
PKBZ	75	29
BLYP	41	40
PBE0	50	28
OLYP	40	25
B3LYP	40	21
VSXC	39	14
HTCT	33	30
τ-HCTH	31	
τ-HCTH-hybrid	26	
TPSS		24
TPSSh		16

(b)

Figure 18.2: A summary of the main classes of functionals with examples (a); and a comparison among them (b). RMS stands for Root Mean Square, MAD for Mean Absolute Deviation. The comparison is performed with experimental data on a set of molecules, data are taken from [219].

Meta-GGA:

The logical extension of GGA methods is to allow the exchange and correlation functionals to depend on higher order derivatives of the electron density, with the Laplacian ($\nabla^2 n(\vec{r})$) being the second-order term. Alternatively, the functional can be taken to depend on the orbital kinetic energy density τ. Inclusion of either the Laplacian or orbital kinetic energy density as a variable leads to the so-called meta-GGA functionals. Examples are: VSXC (Voorhis–Scuseria eXchange–Correlation), TPSS (Tao–Perdew–Staroverov–Scuseria), PKZB (Perdew–Kurth–Zupan–Blaha) that can be considered as the next improvement over the PBE functional, etc...

Hybrid or hyper-GGA:

In hybrid functionals a portion of exact exchange (from HF theory) is incorporated with the rest of the exchange–correlation energy from other sources (*ab initio* or empirical). The exact exchange energy functional is expressed in terms of the Kohn–Sham orbitals rather

than the density (so often is called implicit density functional). One of the most commonly used versions is B3LYP (Becke, 3-parameter, Lee–Yang–Parr), other functionals in this class are B3PW91, O3LYP, PBE0 (also denoted PBE1PBE), etc...

Generalized RPA:

In Generalized Random Phase methods also virtual molecular orbitals are considered. These are molecular orbitals that are empty in ground state, but that can be occupied in excited states. The improvement seems to be effective especially for weak interactions such as the van der Waals ones (that are still a problem with conventional functionals). Examples are: Optimized Effective Potential (OEP) methods (e.g. OEP1, OEP2, ...).

18.5.3 DFT and fitting parameters

An important question is now addressed: "should DFT methods be considered *ab initio* or semi-empirical?" If *ab initio* is taken to mean the absence of fitting parameters, LSDA methods are *ab initio* but gradient corrected methods may or may not be. The LSDA exchange energy contains no parameters and the correlation functional is known accurately as a tabulated function of the density. Some gradient-corrected methods (e.g. the B88 exchange and the LYP correlation), however, contain parameters that are fitted to give the best agreement with experimental atomic data, but the number of parameters is significantly smaller than for semi-empirical methods. Functionals such as VSXC contains a moderate number of parameters (21), while other functionals such as PBE are derived entirely from theory and can consequently be considered *ab initio*. If instead *ab initio* is taken to mean that the method is based on theory, which in principle is able to produce the exact results, DFT methods are *ab initio*. The only issue is that current methods cannot yield the exact results, even in the limit of a complete basis set, since the functional form of the exact exchange–correlation energy is not known.
Although gradient-corrected DFT methods have been shown to give impressive results, even for theoretically difficult problems, the lack of a systematic way of extending a series of calculations to approach the exact result is a major drawback of DFT [219]. The results converge toward a certain value as the basis set is increased, but theory does not allow an evaluation of the errors inherent in this limit. Moreover although a progression of methods such as LSDA, BLYP and B3LYP has provided successively lower errors for a suitable set of reference data, there is no guarantee that the same progression will provide better and better results for a specific property of a given system. Indeed, LSDA methods may in some cases provide better results, even in the limit of a large basis set, than either of the more complete gradient-corrected models. The quality of a given result can therefore only be determined by comparing the performance for similar systems where experimental or high-quality wave mechanics results are available. In this respect, DFT resembles semi-empirical methods.

18.5.4 DFT and basis sets

One approximation in essentially all electronic structure calculation methods is the introduction of a basis set. Expanding an unknown function, such as a molecular orbital, in a set of known functions is not an approximation if the basis set is complete (see also chapter 17). However, a complete basis set means that an infinite number of functions must be used, which is impossible in actual calculations. An unknown molecular orbital can be thought as a function in the infinite coordinate system spanned by the complete basis set. When a finite basis set is instead used, only the components of the molecular orbital along those coordinate axes corresponding to the selected basis functions can be represented. The smaller the basis set, the poorer the representation. Also the type of used basis functions influence the accuracy. The better a single basis function is able to reproduce the unknown

function, the fewer basis functions are necessary for achieving a given level of accuracy.
In mathematical terms each molecular orbital ψ_i is expanded in terms of the basis functions $\{\phi_\alpha\}_\alpha$, conventionally called atomic orbitals:

$$\psi_i = \sum_\alpha c_\alpha \phi_\alpha$$

Thus a molecular orbital is expanded as Linear Combination of Atomic Orbitals (LCAO). The expansion of the molecular orbitals leads to integrals of quantum mechanical operators over basis functions (see also section 17.4), and the ease with which these integrals can be calculated also depends on the type of basis function. And often the accuracy does not correspond to efficiency of calculation.
In principle any type of basis functions may be used: exponential, Gaussian, polynomial, cube functions, plane waves, etc... Nevertheless two very common functions are the Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO). Both STOs and GTOs can be chosen to form a complete basis. Historically, STOs were used due to their geometric similarity to the hydrogenic orbitals. They also remain a common choice in semi-empirical methods. However, they do have a significant drawback since there are no general and numerically stable solutions for the many-center integrals involving Slater-type orbitals. Consequently, Gaussian type orbitals are often the preferred choice. In figure 18.3 an example of STOs and GTOs is reported. The GTOs are inferior to the STOs in two respects. At the nucleus a GTO has a zero slope, in contrast to a STO which has a cusp (discontinuous derivative), and GTOs consequently have problems representing the proper behaviour near the nucleus. The other problem is that the GTO falls off too rapidly far from the nucleus compared with an STO, and the tail of the wave-function is consequently represented poorly. The increase in the number of GTO basis functions, however, is more than compensated for by the ease of which the required integrals can be calculated. In terms of computational efficiency, GTOs are therefore preferred.

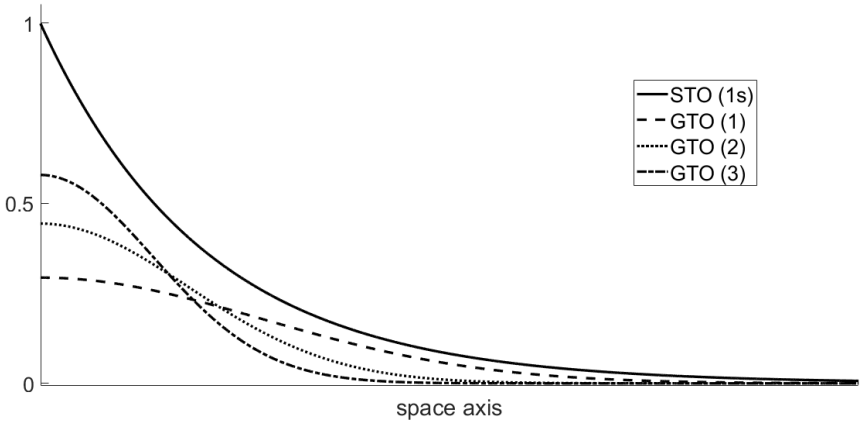


Figure 18.3: A qualitative example of comparison between STO and GTO functions. The exponential 1s STO orbital is compared with three different GTO possible functions.

There are mainly two guidelines for choosing the basis functions. One is that they should have a behaviour that agrees with the physics of the problem, since this ensures that the convergence as more basis functions are added is reasonably rapid. For bound atomic and molecular systems, this means that the functions should go toward zero as the distance between the nucleus and the electron becomes large. The second guideline is a practical one: the chosen functions should make it easy to calculate all the required integrals. To this purpose thinks again to the matrix formulation of quantum mechanics (section 17.4) in which an integral has to be one for each matrix element. Notice also that semi-empirical

methods usually approximate (at least some of) these integrals by means of fitting parameters (usually the so called two-electrons integrals that appears in the electron-electron interaction terms). Exactly like in a Fourier series (indeed it can be seen exactly like a Fourier decomposition) as the number of basis functions increases, the accuracy of the molecular orbitals improves. Nevertheless the computational cost increases exponentially with the number of basis functions. Indeed it is possible to show [219] that (in the large basis set limit) the SCF procedure involves a computational effort that increases (at least) as the number of basis functions to the fourth power.
The first criterion suggest the use of exponential functions located on the nuclei, since such functions are known to be exact solutions for the hydrogen atom. Unfortunately, exponential functions turn out to be computationally difficult. Gaussian functions are computationally much easier to handle, and although they are poorer at describing the electronic structure on a one-to-one basis, the computational advantages more than make up for this. For periodic systems, the infinite nature of the problem suggests the use of plane waves as basis functions, since these are the exact solutions for a free electron (think also that Bloch waves are intimately linked with plane waves - each Bloch wave is constituted by a forward and a backward plane wave).

Basis set classification

The smallest number of functions possible is a minimum basis set. Only enough functions are employed to contain all the electrons of the neutral atom(s). For hydrogen (and helium) this means a single s-function. The next improvement of the basis sets is a doubling of all basis functions, producing a Double Zeta (DZ) type basis. The term zeta stems from the fact that the exponent of STO basis functions is often denoted by the Greek letter ζ. A DZ basis thus employs two s-functions for hydrogen (called $1s$ and $1s'$), and so on...
The importance of a DZ over a minimum basis is clear if different chemical bond types are considered. For example a π-orbital is more delocalized than a σ-one. It will have a more diffuse electron distribution w.r.t. a σ bond. The optimum basis function exponent for representing a more delocalized orbital should be smaller than for representing a localized orbital. Moreover such an optimum exponent may vary with the direction (if e.g. in direction x a σ-bond occurs and in direction y a π-bond occurs). If only a single set of basis function is available (minimum basis), a trade-off will be necessary. Instead in a DZ basis, two sets of basis functions with different exponents are present. Thus, doubling the number of basis functions allows for a much better description of the fact that the electron distribution is different in different directions.
The next step up in basis set size is a Triple Zeta (TZ). Such a basis contains three times as many functions as the minimum basis; then Quadruple Zeta (QZ) and Quintuple or Pentuple Zeta (PZ or 5Z, but not QZ) are possible, and so on.

Polarization function

In most cases, higher angular momentum functions are also important, and these are denoted polarization functions. If methods including electron correlation are used, higher angular momentum functions are essential. Electron correlation describes the energy lowering by the electrons avoiding each other, beyond the average effect taken into account by Hartree–Fock methods. Polarization functions are added to the chosen basis. Adding a single set of polarization functions (p-functions on hydrogens and d-functions on heavy atoms) to the DZ basis forms a Double Zeta plus Polarization (DZP) type basis. Similarly Triple Zeta plus Double Polarization (TZ2P) can be formed.
Mixed basis sets are sometimes used, for example a DZP quality on the atoms in the active/central portion of the device part of the molecule and a minimum basis for the device leads. An important point (not discussed here) is to choose “balanced” basis set is to keep the error as constant as possible. The use of (not at all balanced) mixed basis sets

should therefore only be done after careful consideration. A good treatment is provided in [219].

Plane waves

Rather than starting with basis functions aimed at modelling the atomic orbitals (STOs or GTOs), and forming linear combination of these to describe orbitals for the whole system, one may use functions aimed directly at the full system. For modeling extended (infinite) systems, for example, a unit cell with periodic boundary conditions, this suggests the use of functions with an "infinite" range. For example the outer valence electrons in metals behave almost like free electrons, which leads to the idea of using solutions for the free electron as basis functions. Indeed in periodic systems like crystals the solution of the Schrödinger's equation leads to the arise of energy bands, and the electrons in a band can be described by orbitals expanded in a basis set of plane waves, which in three dimensions can be written as a complex function. To this purpose notice that a Bloch wave-function (solution of steady state Schrödinger's equation) is similar to a plane wave and moreover can be expanded as a superposition of two (a forward and a backward) plane waves.
Therefore plane wave basis functions are ideal for describing delocalized slowly varying electron densities, such as the valence bands in a metal; and indeed plane wave basis sets have primarily been used for periodic systems. Nevertheless they can also be used for molecular species by using a supercell approach, where the molecule is placed in a sufficiently large unit cell such that it does not interact with its own image in the neighbouring cells. Placing a relatively small molecule in a large supercell to avoid self-interaction consequently requires many plane wave functions, and such cases are handled more efficiently by localized Gaussian functions. A three-dimensional periodic system, on the other hand, may be better described by a plane wave basis than with nuclear-centred basis functions. Notice that the set of the all possible plane-waves (with all possible momentum values) can be shown to be a complete set, and for this reason they can be used as basis set.

Pseudo-potentials or effective core potentials

Systems involving elements from the lower part of the periodic table have a large number of core electrons. These are unimportant in a chemical sense, but it is necessary to use a large number of basis functions to expand the corresponding orbitals, otherwise the valence orbitals will not be properly described. This issue can be solved by modeling the core electrons by a suitable function, and treating only the valence electrons explicitly. The function modeling the core electrons is usually called an Effective Core Potential (ECP) in the chemical community, while the physics community uses the term Pseudopotential (PP). These terms results sometimes useful in understanding the principles behind some semi-empirical methods. A good treatment is again provided in [219].

18.6 The Extended Hückel theory and semi-empirical methods

In this section one of the most widely spread semi-empirical method, for the electronic structure calculation, in briefly addressed, namely the Extended Hückel Theory (EHT). Since it is a semi-empirical method the first part of this section is dedicated to an introduction to semi-empirical methods.

18.6.1 The semi-empirical approach

As mentioned in section 18.5.4, the computational cost of performing an HF or a DFT calculation scales formally as the fourth power of the number of employed basis functions

$\{\phi_\alpha\}_\alpha$. This arises from the number of integrals (especially the so called two-electrons integrals - see below) that are required for building the matrix form of quantum mechanical operators (see section 17.4). Semi-empirical methods reduce the computational cost by reducing the number of these integrals.
The first step in reducing the computational problem is to consider only the valence electrons explicitly (the ones in the outer shells that origin the chemical bonds and determine the main chemical-physical properties). The core electrons are accounted for by reducing the nuclear charge or introducing suitable functions for accounting them, that are not part of the basis set. To this purpose see also previously section 18.5.4, the part about the pseudo-potentials PP (or Effective Core Potential -ECP-), that are exactly the functions mentioned above. Furthermore, only a minimum basis set (the minimum number of functions necessary for accommodating the electrons in the neutral atom) is used for the valence electrons (to this purpose see again section 18.5.4). The large majority of semi-empirical methods to date use only s-type and p-type basis functions, and they are usually of Slater type orbitals (again refer to section 18.5.4), i.e. exponential functions.
The central assumption of semi-empirical methods is the Zero Differential Overlap (ZDO) approximation, which neglects all products of basis functions that depend on the same electron coordinates when located on different atoms. To be clearer, said ψ_i the i-th molecular orbital, it is expressed in terms of the basis set $\{\phi_\alpha\}_\alpha$ (the atomic orbitals) as follows:

$$\psi_i = \sum_\alpha c_\alpha \phi_\alpha$$

Then said an atomic orbital on centre A as ϕ_A and an atomic orbital on centre B as ϕ_B, the ZDO approximation corresponds to $\phi_A \phi_B = 0$. Note that it is the product of functions on different atoms that is set equal to zero, not the integral over such a product. The consequences of the ZDO approximation are [219]:

- The overlap matrix S is reduced to a unit matrix: I
- One-electron integrals involving three centres (two that come from the basis functions and one from a quantum mechanical operator) are set to zero (these are product of the kind $\langle \phi_A | \widehat{F} | \phi_B \rangle$, where $\widehat{F}$ is the quantum mechanical operator)
- All three- and four-centre two-electron integrals are neglected. These integrals come out in doing the calculations (in building the so called Fock matrix), and corresponds to products of the kind $\langle \phi_A \phi_B | \phi_C \phi_D \rangle$

To compensate for these approximations, the remaining integrals are made into parameters, and their values are assigned based on calculations or experimental data or from other more accurate theoretical calculations [139]. Exactly how many integrals are neglected, and how the parameterization is done, defines the various semi-empirical methods. The following approximations are done possible:

- Neglect of Diatomic Differential Overlap (NDDO): in this approximation there are no further approximations than those mentioned above. The overlap matrix becomes: $S_{mn} = \langle \phi_m | \phi_n \rangle = \delta_{mn} \delta_{AB}$, where δ indicates the Kronecker delta. Thus S_{mn} is null each time different basis functions are considered $m \neq n$ but also each time different atom centers are considered $A \neq B$.
- Intermediate Neglect of Differential Overlap (INDO): in this approximation all two-centre two-electron integrals that are not of the Coulomb type, in addition to those neglected by the NDDO approximations are neglected.
- Complete Neglect of Differential Overlap (CNDO): in this approximation also the the one-centre two-electron integrals are neglected. Instead the approximations for the one-electron integrals in CNDO are the same as for INDO.

The important point to keep in mind is that all semi-empirical method assume the ZDO approximation (that is the NDDO), and then that successive approximation are possible, and passing from NDDO, to INDO and to CNDO the approximations are worse, that

means that more fitting parameters are present in that method. Fitting parameters that compensate for the neglected integrals and are used to represent the (few) integrals that are performed. Again fitting parameters that are derived from experimental data. Many different versions exist today, and they differ in the exact way in which these parameters are derived. Some of the names are CNDO/1, CNDO/2, CNDO/S, CNDO/FK, CNDO/BW, INDO/1, INDO/2, INDO/S and SINDO1. And moreover better performances are generally obtained with the following methods: MINDO, Modified NNDO, MNDO, AM1, PM5, etc... Typical errors in semi-empirical methods are nowadays only slightly greater than the average DFT ones. For example typical errors on bond lengths are of the order of $10^{-2} \div 10^{-3}$ Å, errors on angles of few degrees.

Semi-empirical methods performances

It is possible to show that semi-empirical methods formally scale as the cube of the number of basis functions in the limit of large molecules [219]. For this reason the current limit of semi-empirical methods at around 1000 atoms [219].
Semi-empirical methods share the advantages and disadvantages of force field methods: they perform best for systems where much experimental information is already available but they are unable to predict totally unknown compound types. Nevertheless, the dependence on experimental data is not as severe as for force field methods. Once a given atom has been parameterized, all possible compound types involving this element can be calculated. Semi-empirical methods are zero-dimensional, just as force field methods are. There is no way of assessing the reliability of a given result within the method. This is due to the selection of a minimum basis set. The only way of judging results is by calibration, i.e. by comparing the accuracy of other calculations on similar systems with experimental data.
Notice that semi-empirical models provide a method for calculating the electronic wavefunction, which may be used for predicting a variety of properties, e.g. the molecule polarization. Nevertheless from *ab initio* calculations that good results require a large polarized basis set including diffuse functions, and the inclusion of electron correlation, that are not considered in semi-empirical methods. Depending on the specific application the semi-empirical results can also be accurate. In this sense the result should be somehow validated by comparison with experimental data or more accurate *ab initio* approaches, as already mentioned above.
Generally semi-empirical methods, like force field ones, are not at all suitable for accurate estimations of molecule energy levels or for molecular geometry optimization, unless some modifications are considered. They instead are very computationally efficient and find applications in many fields in which qualitative results are enough.

18.6.2 The EHT method

The extended Hückel method (EHT) is a semi-empirical LCAO method, which exploits all valence orbitals of the atoms as the basis functions. Contrarily to the original Hückel method, the extended one takes into account not only π-orbitals but also σ ones. These atomic orbitals are approximated with Slater Type Orbitals (STOs) which allows the overlap matrix S_{ij} to be calculated efficiently. As already mentioned (see section 18.5.4), the STOs are orbitals characterized by having a blunt exponential decay centered in the atom nuclei. In this way, by using this basis set, the matrix elements can be defined by very few parameters. In particular, the matrix elements of the Hückel Hamiltonian are described by the following equation [225]:

$$H_{ij} = \begin{cases} -E_i, & \text{if } i = j \\ \frac{c}{2} S_{ij}(H_{ii} + H_{jj}), & \text{if } i \neq j \end{cases} \tag{18.12}$$

where the diagonal elements ($i = j$) of the Hamiltonian are approximated with the valence orbital ionization energies E_i, taken either from experimental data or calculated by means of more advanced methods. The off-diagonal elements ($i \neq j$) are proportional to the overlap of the *i-th* and *j-th* orbitals weighted by a constant c that is usually taken equal to 1.75. Thus, given a molecular geometry, the overlap matrix and thereafter the Hamiltonian can be calculated.

Several variants are possible, but often the overlap integrals are actually calculated, i.e. the ZDO approximation is not invoked. It is possible having EHT methods with NDDO, INDO or CNDO. If the CNDO approximation is assumed, performances may be limited since it does not take into account electron-electron repulsive interactions.

Usually EHT methods are non-iterative. Since the diagonal elements only depend on the nature of the atom (i.e. the nuclear charge), this means for example that all carbon atoms have the same ability to attract electrons. But in general, it is unlikely that all carbon atoms have the exact same charge, i.e. owing to the different environments their ability to attract electrons is no longer equal. In order to overcome this drawback self-consistent EHT calculations can be performed. Such methods are called self-consistent (or charge-interaction) Hückel methods, and in these methods the matrix elements are modified by the calculated charge.

The main advantage of extended Hückel theory is that only atomic ionization potentials are required, and it is easily parameterized to the whole periodic table. Extended Hückel theory can be used for large systems involving transition metals, where it often is the only possible computational model [219]. Moreover in many cases show the correct trend for geometry perturbations corresponding to bond bending or torsional changes, and thus qualitative features regarding molecular shapes may often be predicted or rationalized from EHT calculations [219].

In summary, this approach requires only a fraction of the computational effort needed by the more elaborated *ab initio* methods and often provides reasonably accurate quantitative results that give insight into the essential physics (or chemistry) [226], [225].

Finally notice that recent work in this area has used an approach to parameterize against density functional results, thereby providing a computationally very efficient model capable of yielding fairly accurate results [219].

18.7 Intermolecular interactions

In the practical part (part II) of this work I will consider gas sensor applications. The structure will be mainly the one of a molecular wire, and depending on the presence of a target molecule in proximity to the sensor, the conductance of the molecular channel will be changed. The physical principle of the device is thus linked with weak intermolecular

interactions, such as the van der Waals ones, that modify the channel molecular orbitals such that the transport is significantly changed. The aim of this section is to provide a review of the van der Waals intermolecular interactions and to briefly highlight how to consider them in a electronic structure calculation (mainly considering the DFT approach).

18.7.1 van der Waals and intermolecular interactions

In order to review the van der Waals forces it is better to firstly review briefly the different kinds of chemical bonds and chemical interactions between atomic elements. References can be [227], [228].
Valence electrons in atoms (i.e. the electrons in the most external shells of an atom) are those that generate the chemical bonds, and give rise to many chemical-physical properties of the considered element. Accordingly with Lewis theory, these electrons are those belonging to an incomplete octet, and the chemical bonds arise for the attempt of completing it. Covalent bonds are created between atoms with enough high electronegativity, in these kind of chemical bonds each atom provide one electron, such that two electrons are shared between them and constitute the bond. Covalent bonds can be pure (between atoms of the same element) or polar (between atoms of different elements). Moreover it can also be dative if a single atom provides both the electrons, and a double covalent bonds if two pairs of electrons are provided. Analogously a triple covalent bond involves three pairs of electrons. The triple is stronger w.r.t. the second that is stronger w.r.t. to the simple covalent one. This means that a greater amount of energy (binding energy) should be provided to break a triple bond w.r.t. a second one, and moreover the bond length of the triple will be lesser than the one of double (all analogous between double and simple bond). The covalent bond length is of the order of $1 \div 2$ Å. Moreover ionic bonds are also possible, usually if the electronegativity difference between the two element is greater than 1.9. In this case an element provides one or more electrons to the other, and both becomes ions (positive ion the first one and negative the second one). The donor atom is called cation while the other is called anion. The ionic bond is the result of the electrostatic interaction between the two ions. It is usually the weaker among the mentioned bonds, but it is stronger than intermolecular interactions. Notice that usually a "clean" ionic bond is not possible and instead there is "a bit" of electron sharing between the anion and the cation.
In addition to these bonds it is possible to have also intermolecular interactions, i.e. physical-chemical interactions among molecules. Different molecules can undergo to attractive (or also repulsive) interactions. The intermolecular interactions are four in total: the hydrogen bond and the van der Waals forces, that are in number three since they includes three different cases. These are at the origin of chemical bonds in hydrogen-bond solids and molecular solids respectively. Notice that they are weaker than intramolecular bonds described above (i.e. their binding energy is lower).

Hydrogen bond

The hydrogen bond is a particular case of electrostatic dipole-dipole interaction. It is formed between molecules in which an hydrogen atom is covalently bond with an atom with high electronegativity (and small dimensions), typically fluorine, nitrogen and oxygen. Due to the high electronegativity the electron has more probability of being found nearby the higher electronegativity atom. This means that its wave-function or orbital (the molecular orbital or the electron cloud) is spatially deformed toward the more electronegative atom. Consequently the molecule present a permanent electric dipole, with a small negative charge δ^- close to the high electronegative atom and a small positive charge δ^+ close to the hydrogen atom. An example can be the water molecule H_2O. Because of this permanent electric dipole different molecules can be bond together, due to the electrostatic interaction. Thus there is the formation of weak (if compared to the above intramolecular bonds such as the covalent bonds) chemical bonds between the hydrogen atom of one molecule and the

electronegative atom (e.g. F, N, O) of another molecule. Its binding energy is of the order of $20 \div 50$ kJ/mol, and notice that this value is comparable (or even less!) with the typical errors of the DFT methods reported in figure 18.2 (b). This means that for these kind of weak interactions such a methods are not suitable, and indeed corrections must be used (as described below - subsection 18.7.3). The origin of this interaction is (in a semiclassical picture as the one just provided) an electrostatic interaction between two electrical dipoles. The fact that the hydrogen atom is very small, and the fact that it contains only one electron (thus there are no electrons that can somehow screen the electric field and thus the interaction) make the hydrogen bond particularly intense w.r.t. other intermolecular interactions (van der Waals), thus making the binding energy particularly high. Because of the reasons just mentioned this kind of intermolecular interaction is particularly important, and since it is possible only with hydrogen atoms it is called hydrogen bond. Notice that the hydrogen bond is directional (like the covalent bonds). Finally notice that it is possible to have hydrogen bonds even within the same molecule (if the molecule is large, i.e. made by an enough number of atoms), usually it happens between different functional groups composing the molecule.

van der Waals forces

Intermolecular interactions (i.e. interactions among molecules) are attractive or repulsive forces that mediate the interaction between molecules, they can occur between both polar and non-polar molecules. They have again an electrostatic nature, and they are usually referred as van der Waals forces. Intermolecular interaction (or binding) energy is very low ($0.1 \div 10$ kJ/mol) if compared with covalent or ionic bonds ($100 \div 1000$ kJ/mol), moreover the interaction distance is usually short (of the order of 4 Å) and the strength of interaction decreases rapidly with distance. The van der Waals forces are responsible of several interesting chemical and physical properties of materials, especially in changes of the state of aggregation of matter (solid to liquid, liquid to gas and vice versa). The van der Waals forces (named after the Dutch physicist Johannes Diderik van der Waals) include actually three different types of interactions:

- Keesom's interactions:
  The Keesom interactions involve permanent dipoles. They result from the interaction of a permanent dipole with another permanent dipole. Indeed it is possible to have an electric permanent dipole in a polar molecule, and as already mentioned this may happen if there is one (or more) atom with a significantly higher value of electronegativity w.r.t. to the others composing the molecule. In this case the electron(s) has an higher probability of being found close to the high electronegativity atom, creating a (small) negative charge δ^- around it. Consequently a (small) positive charge δ^+ is induced around low electronegativity atoms. As known from basic physics an electrical dipole generates an electric field that allows the polar molecules to interact. In particular in nature there always the trend of achieving a stable configuration (if no external stimuli are applied), that coincides with the achievement of the thermodynamic equilibrium (i.e. minimum potential energy configuration). If many permanent dipoles are present in the same volume one close to the other then they will align with opposites poles (i.e. δ^+ and δ^-) close the one to the other. Usually Keesom's interactions are negligible in gases due to the high kinetic energy of particles that allows for overcome the Keesom's interaction energies. Notice that it is also possible that the permanent dipoles have a different origin from the one specified above, for example they can be due to the presence of molecular ions or molecular multipoles (nevertheless the treatment does not change).
- Debye's interactions:
  This kind of interaction occurs between a polar and a non-polar molecule. If a non-polar molecule undergoes to an electric field then the electron clouds (or molecular

orbitals) are distorted. Indeed the electrons have the inclination to go in opposite direction to the field. As a result the molecule is temporarily polarized. Thus an induced dipole is created. The same effect is obtained if a polar molecule is nearby the non-polar one. In that case opposite charges are attracted and the non-polar molecule results polarized, until the distance between the two molecules is not so big to make the interaction negligible. The intensity of the interaction is of course proportional to the strength of the permanent dipole of the polar molecule.

– London's interactions:
These interactions occur between a fluctuating dipole and an induced dipole. Indeed non-polar molecules can instantaneously become polar, due to motion of the electrons around the nuclear centers. Usually in a (neutral) atom the electron cloud is symmetric w.r.t. to the nucleus, nevertheless for short time intervals the electrons can be localized in a certain region of space giving rise to an instantaneous electrical dipole. This is called fluctuating dipole. Again the instantaneous electric field that is generated by this dipole can have effects on another atom/molecule nearby this one, originating an induced dipole. In response to the induced dipole the first atom/molecule can change its configuration, such that the total potential energy of the system is minimized. This means that opposite charges are attracted to the same side (as explained previously). Nevertheless this can have again effects on the induced dipole that can change again orientation (passing in between the neutral configuration). This mechanism is at the basis of the London interactions. The London interaction strength increases with the molecular molar mass.

It is possible to derive classical expressions for the Keesom, the Debye and the London interactions (not reported since non-relevant for the quantum mechanical treatment - see next section 18.7.2). Moreover all the intermolecular interactions are anisotropic, that means that they depend on the relative orientations of the involved molecules. This is obvious if their physical origin (above) is considered.
Notice that it is possible to have other intermolecular interactions, not included in the van der Waals one. Nevertheless they all have essentially an electrostatic origin. In particular it is possible to have ionic bonds/interactions between molecules. The principle is analogous to the one of ionic bonds reviewed before: a molecule (or a functional group) provides all the electrons of the bond (and becomes the cation) while the other is the anion. In addition other kind of dipole–dipole interactions are possible between molecules which have permanent dipoles. Also ion–dipole and ion–induced dipole interactions are possible. Notice that generally in the latter case ions are involved, giving rise to forces that are stronger than dipole–dipole interactions because the charge of any ion is greater than the charge of a dipole moment.
A mention is also due to the so called steric effects [227], [228]. They are non-bonding interactions that influence the shape (conformation) and reactivity of ions and molecules. Notice that steric effects result from repulsive forces between overlapping electron clouds. An example can be the real shape of benzene molecule that is non-planar, this because the repulsive forces between electrons (that correlate their motion to avoid each other) cause it to deform and assume the so called chair-like shape (bent instead of planar). These effects are in great majority again a consequence of electrostatic interactions and correlation between electrons.
Nevertheless the point that now should be clear is that all these kinds of interactions have essentially an electrostatic origin. For this reason they all have the same quantum mechanical treatment, that in practice is very simple: all these kind of interactions are automatically included in the considered system Hamiltonian (to this purpose refer also to the next section 18.7.2). Indeed once the geometry is chosen the Hamiltonian can be built following the methods described in the previous sections of this chapter, and the wave-functions satisfying the steady state Scrhödinger's equation are automatically accounting for all these effects, that indeed appear within the Hamiltonian as potential

energy contributions, either relative to electron-nuclei interactions or to electron-electron interactions. Notice that choosing the geometry means to fix and optimize it by means of an optimization procedure aimed in minimizing the potential energy in order to obtain the equilibrium configuration. The problem in modeling all the intermolecular interactions is the fact that they have small interaction energies (van der Waals forces are of the order of $0.1 \div 10$ kJ/mol), that are often below or of the order of the root means square or maximum deviation errors of the DFT and the EHT methods (refer to figure 18.2 to have an idea). This means that some kind of corrections are needed, and they are the subject of the next sections (refer especially to section 18.7.3).

18.7.2 Quantum modeling of intermolecular interactions

So far the classical models for intermolecular interactions were briefly reviewed. As mentioned, intermolecular forces between two molecules or atoms can be attractive or repulsive and occur from either momentary interactions between molecules (e.g. London dispersion force) or permanent electrostatic interactions between dipoles. The aim of this section is to point out how they are modeled with the quantum mechanical approach.
It was already noticed at the end of the previous section that the origin of intermolecular interactions is essentially the electrostatic attraction or repulsion between permanent or momentary dipoles or charges. Moreover it was highlighted that they result from the (both static and dynamic) geometry of the wave-functions, in the sense that the shape of the electron cloud and the electron correlation can somehow influence the interaction. Think for example to steric effects or to the deformation of the electron cloud due to the proximity of a charge/dipole/molecule. As already mentioned all these interactions are somehow already included within the molecular Hamiltonian.
The standard analytical approach for considering intermolecular interactions is by means of the so called Rayleigh-Schrödinger perturbation theory [229]. In this model the (weak) intermolecular forces are treated by means of a perturbational approach. In particular the total system Hamiltonian (for example of the system of two interacting molecules) is written as the sum of an unperturbed Hamiltonian and a perturbative term. In the case of two interacting molecules the first term is the sum of the two equilibrium (and unperturbed) molecular Hamiltonians, while the latter if the contribution considering the interactions between them. Once the total system Hamiltonian is known it is possible to proceed as usual in solving the Schrödinger's equation. Since it is a perturbative approach it holds true if the intermolecular interactions are weak w.r.t. to intramolecular ones (that are e.g. covalent bonds), and from the above discussion it is known that this hypothesis is usually true (van der Waals interactions for example involve energies in the range $0.1 \div 10$ kJ/mol much lesser than covalent/ionic bonds $100 \div 1000$ kJ/mol). This approach allows for evaluating the intermolecular forces, and it can be useful in considering them in wave-function based methods like electron-correlation (post-HF) methods presented previously (see section 18.7), e.g. the Møller–Plesset Perturbation theory based methods [229]. A remark on the so called dispersion forces will be useful in the next section 18.7.3. Long-range intermolecular forces, also known as London or dispersion forces (see previous section 18.7.1), are well represented by the Taylor's second order expansion of the perturbative term. This procedure leads to an attractive potential with a leading term of the kind: $-C_6 R^{-6}$ (where C_6 is a constant and R is the distance). The classical explanation for this term was already provided in the previous section 18.7.1, and it involves the creation of temporary dipoles. London coined the name "dispersion effect" (from which the name "dispersion forces" or "London forces") for the attraction between noble gas atoms after he noted the presence of aforementioned terms that appeared in the equations for the intermolecular interactions, and that were analogous to oscillator strengths.

Other methods that include a perturbative approach in modeling of intermolecular inter-

actions can be other electron-correlation methods (like the coupled cluster one) or the Quantum Monte Carlo (QMC) one.
Nevertheless all such electron correlation methods and the QMC one, have the great disadvantage of being unusable for large systems with more than few tens of atoms (let's say 50 atoms). Nevertheless a complete device, for example a molecular wire or transistor, can widely overcome that limit, also because the contacts should be considered as atomistic in order to make plausible predictions. For this reason DFT or semi-empirical methods are usually preferred. In such methods it is possible to consider intermolecular interactions by means of the so called "supermolecular approach". It is referred to all quantum chemical standard methods (e.g. DFT, EHT, etc...) that include some kind of correction aimed in modeling intermolecular interactions. Notice that the electron correlation should be considered in these approaches in order to get reliable results (especially for dispersion/London forces) [229], [230]. An important issue in supermolecular approaches is the basis set superposition error (BSSE). This is the effect that the atomic orbital basis of one molecule improves the basis of the other molecule, and it can give rise to artifacts, since the intermolecular interactions are distance dependent (see e.g. the above result on London forces). More details are provided in the next section.

18.7.3 Supermolecular approach and DFT corrections

The way of accounting for intermolecular interactions in electron correlation methods was briefly mentioned in the previous section. This section instead briefly introduces the supermolecular approach. Such an approach consists in introducing suitable corrections to both DFT and semi-empirical methods in order to account for intermolecular interactions (like hydrogen bond, van der Waals and London dispersion forces). It found recently applications in accurately predicting some large supramolecular complexes or bio-macromolecular geometries such as the simulation of protein folding [230].
Different methods (namely HF, DFT, and semi-empirical) are now considered and it is mentioned how to consider intermolecular interactions within them. A comparison among them is also proposed. The material of this section is mainly adapted from [230].
Hartree–Fock method can provide an adequately accurate for hydrogen-bonding interactions (with, however, significant underbinding) in some organic compounds. Nevertheless it fails spectacularly for considering dispersion (London) in bound complexes. A dispersion corrected Hartree–Fock method (the HFD method) was proposed [230], but it has important limits and thus the DFT-based methods are usually considered much better in terms of performance with comparable computational performances.
Traditionally, it has been generally believed that semi-empirical methods are not particularly well suited to hydrogen bonding problems [230]. Some semi-empirical methods have been re-parameterized and additional empirical terms were added to increase the accuracy for describing weak intermolecular interactions [230]. Others instead adopt some of the below mentioned DFT-D corrections in the semi-empirical environment [230]. In general is a good practice to consider unreliable a semi-empirical method for accurate estimation of weak intermolecular interactions, unless it is not "validated" for the specific case of interest by means of comparisons with experimental data or DFT methods (or more refined *ab initio* methods).
The situation is even worse for molecular mechanics (force fields) based methods. They are not generally high accuracy methods, even if the presence of many accurate experimental data can give good values for fitting parameters and thus fair results, but only in specific particular cases.
Instead DFT-based methods allow for the usage of corrections to consider electron correlations and weak intermolecular approaches that usually lead to very good results sometime comparable with the ones of post-HF or perturbative methods [230]. In general LDA accuracy can only be described as sporadic, and therefore its application in intermolec-

ular interactions is not recommended. GGA functionals typically show repulsive behavior for dispersion bound complexes similar to Hartree–Fock method, with the corrections mentioned below they provide good results. Meta-GGA functionals show only marginal improvements over the GGA ones, while some hybrid-GGA functionals give reasonable answers for particular problems of interest, but should be used with extreme caution [230]. The main available DFT corrections for intermolecular interactions are:

- DFT-D (Grimme-D2 / Grimme-D3):
  These class of corrections include empirical methods to account for dispersive interactions in practical calculations within the DFT framework. They were tested for a wide variety of molecular complexes resulting in good predictions. For example for stacked aromatic systems the DFT-D (BLYP) model seems to be even superior to standard perturbative approaches. The good results obtained, in a variety of diverse examples [230], suggest that the DFT-D approach is a practical tool for describing weak intermolecular interactions [230]. The dispersive energy is described by a damped interatomic potential of the form $-C_6R^{-6}$ (see previously section 18.7.2). This simple approach has recently been improved regarding accuracy introducing less empiricism with the so called DFT-D2 (or Grimme-D2) model [230]. In particular in this new version the most important parameters (like R and C_6) are computed *ab initio*, and general applicability to most elements of the periodic table was proved. An important change is present in the so-called DFT-D3 (or Grimme-D3) method [230], in which the C_6 dispersion coefficients are dependent on the molecular structure which accounts for subtle effects, e.g., the hybridization state of an atom changes. The DFT-D methods do not only work for molecular complexes and intramolecular dispersion effects but are rather general [230].
- Range Separated and Dispersion Functionals (vdW-DF):
  In this method the non-local correlation is calculated explicitly in a non-empirical manner[230]. It is typically used with standard GGAs like PBE. The overall quality of vdW-DF results for typical non-covalent complexes is not better than with DFT-D but this method has the great disadvantage of requiring substantially more computation time (w.r.t. DFT-D ones).
- Dispersion Calibrated Effective Potentials (DCEP):
  It is a novel approach whereby the missing dispersion effect in traditional DFT methods is introduced using a suitably engineered effective core potential (ECP - see also section 18.5.4).

In the practical of this work (part II) the first correction, namely the DFT-D (in its Grimme-D2 and Grimme-D3 verions) will be used for modeling molecular electronic gas sensors.
Important considerations concern the basis set. The electron density around one nucleus may be described by functions centred at another nucleus. This is especially troublesome when calculating small effects, such as energies of van der Waals or intermolecular interations or even hydrogen bonds, as explained in a while. It was said in section 18.5.4 that a complete basis should include an infinite number of basis functions and this is of course impossible. Moreover in order to have accurate results a great number of basis functions should be consider, but again this is often unfeasible in practice, due to the fact that the computational effort increases as the fourth power of the number of chosen basis functions. Thus the truncation of the basis set leads inevitably to an error (in particular an overestimation of the energy levels is often obtained). Usually the interest is in relative energies, e.g. ratio between energy levels, thus these errors are not so relevant in many practical cases (even if absolute errors are large -e.g. hundreds of kJ/mol- the relative ones are not). A balanced basis set (as mentioned in section 18.5.4) is able to provide a (more or less) constant error, thus leading to small relative errors (on ratio of energies). Moreover fixing the position of the basis functions to the nuclei allows for a compact basis set [219], otherwise sets of basis functions positioned at many points in the geometrical space would be needed, leading to a too large number of basis functions for practical implementations. When com-

paring energies at different geometries, however, the nuclear fixed basis set introduces an error, because as mentioned previously the various basis functions overlap in the nuclei. To better visualize the problem let's consider two water molecules. They may give rise to an hydrogen bond, and its energy can be estimated for example by considering the total energy of the system of two water molecules to which it is subtracted two times the energy of the single water molecule. Nevertheless different nuclear geometries leads to different superposition of the basis functions, thus different overlap errors, and in particular with more nuclei the basis set results improved because of basis functions overlap in the nuclei that lead to a whole amelioration of the completeness of the basis set (it is thus "more complete" than the basis set of the single water molecule). Consequently the system of two water molecules will have a more realistic energy w.r.t. to the isolated water molecules, that is a lower energy (the basis set truncation error results in an increase of the energy levels, since now a "more complete" basis set is used this truncation error is lesser and the energy is lower). And thus an overestimation of the hydrogen bond energy occurs, but it is wrong. This effect is known as the Basis Set Superposition Error (BSSE). In the limit of a complete basis set, the BSSE will be zero, and adding more basis functions will not give any improvement. The conceptually simplest approach for eliminating BSSE is therefore to add more and more basis functions, until the interaction energy no longer changes (that is: the two molecules basis set is no "more complete" than the single molecule one). Unfortunately, this requires very large basis sets and is unpractical. Mainly two corrections for BSSE are possible, namely the chemical Hamiltonian approach (CHA) and the Counterpoise (CP) correction. In the first case the basis functions mixing is prevented *a priori* in the way in which the Hamiltonian is built. In the latter the BSSE is reduced by re-performing all the calculations with the mixed basis set (in the previous example in the case of two water molecules). The two methods tend to give similar results [219].
In conclusion to this section, few recommendations are cited from [230]:

- Whenever possible, approximate methods (force field, semi-empirical or HF) that are unavoidable in many practical applications should be checked or benchmarked against large (or complete) basis set results or experimental/*ab initio* more accurate methods.
- Electron correlation or post-HF methods work fine but are feasible only for small systems (¡50 atoms)
- When the issue of London dispersion interactions is carefully considered within DFT, this yields a similar or often even better performance than electron correlation or post-HF methods. The DFT-D method has proven as an accurate and robust computational tool. The main advantage of DFT-based approaches is that they can be applied also in electronically complicated situations (e.g. a complete molecular electronic device).
- The biggest problem in DFT is the choice of the functional approximation. Computationally cheap (meta-)GGAs (e.g., BLYP, PBE, or TPSS) can be recommended. The currently highest level of approximation in DFT is represented by double-hybrid functionals (e.g. B3PLYP) that also perform very well for non-covalent interactions, even if they should used carefully in some particular situations (see previous discussion).
- Semi-empirical methods (and this also holds for many force fields) mainly suffer from a poor description of the electrostatic interactions which is demonstrated by their notoriously bad performance for hydrogen-bonded systems (for which even the simplest GGAs work very well). Because also their description of polarization (induction) contributions to binding is deficient, these methods can only be recommended for non-polar systems or alternatively employed after a careful benchmark validation is provided.

CHAPTER 19

An introduction to transport in molecular devices

In the introductory section **??** it was briefly introduced the concept of molecular electronics, in which single molecules (or small packets of them) are exploited to create whatever electronic device. Then in section 5.1 the definition of molecular electronic sensors and their working principle were briefly introduced, only qualitatively. The purpose of this chapter is provide an overview of the transport features in such systems, while a rigorous introduction to transport is left to the next chapter. In section 19.1 an oversimplified model for transport through single molecule devices, or more generally through quantum dots (i.e. zero-dimensional systems), is provided. It is useful to catch the main physical insights of transport features in quantum dots. In section 19.2 it is introduced the concept of mesoscopic systems, to which quantum dots (and molecular) devices usually are considered to belong. In section 19.3 is very briefly addressed the Landauer-Büttiker transmission formalism for transport in mesoscopic systems. It is useful in a deeper understanding of the simple transport model presented in section 19.1, and in understanding some of the features of the more advanced Non-Equilibrium Green's Function (NEGF) formalism. The purpose of section 19.5 is to provide an introductory overview of the advantages of using the NEGF formalism for modeling transport in mesoscopic systems (while the formalism is presented in the next chapter).

19.1 A simple model for transport through quantum dots

The aim of this section is to provide an oversimplified model for the transport through quantum dots, i.e. zero-dimensional (0D) systems. Such a model is the one presented in the course "Micro & nano systems", taught at Politecnico di Torino during the academic year 2018/2019 by professors Mariagrazia Graziano and Gianluca Piccinini. It is analogous to the "toy model" introduced in the first chapter of [44], and moreover it is essentially the same of the model presented in [45], exactly concerning the transport through quantum dots. In this section I will adopt the notation used in the above mentioned "Micro & nano systems" course, that can be slightly different from the ones of [44] and [45]. Even if this model is very simplified, the main physical insights of electronic transport in quantum dots can be easily understood, also in a quantitative way. For this reason I think it can be a useful preparatory treatment, before entering into the details of the more complete Non-Equilibrium Green's Functions Formalism (subject of next chapter).

A quantum dot, or zero dimensional system, is a system in which quantum confinement occurs in all the three spatial directions. Molecules can be thus thought as quantum dots (indeed electrons are confined within the molecule in all directions). A feature of quantum confinement is the arising of quantization, e.g. in energy levels. To this purpose one can think to a 3D quantum well, like presented in section 17.3.7 (the parts about the finite and the infinite height quantum wells). A first approximation can be the so called "bounding box" approximation. In which the quantum dot is though to be 3D infinite height quantum well with zero potential (i.e. flat potential) inside it. This is a strong simplification, but it is useful to highlight the discrete nature of energy levels inside the quantum dot. Indeed by proceeding as already done in section 17.3.7, and thus exploiting the separation of variables, and supposing $L_x = L_y = L_z = L$ it follows (see equations (17.71), (17.72), (17.67)):

$$E = \frac{\hbar^2\pi^2}{2mL^2}(n_x^2 + n_y^2 + n_z^2) \quad , \quad n_x,\ n_y,\ n_z \in \mathbb{N} \setminus \{0\}$$

$$\psi(x,y,z) = \left(\frac{2}{L}\right)^{3/2} sin\left(n_x\frac{\pi}{L}\right)\ sin\left(n_y\frac{\pi}{L}\right)\ sin\left(n_z\frac{\pi}{L}\right)$$

that holds true if $U(x,y,z)$ can be written as the sum of three independent contributions: $U(x,y,z) = U(x)+U(y)+U(z)$. This is an oversimplification and it is wrong for molecules. The correct way of obtaining the energy levels (energy eigenvalues) of a molecule is to employ one of the methods described in the previous chapter 18. Anyway the important point is that a quantum dot, both in the oversimplification (and absolutely inaccurate and wrong) assumption of being a 3D infinite height quantum well, both in more accurate approximations (like the ones discussed in chapter 18), is a system characterized by discrete energy levels, due to quantization phenomena in all the three directions. Thus the electrons are not free to move in no direction, and they result confined. Nevertheless, it is possible to place a quantum dot in between two contacts and to have a current flowing between them. But as will be discussed in while, the current-voltage characteristics of such a system will have peculiar features, typical of 0D systems only, arising, precisely, from quantum confinement.

For the moment let's consider a quantum dot characterized by having a single energy level, denoted with E_L, placed in between two contacts (namely the source and the drain). Moreover it is assumed to have a strong coupling between the two contacts and the single-level quantum dot. This means that a strong chemical interaction is present, e.g. a covalent bond. Consequently electrons have no difficulty in passing from contacts to dot and vice versa. Indeed the electron wave-function at contact-dot interface (because of the presence of the chemical bond) envelopes both atoms in the contacts and in the dot/molecule, such that the electrons populating that orbitals have high probability of being localized both in the dot and in the contacts, making easy the passage between the two. If not so, it is said that the dot is weakly coupled to the contacts. The latter case corresponds to wave-functions that are not delocalized, thus making difficult for the electrons being exchanged between the contacts and the dot (since they are localized either inside the dot or the contact). In this case the electrons are required to "jump" from the contacts to the dot and vice versa in order to have an electrical current. This is possible whenever the electrons change their state for some reason and go in another state, that in terms of wave-functions means: the electrons "jump" from a wave-function to another. These physical processes are referred as "sequential tunneling" (or hopping - see also chapter 20), in the sense that two completely uncorrelated tunneling phenomena are required to make the electrons "jump" from the source to the dot and then from the dot to the drain. Nevertheless, as mentioned above, it is now assumed that a strong coupling occurs between contacts and dot such that the electrons can easily be transferred between the two systems without changing state. This corresponds to have "coherent tunneling", i.e. a single tunneling process that allows electron to flow from source to drain passing through the dot. It is possible to think to this single tunneling mechanism as a single plane wave that "carries" the electron from the

source to the drain passing through the dot (indeed within the contacts -periodic structures- Bloch waves exist, that are not much different from plane waves in principle).
It is assumed to have an applied voltage V_{DS}, and that the number of electrons that the single energy level E_L of the dot is hosting is equal to N (notice that N can be only 0, 1, or 2). Notice that (since the contacts are assumed to be in equilibrium -due to their huge dimension when compared to the dot- the Fermi levels within the contacts are constant and differ of quantity $-qV_{DS}$). The energy level E_L is assumed to be within the so called bias window, i.e. in between the drain Fermi level E_{FD} and the source one E_{FS}.
Because of charge conservation at steady state the two net fluxes of electrons from the source to the dot $\Phi_S = \Phi_{S,dot} - \Phi_{dot,S}$ and from the drain to the dot $\Phi_D = \Phi_{D,dot} - \Phi_{dot,D}$ should compensate, giving rise to (see also figure 19.1):

$$\Phi_S = -\Phi_D \quad \Leftrightarrow \quad \Phi_{S,dot} - \Phi_{dot,S} = -\left(\Phi_{D,dot} - \Phi_{dot,D}\right)$$

Notice that charge conservation is in general a consequence of the matter conservation, indeed the previous equation corresponds in saying that all the electrons that enter the dot from the source, exit from the dot toward the drain (no electron is created or annihilated within the dot); that is in agreement with the assumption of having coherent tunneling from source to drain.

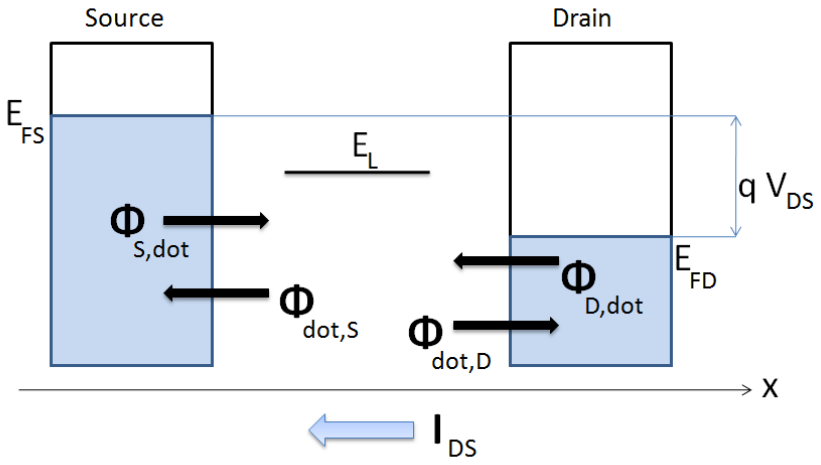


Figure 19.1: Electron transport through a single level quantum dot. Four electron fluxes are considered: $\Phi_{S,dot}$ from source to dot, $\Phi_{dot,S}$ from dot to source, $\Phi_{D,dot}$ from drain to dot, $\Phi_{dot,D}$ from dot to drain. Two net fluxes are then defined: the net flux from source to dot Φ_S, and the net flux from drain to dot Φ_D. The x-axis is from source to drain, while the drain to source current is conventionally flowing from drain to source.

In order to write explicit expressions for the four fluxes of figure 19.1 one more concept is needed. As mentioned it is supposed to have strong coupling between the contacts and the dot. The quality of the contact-dot interfaces is represented by means of the contact intrinsic times. The contact intrinsic times can be defined as the average amount of time required to an electron to move from the contact (the source or the drain) into the dot or vice versa to move from the dot into the source/drain contact. They are indicated with τ_S (referred to source) and τ_D (referred to drain). The smaller is the intrinsic time the stronger is the coupling between that contact and the dot. Indeed a small intrinsic time indicates that the electrons can be easily moved between the contact and the dot, indicating a strong coupling. In addition it is possible to define two energies: the so called contact coupling factors as follows:

$$\gamma_S = \frac{\hbar}{\tau_S} \quad , \quad \gamma_D = \frac{\hbar}{\tau_D}$$

that are the source coupling factor γ_S and the drain one γ_D. Notice that they have dimensions of energy since they are equal the reduced Planck's constant over a time. As mentioned, the lower is the intrinsic time the stronger is the coupling to the contact, that in terms of coupling factors becomes: the greater is the coupling factor the stronger is indeed the coupling between the contact and the dot.
An important remark is due concerning these two concepts. The intrinsic time can be seen as the average time interval needed to an electron to "jump" from a contact into the dot. But as mentioned it can analogously be seen as the average amount of time needed to an electron to "jump" from the dot into the contact. Considering the latter statement the intrinsic time acquires the meaning of average lifetime of the electrons inside the dot. Indeed if every τ_S (or τ_D) an electron escape from the dot to go into the source (or drain) contact, then from the standpoint of the dot every τ_S (or τ_D) an electron disappears (annihilates). In this optics the coupling factors γ_S and γ_D can be interpreted as an uncertainty on the energy of such electrons. To this purpose see also section 17.2, in which the time-energy uncertainty relation $\Delta E \Delta t \sim \hbar$ was discussed. Indeed the fact that an electron can stay within the dot (thus in the energy level E_L, or better in the state with energy E_L) only for an (average) amount of time τ_S (or τ_D) means that the exact value of the energy level of that state has an uncertainty (a standard deviation) equal to $\gamma_S = \hbar/\tau_S$ (or: $\gamma_D = \hbar/\tau_D$). If instead the uncertainty on the energy eigenvalue were null (no uncertainty on the energy level thus it is exactly known) then the lifetime should be infinite, that implies no exchange of electrons between the contacts and the dot (isolated dot, with $\gamma_{S,D} = 0$ and $\tau_{S,D} \to \infty$). An uncertainty on the energy level means a broadening of the energy level, that is no more known exactly but only within a range of values centered in E_L (standard value / variance). And since the stronger is the coupling, i.e. the stronger is the chemical bond, between the contact and the dot, the greater is the coupling factor, a strong coupling means also more broadening of the dot quantized energy levels. These topics will addressed again later.
At this point it is possible to write the expressions for the electron fluxes of figure 19.1. In general a flux of charged particles (electrons in this case) is given by the number of particles that pass through a given cross-section area (contact area) in the unit time. Here the flux can be considered as the number of electrons that pass from the contact to the dot (or vice versa) in the entire area (thus not per unit area, but like it was already integrated over the cross-sections) in the unit time. Thus the flux of electrons from the source to the dot is given by the number of electrons that pass from the source to the dot over the source intrinsic time τ_S. the number of electrons that with a unique tunneling process (i.e. coherent tunneling process) can pass from the source to the dot is given by two times (spin degeneracy) the Fermi-Dirac function referred to the source Fermi level E_{FS} evaluated at energy level E_L: $f_{FD}(E_L, E_{FS})$. Indeed two is the maximum number of electrons that can be hosted in E_L in the dot, and thus is the maximum number of electrons that can flow from the source toward the dot. The Fermi function evaluated for energy equal to E_L provides instead the probability of occupation of the energy level E_L within the source (since it is supposed coherent tunneling this is also the probability of having tunneling from source to the dot). The time in which an electron on average "jumps" from the source to the dot is τ_S and thus it is the unit time. Thus:

$$\Phi_{S,dot} = \frac{2 f_{FD}(E_L, E_{FS})}{\tau_S} = 2\frac{\gamma_S}{\hbar} f_{FD}(E_L, E_{FS})$$

Instead the flux of electrons from the dot toward the source is given by the number of electrons hosted in the dot N over the average time at which electrons pass from the dot to the source, τ_S:

$$\Phi_{dot,S} = \frac{N}{\tau_S} = \frac{\gamma_S}{\hbar} N$$

Analogous reasoning for the drain contact lead to:

$$\Phi_{D,dot} = \frac{2 f_{FD}(E_L, E_{FD})}{\tau_D} = 2\frac{\gamma_D}{\hbar} f_{FD}(E_L, E_{FD}) \quad , \quad \Phi_{dot,D} = \frac{N}{\tau_D} = \frac{\gamma_D}{\hbar} N$$

The net fluxes are defined as incoming fluxes (figure 19.1), thus they are given by:

$$\Phi_S = \Phi_{S,dot} - \Phi_{dot,S} \quad , \quad \Phi_D = \Phi_{D,dot} - \Phi_{dot,D}$$

At steady state the sum of the two net fluxes must be zero, because of charge (or matter) conservation, as stated previously. Thus:

$$\Phi_S + \Phi_D = 0 \quad \Leftrightarrow \quad \Phi_{S,dot} - \Phi_{dot,S} + \Phi_{D,dot} - \Phi_{dot,D} = 0$$

From which it is possible to find the number of electrons N in the energy level E_L:

$$\begin{aligned} & \Phi_{S,dot} - \Phi_{dot,S} = -\left(\Phi_{D,dot} - \Phi_{dot,D}\right) \\ \rightarrow \quad & \frac{\gamma_S}{\hbar}\left[2f_{FD}(E_L, E_{FS}) - N\right] = -\frac{\gamma_D}{\hbar}\left[2f_{FD}(E_L, E_{FD}) - N\right] \\ \rightarrow \quad & N = \frac{2}{\gamma_S + \gamma_D}\left[\gamma_S f_{FD}(E_L, E_{FS}) + \gamma_D f_{FD}(E_L, E_{FD})\right] \end{aligned} \tag{19.1}$$

Once that N is known it is possible to write an expression for the electrical current flowing from drain to source. Since the net flux of electron from source to dot Φ_S was already calculated as the flux of electrons over the entire cross section (it is already the actual number of electrons flowing from the source contact into the dot and then toward the drain), the electrical current can be written by directly multiplying it by the electron charge $-q$. At steady state the drain to source current is (the first minus sign is due to the fact that x-axis goes from source to drain while the current from drain to source):

$$I_{DS} = -(-q)\Phi_S = +q\left[2\frac{\gamma_S}{\hbar}f_{FD}(E_L, E_{FS}) - \frac{\gamma_S}{\hbar}N\right]$$

and considering the above expression for N:

$$\begin{aligned} I_{DS} &= \frac{q}{\hbar}\gamma_S\left[2f_{FD}(E_L, E_{FS}) - \frac{2\gamma_S}{\gamma_S + \gamma_D}f_{FD}(E_L, E_{FS}) - \frac{2\gamma_D}{\gamma_S + \gamma_D}f_{FD}(E_L, E_{FD})\right] = \\ &= \frac{q}{\hbar}\frac{2\gamma_S}{\gamma_S + \gamma_D}\left[\gamma_S f_{FD}(E_L, E_{FS}) + \gamma_D f_{FD}(E_L, E_{FS}) \;+\right. \\ &- \left.\gamma_S f_{FD}(E_L, E_{FS}) - \gamma_D f_{FD}(E_L, E_{FD})\right] \;= \\ &= \frac{q}{\hbar}\frac{2\gamma_S\gamma_D}{\gamma_S + \gamma_D}\left[f_{FD}(E_L, E_{FS}) - f_{FD}(E_L, E_{FD})\right] \end{aligned}$$

In conclusion the electrical current through a single energy level quantum dot is given by:

$$I_{DS} = \frac{q}{\hbar}\frac{2\gamma_S\gamma_D}{\gamma_S + \gamma_D}\left[f_{FD}(E_L, E_{FS}) - f_{FD}(E_L, E_{FD})\right] \tag{19.2}$$

This expression holds for any temperature $T \geq 0$ K but it is not general at all. Indeed it was mentioned before that the meaning of the coupling factors $\gamma_{S,D}$ is to broaden the energy levels, since they introduce an uncertainty over the energy level value. This issue will be discussed in a while, in the next section 19.1.1. Before discussing it a last remark is provided.

One may ask when the single level quantum dot conduct. In order to answer this question let's consider three cases:

1. In the first case the energy level E_L within the dot is supposed to be placed well above both the source Fermi level E_{FS} and the drain Fermi level E_{FD}. In this case both the Fermi-Dirac functions $f_{FD}(E_L, E_{FS})$ and $f_{FD}(E_L, E_{FD})$ will be essentially null, leading (see eq.(19.2)) to a null I_{DS}. Indeed in this case there are no occupied states in the source, thus no electron can flow into the dot (and proceed up to the drain). In this case $N \sim 0$ (empty) and no electrical current flows.

2. In the second case the energy level E_L within the dot is supposed to be placed well below both the source Fermi level E_{FS} and the drain Fermi level E_{FD}. In this case both the Fermi-Dirac functions $f_{FD}(E_L, E_{FS})$ and $f_{FD}(E_L, E_{FD})$ will be essentially equal to 1, leading (see eq.(19.2)) to a null I_{DS}. Indeed in this case there are only occupied states both in the source and in the drain. Consequently there will be no possibility for the electrons to flow in free states (since there are no free states) within the dot or in the drain. In this case $N \sim 2$ (occupied) and no electrical current flows.
3. In the last case the energy level E_L within the dot is supposed to be placed in between the source and the drain Fermi levels, like depicted in figure 19.1. In this case the source has occupied states at energy E_L, thus $f_{FD}(E_L, E_{FS}) \sim 1$, while the drain has unoccupied states at the energy E_L, thus $f_{FD}(E_L, E_{FD}) \sim 0$. From eq. (19.2) there is a non-null electrical current I_{DS}, and it is said that the energy level E_L within the dot is inside the bias window, where the bias window indicates exactly the energy interval qV_{DS} that separates the source Fermi level from the drain one. In this case there is conduction through the energy level E_L. Notice that if $f_{FD}(E_L, E_{FS}) \sim 1$ and $f_{FD}(E_L, E_{FD}) \sim 0$ the current saturates at the saturation value for the level within the bias window, that is:

$$I_{DS} = \frac{q}{\hbar} \frac{2\gamma_S \gamma_D}{\gamma_S + \gamma_D}$$

This is a general feature of transport through each 0D, but also 1D, 2D and 3D system: the electron states that take part to the conduction mechanism are only those in between the source and the drain Fermi levels. Otherwise, if no electron state is present in that energy interval, no electrical current originates.

19.1.1 Levels broadening

It was already mentioned in the previous section that the strong coupling between contacts and dot implies that the energy levels inside the dot are broaden w.r.t. the isolated dot. The physical origin of this phenomenon is indeed in the chemical bond that is generated between the dot and the contacts, that allows the electrons to easily move from the contacts to the dot and vice versa. In other words the energy-time uncertainty principle holds true and since the lifetime of an electron within the dot is limited (since it can escape from it), then an uncertainty on its energy eigenvalue is present. Again, it is possible to see the same physical phenomenon from another perspective. Let's consider the general solution of the time-dependent Schrödinger's equation, i.e. eq. (17.49) in section 17.3.5; that is reported here for clarity:

$$\psi(\vec{r}; t) = \sum\!\!\!\!\!\!\int dE\ C(E)\ \Psi_E(\vec{r})\ e^{-\frac{i}{\hbar}Et}$$

The steady state with an infinite lifetime have essentially no time dependence, since the exponential is a complex exponential (this feature was already discussed in section 17.3.5 and 17.3.6). Nevertheless in the present case it was said that the electrons can escape from the dot, giving rise to a finite lifetime. Thus the time dependence could be given by an exponential decay in time, indeed from the general solution of the Schrödinger's equation it is known that the time-dependent solution (factorized) should be:

$$\phi(t) = e^{-\frac{i}{\hbar}Et}$$

and if a finite lifetime is considered it is intuitive thinking that the exponential could become real, giving rise to the exponential decay in time, meaning that the (position) probability of finding the electron in such a level decreases with time. Indeed this is exactly what happens, and said τ the electron lifetime within the dot, the time dependence is of kind:

$$\phi(t) = e^{-t/\tau} = e^{-\frac{\gamma}{\hbar}t}$$

A full demonstration of this equation will provided in the next chapter (see section **??**); nevertheless what happens in doing the calculations is exactly that an imaginary part of the

energy comes out, representing the exponential decay of the previous equation. In particular a complex energy $E = E_L - i\gamma$ is considered, and when substituted into $\phi(t) = e^{-\frac{i}{\hbar}Et}$ the following time-dependence is found:

$$\phi(t) = e^{-\frac{i}{\hbar}E_L t} e^{-\frac{\gamma}{\hbar}t} = e^{-\frac{i}{\hbar}E_L t} e^{-t/\tau}$$

Notice that the energy as physical observable will be always a real number. So one may ask what is the meaning of a complex energy, is it correct to consider it? The answer is yes. The energy will be always real when measured. The imaginary part of energy is thus only a model, a mathematical artifact for representing the effects of the contacts. Indeed if one consider all the entire system molecule plus contacts, there is no reason of using a complex energy, since the electrons do not disappear from the "dot + contacts" system, but simply move through it, flowing from the dot towards the contacts or vice versa. Nevertheless, in analogy with conventional electronic devices modeling, it is usually preferred to make calculations only on the isolated device, and to consider the effects on the contacts by means of suitable boundary conditions. This approach (that will be widely investigated in the next chapter) leads to significant simplifications in the analytical treatment and in practical calculations (as it will be explained). The way of modeling the fact that the electrons can escape from the device, without explicitly considering the contacts is exactly by means of an imaginary part of the energy, that can be thus intended as a boundary condition that represents the fact that the dot is not isolated but it is in between the two contacts, and the electrons can escape from the dot (open system). Indeed, as already clarified above, introducing an imaginary part of the energy allows to write the general solution of the Schrödinger's equation as multiplied by a real exponential in time, i.e. a finite lifetime state, like the physics intuitively tells us. The time constant of the finite lifetime state is called lifetime, and in the previous treatment is exactly the intrinsic time τ_S for the source (or τ_D for drain). Thus modeling the energy as a complex entity has the advantage of making the calculations only on the quantum dot (the active portion of the device), but considering as it is in the real world: in between the two contacts. From the standpoint of the device this means that the electron can "disappear" (annihilate) after a time τ, and the probability of finding them in the dot is exponentially decaying with time. An infinite lifetime state has a lifetime that is infinite: $\tau \to \infty$, and it is possible to exactly know its energy, that previously was said to be E_L. Thus the density of states per unit energy of a infinite lifetime state will be a delta Dirac, in energy domain, centered in E_L. Instead in the case of a finite lifetime state, there is an uncertainty on the exact value of the energy of the electrons in such a state. Thus the density of state per unit energy will be broadened (it is only qualitatively illustrated in figure 19.2). In practice, whenever an exponential decay is considered in time domain a Lorentzian distribution is recovered in energy domain. Thus the broadened density of states per unit energy will have a Lorentzian shape, of the kind:

$$D_{EL}(E) = \frac{\gamma/(2\pi)}{(E - E_L)^2 + (\frac{\gamma}{2})^2} \tag{19.3}$$

where $\gamma = \gamma_S + \gamma_D$ and E_L is the energy level in the dot. Notice that it is possible to demonstrate the previous expression. In order to do that one should make the Fourier transform of the electron wave-function, considering the (real) decaying exponential in time, and then normalize it to 1 [45]. Notice also that the Fourier transform w.r.t. to time corresponds to the following correspondence of variable (between time variable t and angular frequency ω):

$$t \quad \overset{\mathfrak{F}}{\longleftrightarrow} \quad \omega = 2\pi f = 2\pi \frac{E}{h} = \frac{E}{\hbar}$$

From which it is clear that by Fourier transforming a wave-function that depends on time corresponds in going into energy domain, that is where the density of states per unit energy is defined (further details on this will be clarified in the next chapter). Notice that it is

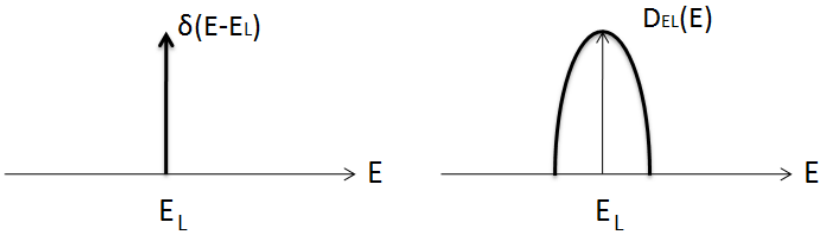


Figure 19.2: Qualitative pictorial representation of broadening of an energy level. Usually the broadening assumes a Lorentzian shape, thus $D_{EL}(E)$ is a Lorentzian distribution.

said that the time and energy domain are isomorphous (see also section 17.3.3), since they are biunivocally linked by means of linear application (the Fourier transform).
A measure of the level broadening is provided by $\gamma = \gamma_S + \gamma_D$, indeed notice that if in equation (19.3) it is made the limit for $\gamma \to 0$ then the denominator goes to zero more rapidly than the numerator (squared γ is present), such that the remaining function tends to infinite when $E \to E_L$ (it is essentially a delta Dirac centered in E_L).
Notice that the integral over all the energy axis of equation (19.3) is again equal to 2 (due to spin degeneracy); indeed the integral w.r.t. energy of a density of states per unit energy should provide the number of states; and since $D_{EL}(E)$ is the effect of the broadening of a single energy level E_L then it should again host two electrons (since it is a single electron state with spin degeneracy).
In order to consider the broadening in the previous expression for the electrical current through the molecular/dot wire, equation (19.2) should be modified as follows:

$$I_{DS} = \frac{q}{\hbar}\frac{2\gamma_S\gamma_D}{\gamma_S+\gamma_D}\int_{-\infty}^{+\infty} D_{EL}(E)\left[f_{FD}(E,E_{FS}) - f_{FD}(E,E_{FD})\right]dE \tag{19.4}$$

in which the integral over all the energies is aimed in recovering the full probability of occupation of the broadened level E_L. Notice again that the stronger is the coupling between the contacts, i.e. the greater is γ, the wider is $D_{EL}(E)$.
If it is defined the transmission functions as:

$$T(E) = 2\pi\frac{\gamma_S\gamma_D}{\gamma_S+\gamma_D}D_{EL}(E) \tag{19.5}$$

then the las equation becomes:

$$I_{DS} = \frac{2q}{h}\int_{-\infty}^{+\infty} T(E)\left[f_{FD}(E,E_{FS}) - f_{FD}(E,E_{FD})\right]dE \tag{19.6}$$

19.1.2 Multi-level quantum dot

Let's now consider the case of a two-levels quantum dot. First of all a remark concerning the position of the Fermi level of the dot w.r.t. source and drain Fermi levels. In the moment in which the molecular wire is created a (small) amount of charge exchanging between the contacts and the dot occurs such that after a given time the thermodynamic equilibrium is reached. This means that the Fermi levels in the source, the drain and the dot are all aligned. In the moment in which there is an applied bias one may ask where the dot Fermi level is positioned w.r.t. to Fermi levels of source and drain contacts. The answer is not easy in general, and some *ab initio* calculations, regarding the entire device (dot/molecule plus contacts), should be performed to answer it. Nevertheless if it assumed

of having similar coupling factors $\gamma_S \cong \gamma_D$ then it is quite probable of having the Fermi level of the dot in the middle between the source and drain ones. This is in general an assumption, and for the moment no further discussion is provided. This topic is again considered under a different perspective in section 19.1.4.
Thus for the moment the considered case is the one of a molecular quantum dot with two energy levels, namely the HOMO (Highest Occupied Molecular Orbital) and the LUMO (Lowest Unoccupied Molecular Orbital), respectively placed below and above the dot Fermi level, that is supposed to be in the middle between the source and drain Fermi levels as mentioned above. The situation is summarized in figure 19.3. The relative positions of the LUMO and the HOMO w.r.t. the dot Fermi level are supposed to be known from electronic structure calculations (e.g. performed by means of the methods discussed in chapter 18).

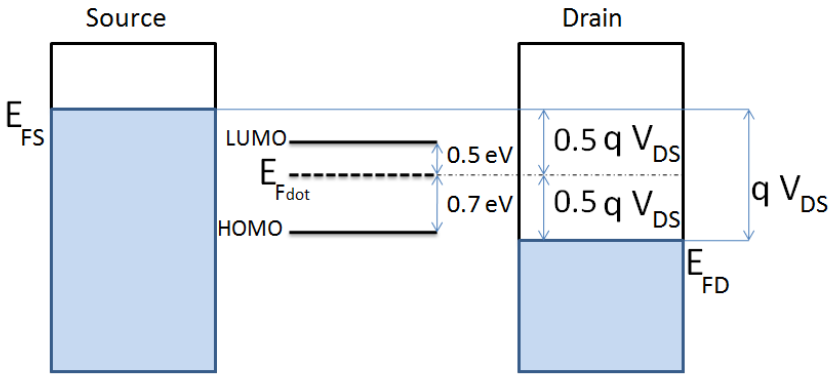


Figure 19.3: Qualitative representation of a two level quantum dot. The distances between the Fermi level in the dot E_{Fdot} and the two levels (HOMO and LUMO) are supposed known from electronic structure calculations.

Let's start by considering the thermodynamic equilibrium case. As mentioned above in this case all the Fermi levels are aligned. In the moment in which a small V_{DS} is applied neither the LUMO nor the HOMO are inside the bias window, thus the electrical current is null. Instead when the bias window reaches the value of 1 eV (i.e. $V_{DS} = 1$ V) then the LUMO level starts conducting. This because the applied bias drops half between the source and the dot and half between the dot and the drain. Indeed it was supposed to have the dot Fermi level in the middle between the source and drain ones. Consequently a voltage equal to two times the difference between the LUMO and the Fermi level of the dot should be applied in order to have the LUMO within the bias window and thus an electrical current flowing from drain to source. If the applied voltage is further increased then when it reaches the 1.4 V the HOMO level starts conducting, again this is a consequence of the fact that it was assumed that the applied voltage drops equally from source to dot and from dot to drain. In summary, without broadening, the current-voltage characteristics will be similar to the one of figure 19.4 (dashed line). If broadening is considered then the solid line of figure 19.4 will be recovered. Indeed in the picture of figure 19.3 the energy is on ordinate axis and the effect of broadening is to make the levels entering before and exiting later from the bias windows, thus leading to a smoother characteristics. The greater the broadening the smoother (i.e. more linear) will be the current-voltage characteristics.
The important point is that in this oversimplified model for conduction through quantum dots the two energy levels contribute to conduction independently when they enter in the bias window. Thus, under these hypotheses, the extension of equation (19.4) to a generic number of discrete energy levels within the quantum dot is trivial: indeed it is sufficient to superimpose their effects by summing their density of states. This is exactly what is done with the following equation, in which the multi-level transmission function is defined as the

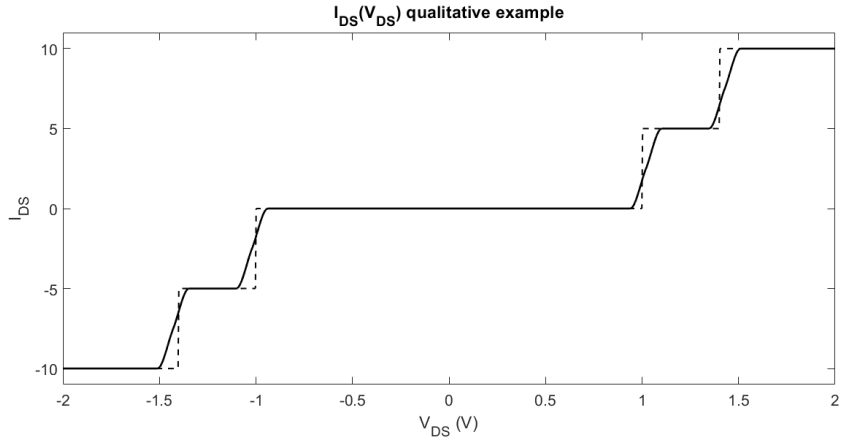


Figure 19.4: Qualitative example for the two level quantum dot of figure 19.3. The LUMO level causes the first plateau at ±1 V while the HOMO the second at ±1.4 V. Notice the conductance gap around the zero bias condition (0 V), and notice that it is equal to the double (for both positive and negative biases) of the energy interval that separates the closest dot energy level to the dot Fermi level.

sum over all the *i-th* levels in the dot:

$$T(E) = \sum_i 2\pi \frac{\gamma_{Si}\ \gamma_{Di}}{\gamma_{Si} + \gamma_{Di}} D_{ELi}(E) \tag{19.7}$$

where different coupling factors γ_{Si}, γ_{Di} are in general possible for the different energy levels (it is not said that all the energy levels -i.e. all the dot orbitals- have the same coupling with the contacts -i.e. wave-function hybridization or chemical bond). The final expression for the electrical current through a multi-level quantum dot, even in the case of broadening (and at whatever temperature, that is account with the Fermi-Dirac distributions) is the following:

$$I_{DS} = \frac{2q}{h} \int_{-\infty}^{+\infty} T(E) \left[f_{FD}(E, E_{FS}) - f_{FD}(E, E_{FD}) \right] dE \tag{19.8}$$

where the transmission spectrum (or function) $T(E)$ is defined by eq. (19.7). Notice that the last expression resembles eq. (19.6).

19.1.3 Quantum conductance

If in equation (19.8) it is assumed to have perfect transmission, i.e. $T(E) = 1$ for each energy value E, the following expression is obtained:

$$I_{DS} = \frac{2q}{h} \int_{-\infty}^{+\infty} \left[f_{FD}(E, E_{FS}) - f_{FD}(E, E_{FD}) \right] dE$$

Moreover by supposing of being at zero kelvin for simplicity, the two Fermi functions are 1 below the source and drain Fermi levels respectively while they are 0 above. Thus the integral becomes simply the difference between the contact Fermi levels:

$$I_{DS} = \frac{2q}{h} \left(E_{FS} - E_{FD} \right) = \frac{2q^2}{h} V_{DS}$$

where it was exploited the bias window definition: $E_{FS} - E_{FD} = -qV_{SD} = +qV_{DS}$.
Notice that even if the transmission is unitary, i.e. all electrons are always transmitted by hypothesis through the quantum dot, the electrical current presents a finite slope, i.e. a well defined (and finite) conductance, that is (the factor two arose from spin degeneracy):

$$G_Q = \frac{2q^2}{h}$$

This conductance value is called "quantum conductance" or "contact conductance". It arises from the interface between the quantum dot and the contacts, indeed it is present even if a full transmission is supposed. The contact conductance limits the maximum slope of the current that cannot be infinite as predicted with classical physics for a conductor length that tends to zero. To this purpose remember that the classical (ohmic) conductance is: $G = \sigma W/L$, thus for a small conductor ($L \to 0$) it is expected a huge conductance, especially if transmission is unitary: $T(E) = 1$. Nevertheless the actual conductance has an upper limit, the so called quantum limit to conductance, that is indeed given by the quantum conductance value: $G_Q = \frac{2q^2}{h} \sim (12.49\text{k}\Omega)^{-1}$. Early experiments on mesoscopic systems (see also section 19.2 for a precise definition of mesoscopic systems), i.e. small conductors, underlined that the actual conductance was approaching to a limiting value, that was indeed G_Q. Notice that the physical origin of the quantum conductance, despite the name, has no an intimate relation with quantum mechanics [91]. Indeed it arises from the interface between the contact and the quantum dot (or the small conductor). Within the contact the current is carried by a huge number of possible electron states, or contact proper modes. Indeed a contact is usually a metal contact, and it presents allowed energy bands, composed by a set of continuum states, i.e. a continuum infinite number of states/modes. Instead within the quantum dot the current can flow only through the allowed discrete (and few!) electron states. This fact requires a redistribution of the current among the current-carrying modes/states leading to an interface resistance (the dot is like a sort of "bottleneck") [91]. It is possible to get rid of this resistance by making the contacts identical to the channel (i.e. the quantum dot). Nevertheless this has no sense because the contacts are by definition a large reservoir of electrons, and they should be "infinitely" more conducting than the conductor in order to justify the assumption that the applied voltage drops entirely across the channel.

19.1.4 Electrostatic source and drain capacitances

In the moment in which there is a non null applied voltage conventional electrostatic capacitances should be considered. Indeed the movement of charges in response to the applied voltages would lead to capacitive effects. In total there are two effects on the channel potential that are linked to electrostatic capacitances:

- the electrostatic effect: quite conventional, it follows from the capacitance definition, i.e. from the fact that an applied bias would produce a variation in the channel (i.e. in the dot) charge.
- the charging effect: it is an effect that in general occurs also in conventional electronic devices, but is so small that it is in general negligible. Instead in nano-devices it is comparable to the electrostatic effect and must be considered in the system modeling. It is the subject of the next section.

Notice that in general the electrostatic effect and the charging effect are linked, as will be better clarified in the next section, and a self-consistent loop is necessary to ensure convergence.

Once the source and drain electrostatic capacitances are supposed known, the simple capacitive model of figure 19.5 can be considered. This is a linear (or better linearized) model, and conventional circuit theory can be used to find the total channel potential within the dot. Since it is a linearized model it holds only for small variations, or analogously small V_{DS} values. If not so a non-linear model should be considered, indeed the actual values of C_S and C_D are dependent on the applied voltage.

The equivalent circuit (figure 19.5 - right) is a simple capacitive divider, and since C_S and C_D are supposed known it is possible to easily estimate the effect of an applied V_{DS} on the channel potential. Notice that this model make also the additional assumption of having a zero-dimensional quantum dot, i.e. the physical extension of the quantum dot is completely neglected. Consequently it is also neglected the potential shape within the dot

(that instead in general varies from point to point accordingly to the Laplace equation). And notice also that the source potential is taken as reference potential.

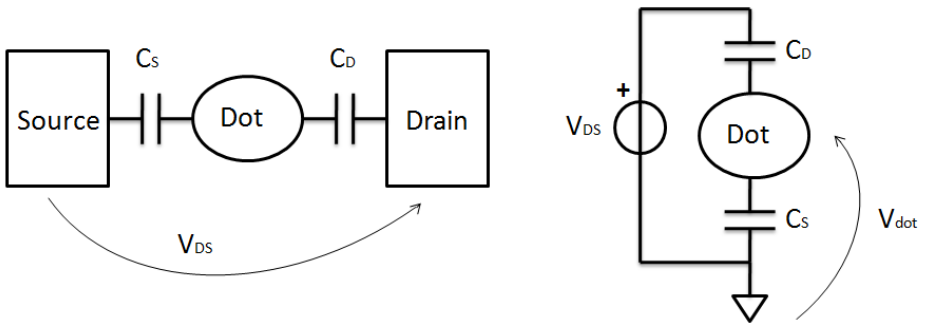


Figure 19.5: Capacitive (linearized) model for an ideal quantum dot with no physical extension (zero dimensions). Left: a representation of the molecular/dot wire physical structure; right: the equivalent linearized circuit.

The consequence of an applied V_{DS} voltage is the one of shifting the quantum dot energy levels. In particular the quantum dot Fermi level will be no more the one at thermodynamic equilibrium but instead it will be shifted in energy of the amount $U_{Vds} = -qV_{dot}$, where from the capacitive divider:

$$V_{dot} = V_{DS}\frac{C_D}{C_S + C_D} \quad \rightarrow \quad U_{Vds} = -qV_{dot} = -q\frac{C_D}{C_S + C_D}V_{DS}$$

In conclusion, the effect of an applied V_{DS} will be to create a bias window $-qV_{DS}$ as explained previously, but also to shift the quantum dot energy levels (the quantum dot Fermi level) of the amount U_{Vds}. Usually the capacitive ratio is called "voltage division factor", and indicated with:

$$\eta = \frac{C_D}{C_S + C_D} = \frac{C_D}{C_{ES}}$$

where C_{ES} is the total electrostatic capacitance, defined as: $C_{ES} = C_S + C_D$. Notice that if $C_S = C_D$ then $\eta = \frac{1}{2}$, that is exactly what was supposed in the previous sections by saying that the dot Fermi level was assumed to be in the middle between the source and drain Fermi levels. Indeed $\eta = \frac{1}{2}$ implies that the Fermi level of the dot is $1/2qV_{DS}$ higher than the source potential (that is the reference), and $1/2qV_{DS}$ lower than the drain potential.

19.1.5 Charging effect

The charging effect arises from the motion of charged particles, i.e. electrons, toward/from the channel. In particular the amount of charge that is moved toward/from the channel alters the energy levels in the channel itself. In general the charging effects occurs also in conventional macroscopic electronic devices, where the presence of an extra electron can alter somehow the band diagram. Nevertheless in such devices the charging effect is so small that in general it results negligible. Instead in nano-devices it becomes comparable to the electrostatic effects and thus it must be considered in the system modeling. Indeed the perturbation of a quantum dot energy levels due to the presence of extra electrons can be large, of the order of the eV, and thus comparable to the effects of the applied bias.
In response to a variation in the applied voltage, a given amount of charge $Q_n = -q\Delta N$ is moved toward/from the channel, accordingly with the total electrostatic capacitance

C_{ES} (see also previous section). Notice that ΔN indicates the variation of the number of electrons inside the dot due to the applied bias, while the corresponding total charge is Q_n. As a consequence a voltage is established across the total capacitance of the system, indeed it is the result of the charge transferred from the contact(s) into the dot and the result affects the entire system:

$$\Delta V_{charging} = \frac{\Delta Q}{C_{tot}} = \frac{Q_n}{C_{ES}} = \frac{-q\Delta N}{C_S + C_D}$$

The potential energy corresponding to $\Delta V_{charging}$ is called "charging energy", and it is equal to:

$$U_{charging} = -q\Delta V_{charging} = \frac{q^2}{C_{ES}}\Delta N \tag{19.9}$$

This potential contribution corresponds to the potential energy within the dot that is due to the transfer (through $C_S + C_D$) of a given number of extra electrons $\Delta N = N - N_0$, where N indicates the current number of electrons within the dot and N_0 the number of electrons in the dot at thermodynamic equilibrium. If a single electron is transferred toward the dot then $\Delta N = N - N_0 = 1$, that means a single extra electron is present, and the charging energy is:

$$U_{charging\ 1\ eln} = \frac{q^2}{C_{ES}}\Delta N = \frac{q^2}{C_{ES}} = U_0$$

Consequently it is possible to rewrite the charging energy as: $U_{charging} = U_0\Delta N = U_0(N - N_0)$. Notice that ΔN can also be negative (meaning that the electrons are moving from the dot toward the contacts as a consequence of the applied bias).

The effect of the charge transferring toward/from the dot thus has again the effect of shifting the total potential of the dot of the amount $U_{charging}$; i.e. the Fermi level of the dot is shifted of $U_{charging}$. To this purpose it is possible to have a dynamic level shifting of the dot levels. Indeed, if, on average, an extra electron is present within the dot due to the external bias V_{DS}, then the Fermi level of the dot is shifted upward of the (positive) quantity U_0, that means that both the LUMO and the HOMO are shifted upward of U_0. It may also happen that due to the charging effect one level (e.g. the LUMO) that initially was inside the bias window (thus conducting), then is shifted so much that exits from the bias window, thus reducing the total current. This can give rise to negative differential resistances.

Notice that the contribution $U_{charging}$ can be superimposed (i.e. summed) to the electrostatic effect discussed in the previous section to get the total effect.

19.1.6 Self consistent algorithm and total electrostatic effect

The effect of an applied bias is to shift the dot energy levels of an amount $U_{Vds} = -q\frac{C_D}{C_S+C_D}V_{DS}$, and as a consequence of the applied bias a number of electrons N populates the dot, with $N \neq N_0$ in general. Thus a charging energy arises and it has the effect of shifting the dot energy levels of the amount $U_{charging}$. The total effect of the applied bias is thus a superposition of the two aforementioned effects (under the hypothesis of a linear, or better linearized, system - i.e. for small variations or small V_{DS} values):

$$U = U_{Vds} + U_{charging} = -q\frac{C_D}{C_S + C_D}V_{DS} + \frac{q^2}{C_{ES}}\Delta N$$

In order to account of these effect the potential U can be used to shift the transmission function $T(E)$, that becomes $T(E - U)$, thus leading to a total current:

$$I_{DS} = \frac{2q}{h}\int_{-\infty}^{+\infty} T(E-U)\left[f_{FD}(E, E_{FS}) - f_{FD}(E, E_{FD})\right] dE \tag{19.10}$$

from which, by remembering the definition of transmission spectrum in equations (19.7) and (19.5), it follows that the density of states $D(E)$ is shifted of the amount U:

$$D_{ELi}(E) \quad \rightarrow \quad D_{ELi}(E - U)$$

Nevertheless the density of states appears in the evaluation of the number of electrons within the dot N, that from eq. (19.1) is:

$$N_i = \frac{2}{\gamma_{Si} + \gamma_{Di}} \int_{-\infty}^{+\infty} D_{ELi}(E - U) \left[\gamma_{Si} f_{FD}(E, E_{FS}) + \gamma_{Di} f_{FD}(E, E_{FD})\right]$$

where the last expression is a straightforward generalization of the eq. (19.1) for the case in which broadening occurs in the level *i-th*. This holds for each energy level i of the dot (the total N is the sum of the various i since the relations are all linear: $N = \sum_i N_i$). The point is that U depends on N, because of $U_{charging}$ dependence from N, but N depends on U. Thus an iterative self-consistent procedure or algorithm must be used to evaluate the final effect of an applied bias.

19.1.7 Gating the quantum dot: a three terminals model

In the first chapter it was mentioned the possibility of realizing a molecular transistor, by adding a third electrostatically coupled gate terminal, see figure 1.9. In this section the capacitive model presented in the previous sections is suitably modified in order to account for the gating effect.
In figure 19.6 is represented a possible capacitive model including also the gate capacitance that arises from the device structure (again refer to figure 1.9). If the applied voltages V_{GS} and V_{DS}, and the electrostatic capacitances C_G, C_S and C_D are known, the resulting channel potential can be simply found from two capacitive dividers. Exploiting the superposition of effects (i.e. linearized system hypothesis), the total average channel potential (this is the average potential inside the quantum dot, that in this model is considered without physical dimensions) is given by:

$$U_{tot} = U_{Vgs} + U_{Vds} + U_{charging} \tag{19.11}$$

where the two contributions U_{Vgs} and U_{Vds} can be evaluated by superposition of the effects. For estimating U_{Vgs} it is possible to set $V_{DS} = 0$, thus C_S and C_D are in parallel (toward ground since the source is taken as reference) and the effect of an applied V_{GS} is thus the one of shifting the energy levels in the dot of the amount:

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

where $C_{ES} = C_G + C_S + C_D$ is the total electrostatic capacitance.

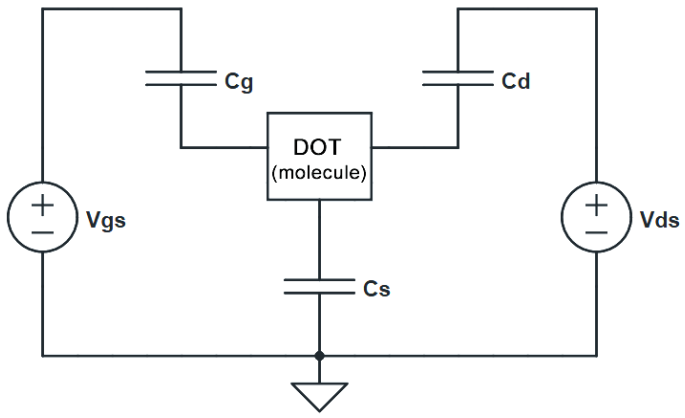


Figure 19.6: Quantum dot capacitive model. In this simplified model the molecular quantum dot -i.e. the channel- is assumed to be without physical extensions, i.e. zero-dimensional.

For estimating U_{Vds} it is possible to set $V_{GS} = 0$, thus C_G and C_S are in parallel and the effect of an applied V_{DS} is thus the one of shifting the energy levels in the dot of the amount:

$$U_{Vds} = -qV_{dot}\big|_{V_{DS}} = -q\frac{C_D}{C_D + C_G//C_S}V_{DS} = -q\frac{C_D}{C_G + C_S + C_D}V_{DS} = -q\frac{C_D}{C_{ES}}V_{DS}$$

where $C_{ES} = C_G + C_S + C_D$ is the total electrostatic capacitance. The charging effect is given by same expression of previous section, i.e. eq. (19.9), that is now reported for clarity:

$$U_{charging} = \frac{q^2}{C_{ES}}\Delta N \tag{19.12}$$

where $C_{ES} = C_G + C_S + C_D$ is the total electrostatic capacitance. Finally notice again that the set of equations presented so far needs to be solved self-consistently.

19.1.8 Quantum capacitance

The quantum capacitance represents the molecule states filling. The easiest way of introducing it (in my opinion) is the approach presented in [45]. In the moment in which the molecule-contact (source or drain contact) junction is created, a re-distribution of charges occurs between the molecule and the contacts and the thermodynamic equilibrium is achieved, with an alignment in the Fermi levels. A contact in order to be a "good" contact should have an enormous amount of states around the Fermi level, if compared to the channel ones, and thus it should be an unlimited reservoir of electron states for the channel. Under this approximation the contact Fermi level is essentially pinned. Instead the molecule/dot Fermi level is moved up or down in order to achieve the equilibrium. A small amount of charge can be already enough to produce a huge molecular Fermi level shift. Notice that even a fractional electron charge can be exchanged between the molecule and the contacts. Indeed if an electron is involved in a chemical bond between the contact and the dot/molecule (or the molecular anchoring group), the electron orbital (electron cloud) is partially in the contact and partially in the molecule, thus the electron is for a certain time localized inside the contact and for another time interval inside the molecule. On average, therefore, a charge less than q (less than an electron elementary charge) can be localized inside the molecule and considered as transferred charge in order to achieve the equilibrium. The amount of shift in the Fermi level depends on the amount of states in the dot around the Fermi level. In particular if a dot/molecule has a large density of states (per unit energy) around the Fermi level, the amount of Fermi level shift due to a given amount of charge transfer between the contact and the dot would be lesser than if the density of states would be small. The total number of electrons in the molecule is in general given by:

$$N = \int_{-\infty}^{+\infty} DOS(E) f_{FD}(E, E_F) dE \tag{19.13}$$

where $DOS(E)$ is the (total) density of states per unit energy of the molecule (corresponding to the sum over all energy levels of the $D_{ELi}(E)$ distributions), and $f_{FD}(E, E_F)$ the Fermi-Dirac distribution. At zero kelvin the amount of occupied states (i.e. the number of electrons) is:

$$N = \int_{-\infty}^{E_F} DOS(E) dE \tag{19.14}$$

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

$$\frac{dN}{dE} = DOS(E) \quad \Rightarrow \quad \left.\frac{dN}{dE}\right|_{E_F} = DOS(E_F) \tag{19.15}$$

and by supposing only small variations of Fermi energy (indicated with δE_F):

$$\delta E_F = \frac{\delta N}{DOS(E_F)} = \frac{q^2}{C_q}\delta N = \frac{q}{C_q}Q_n \tag{19.16}$$

where it is defined the quantum capacitance as: $C_q = q^2 DOS(E_F)$, and where the total channel charge is $Q_n = q\delta N$. The notation is analogous to the one of [45], where "δ" indicates a small variation.
Equation (19.16) holds only for small variations and it is an linearized expression, useful for determining the amount of charge transferred between the contacts and the molecule, for reaching the thermodynamic equilibrium, once the Fermi level shift is known (or vice versa to find δE_F if Q_n is known). From eq. (19.16) it is evident that the greater is the quantum capacitance, the smaller is δE_F for a given δn.
The concept of quantum capacitance can be generalized for any (small) variation of the total channel potential due to whatever effect, such as the application of a bias. The concept of quantum capacitance is simple (and strictly holds) for small deviations from equilibrium, i.e. for linearized models. In particular with the notation of [44] (see chapter 7.3 for a more formal introduction of these concepts), said N_0 the number of electrons in the molecule at equilibrium, N the number of electrons in the molecule outside the equilibrium, U_N the value of the channel potential such that $N = N_0$ and U the current value of the channel potential, the following relation holds:

$$\delta N = N - N_0 \approx C_q \frac{U_N - U}{q^2} = C_q \frac{\delta U}{q^2} \quad \Rightarrow \quad -q\delta N = Q_n \approx C_q \delta V_{CH} \tag{19.17}$$

where $\delta U = -q\delta V_{CH}$ is the (small) electrostatic potential variation that occurs in the channel, e.g. due to external bias (notice that in eq. (19.17) q^2 can be thought as $(-q)^2$). A full demonstration of eq. (19.17) is provided in [44] (chapter 7.3) and also in slightly different terms in [45]. Nevertheless it is reasonable thinking at eq. (19.17) like the generalization of eq. (19.16) in which the dot Fermi level E_F is substituted with the channel potential U.
In conclusion, the quantum capacitance is an additional contribution of capacitance that represents the state filling in a nanodevice. More precisely, it is in general a measure of the amount of charge (electrons) that can be transferred into (or from) the molecule in a given bias condition, indeed it is intimately linked to the density of states within the molecular channel or the quantum dot.
In general the quantum capacitance should be considered in molecular devices capacitive balance. In order to do that let's consider once again the expression for the total channel potential shift U, that is given by eq. (19.11) and reported here for simplicity:

$$U = U_{Vgs} + U_{Vds} + U_{charging} = -q\frac{C_G}{C_{ES}}V_{GS} - q\frac{C_D}{C_{ES}}V_{DS} + q^2\frac{\Delta N}{C_{ES}} \tag{19.18}$$

if a small variation of the number of electrons in the dot is considered, then $\Delta N \to \delta N$ and thus:

$$\Rightarrow U = -q\frac{C_G}{C_{ES}}V_{GS} - q\frac{C_D}{C_{ES}}V_{DS} + q^2\frac{\delta N}{C_{ES}} \tag{19.19}$$

Remembering the definition of quantum capacitance and equation (19.17), it is possible to rewrite eq. (19.19) as follows:

$$U = -q\frac{Cg}{C_{ES}}V_{GS} - q\frac{Cd}{C_{ES}}V_{DS} - \frac{C_q}{C_{ES}}U$$

from which:

$$U = -q\frac{Cg}{C_{ES} + C_q}V_{GS} - q\frac{Cd}{C_{ES} + C_q}V_{DS} \tag{19.20}$$

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

$$\delta U = -q\frac{Cg}{C_{ES}+C_q}\delta V_{GS} - q\frac{Cd}{C_{ES}+C_q}\delta V_{DS} \tag{19.21}$$

The important point is that the quantum capacitance C_q should be considered in the same way as the electrostatic capacitances in the capacitive balance of the entire system, indeed it appears in equations (19.20) and (19.21).
A final remark: in 2019 the first experimental evidence of the quantum capacitance was possible [231], thus highlighting its increasing importance as the devices are scaled.

19.1.9 Final considerations

The model introduced so far for the transport through molecular quantum dot is an oversimplified model. In this model the broadening was essentially introduced *ad hoc*, even if an attempt of short qualitative and intuitive discussion was provided. The broadening is an extremely important feature of transport in nano-devices and more importantly it is a natural consequence of the quantum mechanical physical and mathematical modeling of nano-transport, as will be better clarified in next chapter. Moreover in the multi-level picture it was stated that the levels participate to conduction independently the one from the other. This is not always at all correct, as will be clarified in the next chapter. More important is the fact that when more energy eigenvalues are considered within the quantum dot, more coupling factors arise. This suggest to collect them within a matrix, and this is exactly the standard approach in non-equilibrium Green's functions theory in which matrices are considered instead of scalars. A more formally correct and comprehensive of all the effects (also those introduced *ad hoc* here or not yet introduced) treatment is provided in the next chapter. It is a purely quantum mechanical model for transport and it is very general, indeed it holds in all transport regimes, comprising also incoherent transport.
Despite of its limits, this oversimplified model is capable of collecting the main physical insights of the transport at nano-scale. Indeed the current saturation or plateau whenever an entire level (with entire level is intended the broadened $D_{ELi}(E)$) is included in the bias window, and the conductance gap around the zero bias are well explained with this extremely simple model. Moreover, the origin of an asymmetric I-V characteristics is again explained if a voltage division factor η different from one half is taken into account. And again the possibility of a negative differential resistance arises naturally from the concept of having a channel energy level exiting from the bias window. Still the capacitive model is capable of catching many interesting insights in a straightforward manner, especially when also the quantum capacitance is taken into account.
Since the advantages seem to be more than the above discussed drawbacks, one may ask the reason for introducing a more refined model, that is also a little bit more involved in the formalism, instead of continuing using this one, eventually with some *ad hoc* corrections. The answer is that the physical phenomena are very rarely uncorrelated like in this "toy" model. For example, a subtle effect of the applied bias voltage is to modify the density of states within the quantum dot, and also its transmission function (or spectrum) $T(E)$. Transmission peaks due to different energy level can merge together, totally separate, modify in height, width, shape, entering or exiting in unforeseen ways, and all these effects can be correlated. Thus the model presented so far is of extreme importance because it allows to keep always in mind in a straightforward manner the physical intuition of what happens in a real case, but in order to make quantitatively correct previsions on actual transport behaviors of nano-devices a more refined model is needed. As already mentioned, it is the so called non-equilibrium Green's functions formalism (NEGF), that is the subject of the next chapter.

19.2 Mesoscopic systems

It was already mentioned that due to the small dimensions of molecules (of the order of the nanometer), a classical approach to transport cannot catch the essence of the transport features in such a system. Indeed the correct way of proceed should take into account quantum mechanics. In particular, molecular electronics modeling is based on the physical treatment of mesoscopic systems, i.e. that systems whose dimensions are intermediate between the microscopic and macroscopic ones [91]. Mesoscopic conductors are much larger than microscopic objects like atoms (of the order of tens of hundreds of picometers) but not large enough to be "ohmic" conductors [91]. Molecular devices, especially when the considered molecules are composed by some tens of atoms (and this is the typical case), belong to the class of mesoscopic system. The purpose of this section is to point out the features of such systems.

A conductor usually shows ohmic features if its dimensions are much larger than each of the following three characteristic length scales [91]:

- The de Broglie wavelength: it is related to the kinetic energy of the electrons, indeed it is $\lambda = \frac{h}{p} = \frac{h}{mv}$.
- The mean free path: it is the distance that an electron travels before its initial momentum is destroyed. It is the consequence of scattering phenomena, arising from interactions among electrons and phonons, etc...
- The phase-relaxation length: it is the distance that an electron travels before its initial phase is destroyed (think to the phase of the wave-function). Again it is due to interactions among electrons and phonons or between an electron and another electron. In the latter case (under the assumption that the scattering phenomenon between two electrons is elastic - i.e. any the momentum lost by one electron is picked up from the other) the main effect is affecting the phase-relaxation time and thus the phase-relaxation length without affecting the mean free path (the situation is different if a phonon is involved in the scattering process). Notice that rigid (i.e. static) scatterers (such as an impurity atom or a well defined molecular geometry) do not influence the phase-relaxation since the phase relation between electrons interacting or not with these static scattereres is constant in time leading to constant transmission coefficient and constant phase relation. Only dynamic scatterers (electron-electron interactions etc...) lead to phase-relaxation.

These three scale lengths vary widely from one material to another and they are also strongly affected by temperature (that means that a conductor can be ohmic or not depending on temperature), magnetic field, etc... For this reason mesoscopic features were observed in a wide range of dimensions of conductors (see [91] for details). Mesoscopic conductors can present interesting features such as negative (differential) resistance, and so on.

There are mainly two approaches in modeling transport through mesoscopic systems and they are the transmission formalism or Landauer-Bütticker formalism (briefly introduced in the next section 19.3) and the Non-Equilibrium Green's Function formalism (NEGF). The first one has the advantage of being based on trivial concepts, but it has the disadvantage of holding essentially only for ballistic transport (i.e. only for coherent transport). The latter instead has the advantage of being very general and holding also for non-coherent transport. Actually the real power of NEGF can be appreciated when incoherent transport is considered (see later - section **??**).

19.3 The transmission formalism

The purpose of this section is to provide a brief overview about the transmission formalism or the so called Landauer-Bütticker formalism. It is one of the two mainly used models for transport in mesoscopic systems. The other, namely the NEGF one, will be widely addressed in the next chapter.
In the transmission formalism the electrical current flowing through a mesoscopic conductor is expressed in terms of the probability that an electron can transmit through the conductor itself. For the moment let's assume zero temperature, so that the only energy states involved in conduction are only those with energy strictly in between the source and the drain Fermi levels. Moreover let's consider a conductor in between two large contacts. The probability of an electron of being transmitted from one contact (the source) to the other contact (the drain) passing through the conducting channel is indicated with T. The way in which it can be calculated depends on the nature of the considered device/channel, and will be briefly addressed in the next subsection. For the moment let's assume that it is known. Moreover it is assumed that the electrons can exit from the contacts with no reflections: "reflectionless" contacts. If a voltage V_{DS} is applied then the source and drain Fermi levels will be shifted of the amount qV_{DS}. Since the contacts are assumed reflectionless, the total incoming current of electrons from the source into the mesoscopic conductor can be evaluated considering the quantum conductance introduced in section 19.1.3. Indeed no reflections means unitary transmission, and with an applied bias V_{DS} the states in the source contact that are in between E_{FD} and E_{FS} see free states in the drain and try to go there such that the total energy of the system is lowered (they "try" to reach the equilibrium). The incoming current from the source is thus (notice the similarity -for zero kelvin- with the flow from source to dot of section 19.1):

$$I_{S,dot} = \frac{2q}{h}(E_{FS} - E_{FD}) = \frac{2q^2}{h}V_{DS}$$

The amount of electrons (or current) that reach(es) the drain, i.e. the outgoing current from the dot into the drain is thus given by the probability of transmission T through the dot that multiplies the incoming current $I_{S,dot}$:

$$I_{dot,D} = TI_{S,dot} = \frac{2q}{h}T(E_{FS} - E_{FD}) = \frac{2q^2}{h}TV_{DS}$$

The rest of the electron flux, or the incoming current, is instead reflected back to the source contact: $(1-T)I_{S,dot}$. In conclusion the net current flowing in the device is:

$$I_{DS} = \frac{2q}{h}T(E_{FS} - E_{FD}) = \frac{2q^2}{h}TV_{DS} \tag{19.22}$$

from which the conductance is:

$$G = \frac{2q^2}{h}T \tag{19.23}$$

That can be modified in: $G = \frac{2q^2}{h}TM$, if more modes/states within the quantum dot are considered, with M number of modes. In the previous treatment of a multi-level quantum dot (see section 19.1.2) M was "embedded" in $T(E)$.
Equation (19.22) is called Landauer's equation for the electrical current, flowing in a mesoscopic ballistic conductor, and equation (19.23) is referred again as Landauer's equation for conductance. Notice that in this transmission standpoint conductance is essentially equivalent to transmission, indeed they are directly proportional. Sometimes one says that conductance is transmission [44]. Notice that equation (19.23) includes the contact conductance (or quantum conductance - see section 19.1.3).
So far the discussion neglected the effects of the contacts, that indeed were assumed to be reflectionless. A remark is due concerning this assumption. Very often it is reasonable

assuming that the contacts are reflectionless for electrons coming from the dot and going into the contact, since the contact is huge and have many electron states, both occupied and available. Instead in general the reflection of electrons coming from the contact and going into the dot can be quite large, and cannot be neglected. Nevertheless the above formulae are correct since this reflection corresponds exactly to the contact conductance. In particular contact-dot reflections are accounted for by means of the contact conductance, that already considers them.
The extension of Landauer's equation to the case in which the temperature is no more zero kelvin is trivial. In particular instead of considering simply to take part to conduction the electron states between E_{FD} and E_{FS}, the Fermi-Dirac function "tails" should be considered. In particular equation (19.22) becomes:

$$I_{DS} = \frac{2q}{h} \int T(E) \left[f_{FD}(E, E_{FS}) - f_{FD}(E, E_{FD}) \right] dE \tag{19.24}$$

that corresponds exactly to the previously found equations (19.6) for the single level quantum dot and (19.8) for the multi-level case. Indeed the aforementioned equations correspond to the Landauer's formula for a single or a multi-level quantum dot.

19.3.1 The transmission function T

While the model presented above is general and holds for each kind of small conductor or mesoscopic system, the particular expression for the transmission function T depends on the particular system that is considered. In general it is possible to estimate it by exploiting quantum mechanics, and in particular the transmission function T can be obtained from the Schrödinger's equation for the mesoscopic system. Indeed it is possible to define a transmission operator, that, once applied to the eigenfunctions of the quantum dot, provides the transmission probability for each energy level. This is not in principle much different from evaluating the transmission coefficient through a potential barrier (see section 17.3.7). In the transmission formalism all the complexity arising from the quantum mechanical treatment of the states in the dot is thus hidden within the transmission functions T. Notice that is also possible to calculate T starting from the NEGF formalism, indeed there exist a connection between the ballistic NEGF formalism and the transmission formalism (to this purpose see also section **??**; while a good reference is e.g. [91]). Nevertheless it may have no much sense using the NEGF formalism only for evaluating T and then using the more limiting transmission formalism (that remember holds essentially only for ballistic transport, while the NEGF is always holding in general).
A further remark is now provided relative to the transmission operator T. In section 17.4 it was illustrated a way of representing a differential operator in matrix form. It is thus intuitive that the above mentioned transmission operator can be expressed in matrix form. In this case the transmission function T becomes a matrix: the transmission matrix, that is indeed a representation of the transmission operator in a certain basis set. Moreover it is known from the electromagnetic fields courses that it is possible to define the so called scattering matrix S, that provides reflection and transmission coefficients for a multi-terminal (or multi-port) circuit among the various ports. Moreover it is known that there exist a connection between the scattering matrix S and the transmission one T, and it is possible to switch from one to the other by means of suitable transformation formulae. This holds true even in the case of transport through mesoscopic systems, such that the transmission formalism corresponds to the scattering formalism in which each mesoscopic system or sub-system is represented by means of scattering (or transmission) matrices, and a theory quite analogous to the high frequency electronic circuits one is used. In particular, a good choice for the basis set, can be the one of the energy eigenstates of the Hamiltonian operator of the mesoscopic system. In that case said M the number of allowed energy levels, i.e. eigenvalues, it follows that the basis set will have M eigenfunctions (if no degeneracy occur). In that case the transmission and the scattering matrices will be MxM matrices.

The transmission matrix element $T_{i,j}$ will provide the probability of transmission between the *j-th* and the *i-th* proper modes, i.e. eigenstates. This point is further addressed in the next section.

19.3.2 Bütticker's formula

A simple and good discussion and introduction to the Bütticker formalism is provided in [91]. Here I will only mention that Bütticker extended the two terminal Landauer formula to the case of more terminal devices, by summing up the linear responses at each terminal (the responses must be linear to invoke the superposition of effects and sum them - see next section for a brief discussion of linear responses):

$$I_p = \frac{2q}{h} \sum_q \left[T_{q,p} E_{Fp} - T_{p,q} E_{Fq} \right] = \sum_q \left[G_{q,p} V_p - G_{p,q} V_q \right] \quad , \; G_{q,p} = \frac{2q^2}{h} T_{p,q} \; , \; V_p = E_{Fp}/q$$

where the subscripts p and q refer to two terminals p and q, and $T_{i,j}$ is the transmission function from terminal j to i. The electrical current must be zero when the applied voltages at all the terminals are equal, this fact implies that:

$$\sum_p G_{qp} = \sum_q G_{pq} \quad and \quad \sum_p T_{qp} = \sum_q T_{pq}$$

indeed it corresponds to charge conservation or matter conservation principle. Considering the last expressions it is possible to rewrite the current at terminal p as follows:

$$I_p = \sum_p G_{pq} \left[V_p - V_q \right]$$

From basic circuit theory it is known that is possible to represent whatever linear network by means of an impedance matrix. If the network is resistive it corresponds to the resistance matrix, whose inverse matrix is the conductance matrix. Thus in the multi-terminal case the conductaces $G_{p,q}$ can be seen as the elements of a conductance matrix. Moreover it is well known that a linear multi-port or multi-terminal (both lumped or distributed parameter) device admits also a representation in terms of scattering matrix, that is known being related to a transmission matrix. The transmission formalism allows to work with these matrices, and especially the scattering one is very used. Notice again that this is possible if the system is linear, topic briefly addressed in the next section. The usual approach is then to use quantum mechanics, and in particular the Schrödinger's equation, to evaluate the probability of transmission among states, i.e. the transmission rates among the different states. They are then collected within a matrix, that can be seen as a matrix representation for a transmission operator that provides the transmission probability between two different states (an initial and a final one). Then by means of algebraic transformation it is possible to get the scattering matrix, whose elements correspond to the scattering rates between the system states. Then, if a sequence of systems, or a complex mesoscopic system, is considered, it is possible to split it into simple parts or subsystems, recover the transmission or the scattering matrix and then suitably consider the total scattering/transmission matrix as a composition of the subsystem ones. This is analogous to what is done in conventional high frequency circuit. The transmission formalism rely on these few concepts, and "hides" the quantum mechanical complexity within the transmission (or scattering) function.

19.3.3 Linear response

The transmission formalism shows its full power in the moment in which a linear response is considered. This is the case of low-bias. In such conditions the transmission formalism framework allows to write and solve trivial equations, giving, at the same time, a quantitatively correct result in a fast way, and a physical insight in the description of the behaviors

of the mesoscopic systems. These advantages arise from the fact that the transmission function can be written once at the beginning starting from quantum mechanical considerations on the system under study and then it is unchanged, and one works with the Landauer's formalism without handling quantum mechanics. Nevertheless in the moment in which a full current-voltage characteristics is considered things become suddenly more complicated and in particular a self-consistent iterative solutions of the involved equations must be considered. The aim of this section is to specify when a response can be considered linear and when instead it cannot. In the latter case the consequence will be that a self-consistent solution must be provided.
The response is linear whenever the transmission function is approximately constant over the considered bias window. This statement implies two assumptions: first, the transmission function is approximately constant over the considered bias window, but also: it is independent on the applied bias. The latter statement is not valid in general, since the applied bias can modify $T(E)$, as already mentioned in section 19.1 (see final remarks). In the case of a constant $T(E)$, that is: $T(E) \sim T(E_F)$ where E_F is the Fermi level of the channel, and with a $T(E)$ independent on the bias, the Landauer's formula can be rewritten as:

$$I_{DS} \sim \frac{2q}{h} T(E_F) \int [f_{FD}(E, E_{FS}) - f_{FD}(E, E_{FD})] dE \sim \frac{2q}{h} T(E_F)[E_{FS} - E_{FD}]$$

The last equality holds obviously for low temperature, but it is also possible to demonstrate that it holds true also for higher temperatures [91]. A linear response thus occurs if the conductance $G(E) = \frac{2q}{h} T(E)$ is independent on energy in the bias window: $G \sim \frac{2q}{h} T(E_F)$. A sufficient condition for having a linear response is that the applied bias is small if compared to $k_B T$, where k_B is the Boltzmann's constant and T the temperature. Indeed in that case the Fermi levels difference is much less than $k_B T$, i.e. $E_{FS} - E_{FD} << k_B T$, and it is possible to show (by Taylor expanding the Landauer's formula for the electrical current), that the response is linear. Physically this is a consequence of the thermal broadening that affect the device. Due to temperature the electrons have extra kinetic energy w.r.t. to zero kelvin case, and as a result they can occupy higher energy levels (excited states) w.r.t. Fermi levels (ground state). Thus the Fermi functions have a longer tail and a slower exponential decay (in energy). Moreover the integral that defines the Landauer's electrical current can be seen as a correlation integral (a convolution integral in dE), noticing that the Fermi functions are dependent on E but also on Fermi level. This integral thus has the effect of "weighing" the transmission function $T(E)$ over the thermal broadening $k_B T$. In other words, the greater is the temperature the smoother is the integral of the product between $T(E)$ and the Fermi function difference. This appears very clear as soon as formulae are derived, see for example [91]. In summary, the response is linear if $T(E)$ is almost constant in energy: $T(E_F)$, if it is bias independent and if the applied bias is much less than $k_B T$: $E_{FS} - E_{FD} \ll k_B T$ (i.e. large thermal broadening w.r.t. applied bias). In such cases it is possible to formally write:

$$I_{DS} \sim \frac{2q}{h} T(E_F)[E_{FS} - E_{FD}] \quad and \quad I_p = \sum_p G_{pq} [V_p - V_q]$$

It is now addressed the question: "what happens if the applied bias is large?" In order to answer that question it must be said that in general when the bias is large (or not so small) an electric field develops within the channel. Such an electric field modifies in general the transmission functions thus leading to a non-linear response. It means that the transmission function $T(E)$ is not only functions of the energy but also of the bias: $T(E, E_{FS}, E_{FD})$ or $T(E, qV_{DS})$. In that case it is of extreme relevance including the effect of the applied electric field in order to get reliable quantitative results. In some easy cases it can be sufficient to consider a linear electric field inside the device, but in general one has to take into the electron density inside the channel and find the electric field by solving

a suitable Poisson's equation (see SCF procedure - section 18.2). This corresponds in practice in implementing a self-consistent iterative solution, in which the electrostatic of the system is solved self-consistently with the evaluation of the transmission function. This is analogous to what already seen in section 19.1. This also means that the transmission function, that is obtained from the Schrödinger's equation, cannot be evaluated only once at the beginning and then assumed constant. This complicates the calculations with the Landauer's formalism, and each case should be addressed separately in order to understand if it is more convenient performing calculations with the transmission formalism or with the NEGF one. A good treatment is presented in [91].

19.3.4 Non-coherent transport

The previously introduced Landauer-Bütticker transmission formalism rigorously holds true only for ballistic transport, i.e. for coherent transport in which a single wave-function is supposed to "carry" the electron from the source to the drain.
For non-coherent transport a "vertical" flow of electrons occurs, where with "vertical flow" it is intended that incoherent scattering phenomena occur such that the electrons can dynamically change their energy state (i.e. wave-function or state) in consequence of such scattering phenomena, and thus undergo to momentum/energy relaxation.
Nevertheless if one calls "non-coherent elastic transport" the particular transport regime in which the electrons can undergo to phase-breaking phenomena but not to momentum relaxation processes, thus in this transport regime the Landauer-Bütticker's formalism still holds. In that case no vertical flow of electrons happens, since there is no net exchange of energy between the electrons and the surrounding channel, thus no dissipation of energy or relaxation (but only phase-breaking). In this case if an electron "enters" at source in a channel (e.g. a dot eigenstate) with a given energy, then it exits from the same channel at drain side with the same energy (no change of electron state occurs). Notice that another case is the one in which an electron with energy E_1 scatters with an another electron at energy E_2. Again if no net energy is relaxed, i.e. if the scattering process is elastic in nature, at the end of the process the first electron will have energy E_2 wile the second E_1. This means that there is always an electron in the channel E_1 and one in the channel E_2, even if they are exchanged: i.e. a phase-breaking process occurred. It should be now clear the reason why in this "non-coherent elastic transport", in which phase-breaking occurs, the transmission formalism still holds. The picture of "no vertical flow" is sometimes represented by the fact that the electrons can travel in separated channels within the mesoscopic conductor, even if they can interact as mentioned above, by means of phase-breaking phenomena. Nevertheless after each phase-breaking phenomena one may think the electron continues flowing in the same channel (even if two electrons may be exchanged).
An interesting case is the one in which the transmission function is constant in the bias window. In this case the Landauer's formalism still holds. Because even if an incoherent scattering process occurs and a electron changes its energy during the propagation in the channel, since $T(E)$ is constant in energy, its final transmission probability does not change relevantly. Thus the Landauer-Bütticker's formalism provide reasonable quantitatively results, and the committed error is small. Again this corresponds essentially to the linear case. Thus if the response is linear, then the error committed in neglecting incoherent transport is small and the transmission formalism can be used.
If instead the transmission function $T(E)$ has abrupt variations in the energy range included within the bias window, the Landauer-Bütticker's transmission formalism may lead to completely wrong results, if incoherent transport has relevant effects on the total transmission. An example can be the one of a potential barrier that is thick enough to lead small transmission probability. If the temperature is enough high that many highly energetic phonons are present, they can generate a relevant thermoionic current of electrons

that classically overcome the barrier, thus leading to a much greater current w.r.t. the ballistic one. In this cases the transmission formalism is not reliable.
In summary it is possible to use the transmission formalism if:

- purely coherent transport is considered
- elastic non-coherent transport is considered (in which only phase-breaking phenomena occur but not energy relaxation ones)
- inelastic non-coherent linear transport is considered, i.e. the transmission function $T(E)$ is approximately constant in the energy range within the bias window

Instead it must be carefully used in inelastic non-coherent and non-linear transport case, in which the transmission function $T(E)$ varies abruptly within the bias window (e.g. because of localized states). In this last case the incoherent transport mechanisms can be so relevant that neglecting them can lead to completely wrong results.
Notice also that in the case of non-coherent transport it is possible to define an "effective transmission function" $T_{eff}(E)$, that embeds the scattering processes complexity, such that the Landauer's formula can be again be used. Nevertheless in order to derive $T_{eff}(E)$ the NEGF formalism should be used, with the additional complexity of converting the NEGF equations into the Landauer's one with the term $T_{eff}(E)$. Thus usually it has no much sense this approach. A good and detailed treatment of these topics is provided in [91].

19.4 Molecular devices linear response and coherent transport

In section 19.1 it was introduced a simple model for transport through molecular quantum dots. In particular in sections 19.1.4 and 19.1.7 it was introduced a linear (or linearized) capacitive model for zero-extension quantum dots. A question that may arise is when these models hold, at least approximately. The answer comes from the discussion in the previous section regarding the linear response and the transmission formalism. Indeed if the transmission spectrum $T(E)$ is approximately constant in the energy range within the bias window, the current-voltage characteristics will present a linear response, indeed the integral of a constant is a straight line. Notice that also the capacitances should be linearized w.r.t. the applied bias. This occurs if the applied bias is small, or if small variations occur such that they can be linearized around the working point. In order to get a linear response not only $T(E)$ should be constant in the bias window, but also it should be independent on the applied voltage. The latter condition is usually true if the bias is much lesser than k_BT [91].
Nevertheless I would like to point out an important insight concerning the conditions under which $T(E)$ is linear for a molecular device. In particular this happens if the molecular device is strongly coupled with the contacts. In that case indeed it was already mentioned that broadening occurs (section 19.1.1). If it is supposed to have an extremely strongly coupled molecular channel to the contacts, then the coupling factor γ will be very large, thus leading to a very broadened $D_{ELi}(E)$ and consequently to a very broadened $T(E)$. In this case it is easy to understand that $T(E)$ is more constant than in the case of weak coupling. In conclusion, if strong coupling occurs, then a great delocalization of the wave-functions occur, leading to broadened transmission functions and thus a linear response. Otherwise if weak coupling occurs, then the localized states are responsible of a $T(E)$ abruptly varying in energy, thus to a strongly non linear response. Another way of seeing the same concepts is in the current-voltage characteristics: the greater the broadening the more linear appears the current, the lower the broadening the steeper will appear the current variations in the I-V characteristics.
A last remark on incoherent transport is provided. From the discussion in the previous section it is known that a ballistic model for transport can be good if purely coherent trans-

port is present, if elastic non-coherent transport is present, and if inelastic non-coherent transport is present but with a linear response. The latter case corresponds to the case of strong coupling with the contacts (see above). Thus the only case in which non-coherent transport can be relevant in molecular devices is when weak coupling case is present, and the incoherent contribution provides a large thermoionic current that the coherent picture is not capable of catching. This last case corresponds to the case in which a potential barrier arising from a localized (thus weakly coupled state), is enough big to make the thermoionic (classical) transmission probability comparable or greater to the coherent tunneling one. In this case, with incoherent transport models the calculated current can be very different from the coherently calculated one. Nevertheless one should also consider that molecules are small. A typical dimension is less than 1 nm (benzene ring is around 5 Å long while C_{60} has a diameter of the order of 7 Å). Thus the localized states that can prevent the coherent transport, can at most generate potential barrier of few Å, always less (or at most equal) to 1 nm. The coherent tunneling effect is thus not at all negligible through a barrier of that thickness. Moreover one should also consider the probability of interaction between electrons and molecular phonons (or molecular vibrations). Indeed in order to be in the case of relevant incoherent case the electrons should undergo to energy relaxation that is possible if they interact with phonons (or photons); since it was seen previously that the electron-electron interaction give rise to phase-breaking phenomena but not to energy relaxation, unless a third particle, like a phonon, is involved in the interaction.
In order to have an idea of the amount of incoherent transport through molecules (due to phonon interactions), the following reasoning is carried on. An electron at room temperature has a thermal velocity of the order of 10^5m/s, and since the molecule is more or less 1 nm long, the electron transit time within the molecular channel is of the order of $(10^{-9}\text{m})/(10^5\text{m/s})$ that is around 10^{-14}s. A molecular vibration usually has a frequency of the order of 10^{13}Hz, thus the vibration period will be of the order to 10^{-13}s, ten times greater than the electron transit time. This could be an indication of the fact that the incoherent transport is not in general negligible in molecular devices, but anyway the transit time is one order of magnitude lesser than the typical time required to accomplish a complete molecular vibration, thus making the resonance not exactly matched. In other words, the molecule can be considered essentially as a (quasi-)rigid environment, in which electrons travel, since, as mentioned, it requires one order of magnitude less to cross the entire molecule than to accomplish a complete oscillation. And since, as mentioned previously, rigid scatterers do not lead to phase-breaking processes, but only to phase mismatches, they are are accounted in a coherent transport framework by means of the transmission functions (that can be reduced by destructive interference of electron wave-functions). Actually this is true as long as molecules are very short. But there exist molecules that are even far longer than 1 nm. In such cases the transit time can be of the order of the molecular vibration period, or even longer, and consequently the molecule cannot be considered as rigid or static during the electron motion through it. In such cases the electron phase perturbation due to the molecular dynamics can be relevant and important, and it can leads to dynamic interference effects (phase-breaking). In this optics it should be now clear that the interaction probability of an electron traveling through the molecule and a molecular vibration or phonon depends on the time the electron spends inside the molecule. If the molecule is very short and the vibrational frequencies for that specific atomic/nuclear configuration are low (i.e. long vibrational periods), then the transport is in optimal approximation only coherent. Otherwise it may not be so. In section 19.1 it was introduced the concept of intrinsic time or escaping time $\tau = \hbar/\gamma$. It was said that it represents the time one electron on average spends inside the molecule before escaping into one contact. This time thus plays a role also in determining the amount of incoherent transport in the total transport balance [232]. Indeed, the longer is the time spent by an electron within the molecule, the larger the molecular vibrational interaction probability is. But since τ is inversely proportional to the broadening γ, the smaller is the broadening (i.e. the weaker

is the coupling of the molecular channel with the contacts) the longer is τ and the higher is the probability of an electron to undergo incoherent scattering mechanics, that can be a phase-, momentum- or energy-relaxing process. In summary, if a short molecule is strongly coupled with contacts the transport the coherent transport will be essentially the only relevant mechanism of transport through it; if instead a long molecule weakly coupled with the contacts is considered, incoherent transport can be very important [232].
In this optics, [45] provides a discussion in which it concludes that incoherent transport is very improbable in devices shorter (smaller) than around 50 nm, like molecules are, because the transit time is much lesser than the scattering mean interaction time. Nevertheless this argument holds in general for crystalline semiconductor, while for molecules the previous considerations must be kept in mind. The level broadening plays a central role also in determining the amount of coherent transport, beside the more intuitive molecules length. In conclusion the point is that in many cases it may be accurate to completely neglect incoherent transport in molecular devices, especially if they are strongly coupled with contacts and the molecules are short, but there cases in which such a transport regime can be dominant, and by neglecting it quantitatively non-accurate results are obtained. These cases are often linked with a weak coupling of the molecular channel with the contacts, and with long molecules in which the orbital conjugation is broken. Indeed if orbital conjugation is broken the molecular orbitals are not mixed together and thus they are localized, leading to an increase of the permanence time of electron in such part of the molecule, and within the molecule in general. The distance above which incoherent transport can be relevant can be also far below the limit of 50 nm mentioned above, and in particular in [233] it is found to be 2.5 nm while in [234] around 4 nm, both for conjugated molecules but with different properties. Thus the intrinsic geometry of the molecule can affect also the discriminating distance for transport features. These topics are again considered (under a slightly different perspective) in section 20.2.

19.5 The non-equilibrium Green's function formalism

Molecules are small and thus they obey to the laws of quantum mechanics. A purely quantum mechanical approach in modeling transport in molecular devices is the so called Non-Equilibrium Green's Function formalism (NEGF). For this reason it can do justice to all the complex phenomena that occur in such quantum systems.
In the previous sections of this chapter a "toy" model was presented, that despite its simplicity is capable of catching several physical insights of the transport through molecular channels. Nevertheless as already pointed out in section 19.1.9, it is not a general model and several issues are completely neglected, with the risk of getting totally unreliable quantitative results.
Moreover a brief introduction to the transmission formalism was also provided in section 19.3. The main advantage of the Landauer's approach is that the transmission function can be calculated with phenomenological (often simple) approaches, sometimes with semiclassical approaches, and in these cases the transmission formalism results extremely powerful since it easily provides accurate predictions on quantum mechanical systems in an easy way. Nevertheless it is not general and its application, especially for non-coherent transport should be carefully evaluated.
All these problems can be overcome once the NEGF model is considered. It is a rigorous way of treating the transport in mesoscopic systems that holds in general, with no restriction. In particular it holds also in the case of incoherent transport. Moreover by means of the NEGF formalism it is possible to evaluate also the dissipated power within a channel in which incoherent scattering mechanisms occur [91]. Actually the real power of the NEGF formalism is indeed in incoherent transport treatment. In general the NEGF formalism bridges the gap between the fully coherent picture of the Landauer's formalism and the fully incoherent picture of standard electronic devices modeling (that exploits average con-

cepts like mobility, arising from the concepts of incoherent scattering processes). Moreover it is also a rigorous framework for the treatment of strongly non linear current-voltage characteristics, that are instead somehow difficult in conventional mesoscopic transmission based models. It will be addressed in the next chapter of this work.

CHAPTER 20

Transport regimes in nano-devices

The NEGF approach introduced in the previous chapter (here not reported as out of the aim of this Lecture Notes) is a rigorous and powerful framework for the transport calculation in all kinds of nano-devices and not only in molecular devices. It holds in general, provided that the channel electronic structure, the SCF potential, and the various interactions (i.e. electron-electron, and if necessary incoherent ones such as electron-phonon and electron-photon interactions) are accurately described. The purpose of this chapter is to highlight the methods conventionally used for such calculations in different transport regimes. In general the couple DFT + NEGF is very common in literature, even if other solutions are possible and often used, such as the EHT + NEGF approach. Nevertheless in some cases these standard methods are not suitable to obtain accurate results. This happens especially when Coulomb blockade occurs. This phenomenon is introduced in section 20.1. It allows to discriminate between two modeling and transport regimes that require substantially different approaches. In order to appreciate these differences the various transport mechanisms that are possible in a molecular device are briefly summarized in section 20.2. While the possible and common simulation methods for the different transport regimes are briefly addressed in section 20.3, that is aimed in warning the user in selecting and judging the most suitable modeling technique for a specific case.

20.1 Coulomb blockade

In this section the so called "Coulomb blockade" or "single-electron charging" regime of transport is addressed. In this transport regime the electrical current can be strongly affected by the charging effect that was introduced in chapter 19 (section 19.1.5) and in chapter **??** (e.g. section **??**). In particular, energy levels within the conducting channel come in pairs, indeed due to spin degeneracy there is one up-spin and one down-spin electron state with same energy eigenvalue. In order to highlight the effects of the electron charging phenomenon let's consider a channel with a single energy level ε (two spin-degenerate levels), that for example contains only one electron when neutral: $N_0 = 1$ (being N_0 the number of electrons in ε at equilibrium). It is expected that if such a level, that is broadened for example in a Lorentzian-like DOS, is entirely included within the bias window, a current flows, accordingly to the transport model for a single-level quantum dot presented in chapter 19, section 19.1. Nevertheless in some cases, under certain conditions that will be clarified in while, the up-spin and the down-spin density of states splits into two distinct energy levels, separated by the single electron charging energy (see e.g. section 19.1.5). In other words the presence of a single electron causes the removal of the spin degeneracy, and the two electron states have different energies. As mentioned, the two levels are separated by an amount $U_0 = q^2/C_E$, where C_E is the total electrostatic capacitance of the system,

and U_0 is the potential energy that the channel gains because of the presence of the single electron (i.e. $N_0 = 1$). This indeed corresponds to the charging potential energy associated to a transfer of a single-electron (see section 19.1.5). In this latter case if U_0 is large, it can happen that the two levels are for example one above and one below the bias window, thus leading to a very small current flowing through the device (because in such a case only the Lorentzian "tails" fall in the bias window). In this case it is said that the device is in the "Coulomb blockade" regime of transport. The Coulomb blockade has been experimentally observed for systems in which the charging energy U_0 exceeds the broadening γ of the energy level [44].
In the simple SCF picture (see section 18.2) there is no simple explanation for this phenomenon. Indeed it is expected the two levels to be degenerate as long as they "feel" the same SCF potential. Nevertheless this picture, that is the charging model based on the simple Poisson's equation, is a good-zero-order approximation. This is the Hartree's approximation for the SCF potential, as discussed in section 18.2, in which nevertheless the correlation is neglected. As already mentioned there, the point is that an electron does not "feel" any potential due to itself. Indeed assuming that the up-spin level is filled before, then the potential barrier to be overcome to put another electron there is equal to U_0, that indeed represent the energy to be spent in order to place another electron in the same level where an electron is already present (the work). In other words, the empty down-spin level, because of the presence of one electron in the up-spin level, is shifted up of U_0. Nevertheless the up-spin level is not shifted up of U_0 because the electron does not "feel" any self-interaction. The same happens if the down-spin level is occupied before. In that case the up-spin level is shifted up of U_0 while the down-spin one does not. In both cases if U_0 is enough large, the two levels can fall outside the bias window, thus blocking the current.
In general the standard approach for describing the electric current in such a Coulomb blockade regime is based on a completely different theory from the one described so far (the NEGF but also the simple model of section 19.1). It is based on the so called "Master Equation" or "Orthodox Theory" [44]. Nevertheless there are attempts in literature of exploiting the NEGF approach (along with the DFT) to model the transport also in this regime, they are addressed in section 20.3.
Notice that in general there is no net distinction between the Coulomb blockade regime and the coherent regime of transport (the one widely discussed so far). Indeed, a gradual transition between the two and thus all the intermediate cases are in principle possible (between strong Coulomb blockade regime and fully coherent regime - in which no Coulomb blockade happen, i.e. U_0 is very small). In particular, the following three regimes of transport occur [44]:

- SCF regime: if $U_0 \approx k_B T$ and/or $U_0 \approx \gamma$, the SCF approach (section 18.2) can be used. Notice that the SCF method converges correctly if U_0 is small, otherwise it can have convergence problems.
- Coulomb blockade regime: if $U_0 \gg k_B T$ and $U_0 \gg \gamma$, the SCF approach is typically not adequate. In that case high accuracy in estimating the electron-electron correlation and in general the electron-electron interactions must be employed to get quantitatively correct results. In general the so called Master Equation is conventionally used for this case. Nevertheless it has the big drawback of neglecting completely the broadening (see section 20.3 and 20.3.2).
- Intermediate regime: if U_0 is comparable to the larger between $k_B T$ and γ (i.e. thermal or contact broadening), there is no simple approach in general. Indeed the SCF method fails in representing the strong charging effect, while the Master equation fails in representing the broadening. Notice that, as mentioned previously, in general it is possible to have all intermediate cases between the two above cases, thus this situation is not a minor importance case.

A last remark on the nature of U_0 before proceeding. One may ask what determines U_0. The

answer may be the extent of the electron wave-function. If one electron is smeared over the surface of a sphere of radius R then the potential of that sphere will be: $\frac{q}{4\pi\varepsilon_0 R}$, so that the energy needed to put another electron on the sphere will be $\frac{q^2}{4\pi\varepsilon_0 R} \sim U_0$. Well-delocalized wave-functions (large R) have a very small U_0, otherwise if R is small (i.e. the wave-function is highly localized) U_0 can be large. Consequently when there are delocalized wave-functions the difference between the two energy levels (up-spin and down-spin) is smaller, since U_0 is smaller. Thus the more the conductive molecular channel orbitals are delocalized, the better is the SCF approximation. This corresponds to the strong coupling between molecule and contacts, in which there is an high level of hybridization of the wave-function, i.e. the contact and the channel wave-functions mixed together. Instead the charging energy U_0 can be large if there is weak coupling between the molecular channel and the contacts, i.e. if the states are localized and there is no (or small) hybridization of the electron orbitals. Moreover notice that since U_0 is inversely proportional to the total electrostatic capacitance C_E, the charging energy U_0 can be large also if C_E is very small. This is the case of small mescoscopic conductors. In some cases electronic devices/transistors are specifically designed to operate in the Coulomb blockade regime, and in that case they are called "Single Electron Transistors" (or SETs).

I-V characteristic and conductance

Notice that due to Coulomb blockade the current-voltage characteristics usually presents typical step-like shape. This because when the bias is increased (in modulus) from zero, for example a level that is empty in equilibrium enters the bias window originating an abrupt increment of the current, that from almost zero passes to a given finite value. After that, even if the bias is further increased, the current does not increase anymore due to Coulomb blockade. Only when the bias is increased of at least U_0 (w.r.t. to the first conduction peak) a new electron can populate the energy level (with an opposite spin w.r.t. to the previous one). And thus only when the bias overcome U_0 a new step arise in the current characteristics. Obviously the conductance will present peaks of conduction in correspondence of the bias values multiple of the charging potential (the differential conductance is the derivative of the current w.r.t. the voltage). These shapes of the current-voltage characteristics and of the conductance-voltage characteristics are typical of systems affected by Coulomb blockade.

20.2 Transport mechanisms

The purpose of this section is to briefly illustrate the possible physical mechanisms behind the transport in molecular devices, with a particular reference to the review paper [21]. This is useful in understanding when and if the NEGF approach is more or less suitable for modeling transport in such devices. Notice that usually the NEGF approach provides much better and more accurate results than the simple fitting and idealized/simplified formulae presented below. Nevertheless to gain a physical intuition/insight of what happens in the molecular system, such formulae are sometimes used to fit the more precise NEGF results in order to identify the main transport mechanism in a specific system. This could be useful to derive design parameters and figure of merits, and thus to engineering the system in order to optimize some specific transport features, that are better highlighted with a simple intuitive model rather than with the more complete (and complex) NEGF.
In particular, the electron transport mechanisms in molecular junctions depend on the molecular size and structure, the temperature and the applied voltage. Moreover if not a single molecule but a self-assembled monolayer (SAM) connected to two metallic electrodes is considered, the intermolecular interactions like van der Waals interactions or polarization effects may be relevant. Concerning simulations, such effects can be considered within a

DFT framework (SCF / mean field) by means of a supermolecular approach as described in section 18.7. However, in general, the following transport regimes are possible in molecular junctions:

a. Coherent non-resonant tunneling: for short molecules strongly coupled to contacts, without molecular energy levels in the bias window, the transport mechanism is direct non-resonant tunneling from the source to the drain contact. It can be idealized with a rectangular shaped barrier of finite height and small width (of the order of the molecule length), such that an exponential decay of the wave-function allows to get non-null probability of transmission through it (see section 17.3.7). This simplified model is referred in literature as Simmons' model. In this case the electron is coherently transmitted, but, since there is no energy match between the electron energy and an eigenstate within the molecule, an exponential decay with space occurs. In the Landauer's approximation the conductance can be written as [21]:

$$G = \frac{I}{V} = \frac{2q^2}{h} T_S T_D T_{mol} \quad , \quad G = G_{cont} T_{mol} \quad , \quad T_{mol} = e^{-\beta d}$$

Where T_S and T_D are the transmission coefficients relative to source-molecule and molecule-drain barriers, $G_{cont} = \frac{2q^2}{h} T_S T_D$, and the transmission through the molecule is fitted with an exponential decay. In such a regime of transport, sometimes, the experimental data or the more accurate simulation results (e.g. obtained with NEGF) are fitted with an exponential, and an equivalent β decay constant (measurement unit m^{-1}) is derived. This procedure allows to a straightforward interpretation of the transport through the molecule that, as mentioned, is seen as a rectangular barrier. This transport regime is naturally accurately modeled by means of (coherent) NEGF approach. Usually this regime is relevant at low bias V ($qV < E_0$, being E_0 the barrier height). Notice that this mechanism is independent on temperature and it is also independent on temperature.

b. Coherent resonant tunneling: in this case there is matching between the incoming electron energy and a molecular eigenstate, thus there is high transmission probability and small reflection probability. As already commented in chapter **??**, a peak in the transmission spectrum occurs (see also transparency of a potential barrier - section 17.3.7). Such a transport regime is of course included within the NEGF modeling framework for coherent transport. Notice that this mechanism is essentially independent on temperature.

c. Fowler–Nordheim tunneling: if the bias is large (say $qV > E_0$) then the simple model for coherent non-resonant tunneling through the barrier (point (a.) above) is no more valid since the barrier distorts, and it becomes triangular. In such a case it is possible to have an enhancement of the tunneling transmission. Indeed for zero-bias it may happen that the barrier width is to large to allow significant tunneling, while with an enough high bias it tends to become triangular and thus thinner. Consequently the exponential decay of the wave-function can be no more enough to stop electron and a great improvement of the tunneling probability occurs. This is referred as Fowler–Nordheim tunneling. It is a well known process in electronics since there are many commercial memory devices based on this effect. In molecular devices such an effect leads to an enhancement of the tunneling via orbital-mediated tunneling. This transport regime is of course included within the NEGF modeling framework for coherent transport. Notice that this mechanism is essentially independent on temperature. Also in this case fitting parameters are sometimes extracted from experimental data to highlight the voltage enhancement of tunneling, analogously to what mentioned in point (a.).

These three transport mechanisms are coherent tunneling ones, independent on temperature. The NEGF is far more accurate than all these three, nevertheless it was mentioned the reason for introducing them, as conceptual tools to understand how to act in order to engineer the desired transport features. A trivial example is that if it appears that the

transport mechanism is somehow essentially analogous to a Fowler–Nordheim tunneling, and for a given application it would be better to have more current, an increase of the voltage will trivially likely provide such an increased current, with an exponential relation. In [21] are summarized fitting formula for these transport mechanisms.
Higher order interactions (i.e. many body interactions), like the electron-electron, electron-phonon and electron-photon interactions, can affect the transport through molecules. In such cases other transport mechanisms might be relevant in molecular devices. In particular, as discussed in section 19.4, in the strong coupling regime, the electronic states of the electrodes and the molecular orbitals are strongly hybridized, molecular charging effect (Coulomb blockade - see section 20.1) does not take place and elastic coherent tunneling dominates the transport. In this case it is expected to have a mix of transport mechanisms (a.), (b.), (c.). However, if a weak coupling occurs between the molecule and contacts, Coulomb blockade and charging effects may be relevant. In such a case the coherence between the motion of the electron from the left electrode to the molecule and that from the molecule to the right electrode can be completely lost. Indeed as mentioned in section 19.4, the longer is the time that an electron spends within the molecule, the higher is the probability that it interacts with molecular vibrations/phonons, etc... [232]. Indeed if the time an electron stays in the molecule is of the order (or even longer) than the molecular vibration period, the atomic positions cannot be assumed to be static, but instead they act as dynamic scatterers, thus breaking the phase coherence of electrons, leading to both interference among electrons and to inelastic/dissipative interactions (e.g. phonons). In such cases the dominant transport mechanism is the incoherent tunneling or alternatively the electron hopping mechanism (these transport mechanism will be introduced in a while - see below). Moreover, all the possible intermediate situations in between a full coherent picture (see above (a.), (b.), (c.) cases) and a full incoherent picture (see below) are possible. In the intermediate coupling regime a variety of novel phenomena related to electron–electron correlation effects (such as the Kondo effect and co-tunneling) are observed [21], [232], [44].
Coherent tunneling is effective for very short molecules, let's say of the order of 1 nm. In general depending on the nature of the molecule, on the nature of the contact-molecule coupling and on the applied bias range different transport behaviors are possible for different molecule lengths. For example in [233], it is found the limit molecular length for purely coherent transport in p-phenylenevinylene oligomers (conjugated molecules) is of the order of 2.5 nm; while in [234] the same is found to be around 4 nm for other conjugated molecules. Thus a wide variability is possible. The main incoherent transport mechanisms through molecules are:

d. Incoherent or Sequential tunneling: in this case the transport through a molecule is seen in terms of tunneling through multiple potential barriers/wells. This model applies quite well to conducting polymers [21]. The main different w.r.t. coherent tunneling is that the residence time of the electron in a potential well is long enough to affect the phase of the electron. Thus the electrons do not tunnel with same phase (or wavenumber) but each tunneling process, through each barrier, is independent from the previous one (and involves different states, different wave-functions). The whole electron transport process can be thus described as a series of discrete coherent tunneling steps through barriers. It is not temperature-dependent, like the other tunneling processes. This transport mechanism is again accurately modeled with the NEGF approach, even if incoherent scattering mechanism should be included to get accurate results. Indeed a strong electron-electron interaction or alternative dissipative processes can be extremely relevant to estimate the transition probability from one energy level to another while the electron is in transit within the molecule. And this is exactly what I called (with the notation of [91]) "vertical flow" of electrons within the channel, i.e. transitions/flow between different energy values (generally from an higher to a lower energetic channel/level but also vice versa). More details are also provided in the next section (see weak coupling / Coulomb blockade case).

e. Hopping: in contrast to tunneling mechanisms (from (a.) to (d.)), hopping conduction involves electron motion over the barrier. In the sense that it corresponds to electron "jumps" from one molecular orbital to another. These jumps can occur if the electron gains energy (phonon/photon absorption) such that it temporarily overcome the barrier, or for example if the molecule is dynamically bent/twisted such that (in real space) two orbitals get closer allowing for a shorter barrier length and thus higher tunneling probability. In both the above mentioned cases it is noticeable that hopping is a thermally activated process, since high thermal energy is required to excite molecular vibrations such that an insurmountable barrier becomes instead easily crossed by electrons. Generally hopping follows the exponential empirical and classical Arrhenius relation. An explicit expression is provided for example in [21]: $I \propto e^{-\frac{E_a}{k_B T}}$, where E_a is the "activation energy", and the greater it is, the greater is the temperature required to activate that hopping phenomenon. In general it represents the barrier of potential energy that an electron should overcome to transit from a molecular orbital to another. For example in [233], molecular wires made by connecting together π-conjugated molecules are considered. In such a case it is demonstrated that the activation energy E_a fitting the experimental data, corresponds to the barrier for rotation of the π-conjugated aromatic rings, which transiently couples the conjugated aromatic sub-units (the rings). Thus torsional vibrations are responsible of the hopping, but only if they are enough wide. In particular it is verified that such collective torsional vibrations are associated to an energy of the order of E_a, that is thus the thermal energy necessary to stimulate such vibrations and activate the hopping among adjacent orbitals (that are the hopping sites). In general the activation energy is associated with the movements of nuclei. As mentioned above hopping electron transfer can only occur after a rearrangement of nuclei, that in the above example corresponds to a rotation of rings in a molecule into a coplanar conformation. Hopping involves, like incoherent tunneling, a series of transfers between different sites, but shows an inverse distance dependence. Hopping is probable at low bias high temperature. The NEGF approach can model accurately this transport mechanism but only if incoherent phenomena are considered. In particular molecular vibrations and electron-phonon interactions must be considered.

f. Thermionic emission: the last transport mechanism in molecular device, discussed here, is the thermoionic emission. It is present also in classical electronic devices. In this case the electrons overcome the contact–molecule barrier by thermal agitation and the resulting current has a strong dependence on temperature. The barrier height is influenced by the local electric field which results in a nonlinear current dependence on applied voltage. Thermionic emission can be described using the classical Schottky–Richardson relation, that is conventionally introduced in electronic devices courses. An expression referred to molecular devices is present in [21]. Thermoionic is relevant at high temperature and low contact-molecule barrier height. The NEGF approach intrinsically consider also this transport mechanism, especially when incoherent (phonons) scattering processes are accounted for. Notice that indeed electrons that are thermoionically emitted from source are those whose energy overcomes the barrier height. The barrier height is precisely calculated in NEGF thanks to the self-consistent loop by means of which the electrostatics of the system is taken into account, while the number of electrons with suitable energy are estimated by means of the Fermi-Dirac's distributions of contacts. Molecular phonons can enhance this transport mechanism.

In conclusion, even if conceptually very different, all the mentioned transport mechanisms (from (a.) to (f.)) are accurately modeled and considered by means of the NEGF approach. Obviously, if a relevant transport mechanism is non-coherent, the NEGF modeling should include the incoherent scattering mechanisms, and this is possible as described in section **??**. The NEGF formalism results thus to be a completely general and accurate model for transport though molecules. As mentioned, the discussion reported in this section can be

important to gain an intuition about what is happening in a given molecular system, and thus to correctly set up the desired simulations, such as, for example, incoherent features or corrections.
A critical issue that needs a more in-depth and careful discussion is the case of strong electron-electron interaction, that corresponds to large charging effect and Coulomb blockade regime (i.e. weak coupling with contacts). This is further addressed in the next section.

20.3 Modeling approaches: strong vs weak coupling case and corrections

The purpose of this section is to provide a deeper insight in the NEGF modeling critical issues relative to the electron-electron interactions. The topic is addressed by steps, firstly summarizing the standard approach in transport modeling through systems in which it is possible to neglect electron-electron interactions (section 20.3.1) and then considering the possible corrections to such models, or directly alternative models to transport, for system in which the electron-electron interactions are important and cannot be neglected (section 20.3.2).

20.3.1 Strong coupling case

It was already discussed (see section 19.4 and the previous sections in this chapter) that in the case of strong coupling between the molecular channel and the contacts the charging energy is small due to delocalization of molecular orbitals (hybridization). In such a case an SCF mean field approach (see section 18.2) is accurate, Coulomb blockade does not characterize the transport and no particular corrections should be performed. In this case (weak electron-electron interaction or strong coupling with contacts) the standard mean field DFT method can be used for calculating the molecular electronic structure while the standard (single-particle) NEGF, that was introduced in chapter **??**, can be used to calculate the transport features and the full I-V characteristics. In literature it is very widely used the aforementioned couple DFT + NEGF, and this is also the approach considered e.g. in [44]. Another possible approach is to use a semi-empirical method instead of the DFT for the molecular electronic structure calculations. The most used one is the EHT method, and in this case the EHT + NEGF pair is used (usually with SCF self-consistent calculation in EHT). The differences between the two approaches are essentially only due to the differences between the EHT and DFT methods, described in chapter 18. Both the approaches are very accurate and successful as long as the electron charging energy and thus the electron-electron interactions are weak.

20.3.2 Weak coupling case and the Single Electron Transistor

When the electron-electron interaction is strong and thus the electron charging energy is large, the fact that the electron correlation is ignored or not precisely modeled within an SCF framework can lead to confusion and inaccuracies [232]. Indeed the electron correlation is the key to suitably model the interactions among electrons and thus the charging effect (and consequently the Coulomb blockade). In fact it is because of the interactions between electrons that important energy shifts upon charging can happen [232], giving rise to the aforementioned phenomena. In general there is no unique way to overcome the problem. The various solutions can be classified in three wide groups:

a. “Modified” DFT/Electronic structure calculations: the first approach can be to modify somehow the SCF mean field approach to consider also accurately the electron-electron interactions. In this case a “modified” DFT is employed along with a “non-modified” single-particle NEGF. Therefore the complexity and the corrections are included within

the molecular electronic structure calculation method, i.e. the DFT, while the NEGF portion of the self-consistent loop (section **??**) is unchanged w.r.t. the single-particle picture presented in chapter **??**. This means that the effort is in finding an SCF potential that suitably models also the electron-electron interactions, while a single-particle picture is used for transport, considering a single-electron moving in an average potential representing accurately such interactions. It was mentioned in chapter 18 (section 18.4) that there exist electron correlation methods for the calculations of accurate electron correlation/interaction. Such approaches are not so used in literature, or better: some approaches (such as perturbation theories) leads actually to the same of the methods illustrated below in point (b.), while others are not usable as described above because of the excessive computational cost or analytical complexity.

b. "Modified" NEGF: this approach moves the complexity within the NEGF/transport part, i.e. an unaltered standard (SCF mean field) DFT method (or EHT) is used for calculating the electronic structure while an "improved" NEGF is used for considering also the electron-electron interactions. This is not in principle a novel approach, since it simply correspond in the approach described throughout this work, simply with an additional self-energy aimed in modeling electron-electron interactions (accordingly to what mentioned in section **??**), indeed the self-energies (as said in section **??**) are in general used to represent any kind of interaction, comprised the electron-electron one.

c. Completely novel approaches: the last approach can be the one of considering completely new methods aimed in accurately describing such systems.

These three groups of solutions are further discussed in the following subsections. Before going on a last remark on nomenclature. A transistor operating in the Coulomb blockade regime of transport is often referred as Single Electron Transistor (SET).

a. "Modified" DFT/Electronic structure calculations

In this approach, the complexity is added to the DFT such that the electron-electron strong interaction is considered within the SCF approach, i.e. by means of a potential that is solution of a Poisson's like equation. The critical point in using a mean field is indeed the electron correlation: it is difficult to find an accurate exchange-correlation potential (or exchange-correlation functional of the electron density / density matrix) that accurately provides and models the electron-electron interaction (see section 18.2), especially in the moment in which such an interaction is strong (as happens with weakly coupled molecular channels to contact and in the Coulomb blockade regime). If it were possible to find an enough accurate functional the problem will be easily overcome since it would be enough to use such a functional and then set up all the rest of the calculations as presented previously in this work (with single-particle NEGF, etc...).
As highlighted in [235], the problem with conventional exchange-correlation approximations (e.g. LDA, GGA, ...) is that their functionals are essentially "too smooth", and they are not able to accurately account for first derivative discontinuities that might occur in the total system energy eigenvalues. In particular, one may think the charging effect to be an abrupt variation of the charging potential. Indeed the charging energy U_0 arises abruptly as soon as the number of electrons within a given energy value changes of one unity. In other words, consider an energy level weakly coupled with contacts and for this reason very localized and merely broadened, and consider an initial applied bias such that such a level is slightly outside the bias window. Assume also that this level is initially empty. As soon as the applied bias is increased a bit, the energy level enters abruptly in the bias window, since it is merely broadened, and starts conducting. For example in such conditions a single electron populate the level when it is in the bias window. At this point the charging energy is increased of U_0, that by hypothesis is large (see section 20.1). This means that, to put another electron in the same level, the barrier U_0 should be overcome, or analogously that the mean field that is "felt" by a new incoming electron should be increased abruptly of the

order of U_0 at least in the region in which there is such localized orbital. This is possible if the exchange-correlation functional admits a discontinuous first derivative, such that it is possible to the potential energy to have abrupt variations. Nevertheless this is not the case of the standard LDA and GGA functionals, that instead have continuous first derivative [235], [236]. This fact prevent the system total energy and the electron density (or analogously the electron charge) within the channel to have abrupt variations, as it should be in the case of strong electron-electron interaction or strong charging effect. Moreover, the continuity of first derivative of the total energy functionals in standard approximations, is a consequence of the so called Self-Interaction Error (SIE), that is the interaction of an electron with the exchange-correlation potential generated by its own charge [235]. This spurious interaction is at the origin of the DFT failures in case of strong electron-electron interactions [236]. Indeed as mentioned in section 20.1, an electron does not "feel" any potential due to itself.
The simplest conceptual solution is the subtraction of such self-interaction contribution from the functional. In general there are more ways of practically doing it, and these corrections are called Self-Interaction Corrections (SIC) or Atomic Self-Interaction Correction (ASIC). In the commercial tool *Quantum-Wise ATK* it is possible to use a SIC for DFT calculations that is called "DFT-1/2" [237], [238], [239]; in which only one half of electron self-interaction is subtracted, for details [239]. Further details on the usage are provided within the user's manual [237], [238].
Another possible correction in *Quantum-Wise ATK* for considering the strong electron-electron interaction is the so called "Hubbard correction" [237], [238], [240], [241]. This is a semi-empirical correction of the exchange-correlation functional in LDA, GGA functionals. An additive term is added to such functionals, that is null if a molecular orbital is full or empty, but has a non-null value if an orbital is half-filled. This method improves on the deficiencies of the local exchange-correlation functionals discussed above. Further details on the usage are provided within the user's manual [237], [238].
A last approach of refining the electron-electron interaction and exchange-correlation potential within the DFT framework is by means of hybrid functionals (section 18.5), such as the famous B3LYP. Indeed such functionals allow to more accurately estimate the exchange-correlation. In literature there are several examples in which this approach is used also for simulating SETs. In *Quantum-Wise ATK* there is no wide choice for hybrid functionals, nevertheless the user's manual states that the functional TB09 (that is not hybrid but a meta-GGA functional) has performances that are often comparable to hybrid functionals ones [237], [238].

b. "Modified" NEGF

In this approach the complexity is moved inside the NEGF. In particular in section **??** it was mentioned that there several orders of approximation to model more o less accurately the many-body interactions such as the electron-electron, electron-phonon and electron-photon ones. In [44] the lowest order approximation is considered, i.e. the so called Born approximation. More refined approximations are possible in literature. One may think to these approximations as higher order terms in a Taylor expansion. In this optics it is clear that the higher order approximations lead to more accurate models. In particular in this specific case higher order approximations means approximations in which two-particle, or even three-particle or n-particle interactions are considered. The ways of "embedding" such more refined approximations within the NEGF framework are essentially two: either by means of additional self-energy contributions aimed in modeling the specific interaction (i.e. the electron-electron one), or by means of a two-particle treatment (or in general n-particle treatment).

The first is for example the approach of [242]. The authors, starting from a general many-

body Green's function (second order since two interacting electrons are considered), recover an expression for a single-particle (i.e. first order) Green's function of the system (analogous to the one presented in chapter **??**), with an additional self-energy Σ_U that models the two-electron-interaction (two-electron interaction is a second order term - i.e. two particles). Notice that in this procedure there are approximations as well, indeed, as already mentioned in section **??**, in the moment in which the second order Green's function is written (i.e. the one of two-interacting particles), it turns out that it is function of the third order one (three interacting particles), and so on... Thus the "recursion is cut" at a certain point. As one can intuitively think Σ_U is function of the number of electrons at equilibrium and of the number of electrons outside equilibrium.
The latter approach is for example the one used in [243]. In that case the difference w.r.t. the previous method is that the two-particles Green's function is maintained for doing the calculations. In deriving it the three-interacting-particle should be known, as mentioned above, but it is not considered (cut - like in a Taylor expansion), and thus approximations are introduced (the results are accurate for short range Coulomb blockade but for example not so suited for the Kondo effect).

c. Completely novel approaches

In general it is possible to proceed with completely novel approaches in modeling devices in which the transport is governed by Coulomb blockade, like in SETs. The standard approach, historically very used, for SET modeling is indeed the so called "Orthodox Theory" or "Master Equation" [244], [245]. The NEGF method is at the end a perturbative method, indeed as mentioned in section **??**, in the second quantization formal introduction of NEGF, the Green's function is a quantum propagator that propagates a stimulus, i.e. a perturbation, through the system. The Master Equation approach is instead non-perturbative. The disadvantage of the NEGF is that the analytical and computational complexity increases fast if the one-particle picture is abandoned. Instead the Master Equation approach has he advantage of being relatively simple and general, even if, in the words of [244], it is a "quick and dirty" approach. In order to understand it notice that in general a Master Equation is an equation involving probability, or better probability rates or scattering rates. Thus the approach of the Master Equation is to use quantum mechanical laws to determine the transition probability, or scattering rates, among all the possible states of the multi-electron system. The Master Equation is the equation that collect all these transition probabilities, and that is thus able to predict the scattering rates among the different states. For this reason it is "quick"; because it does not requires many additional concepts to the ones of standard quantum mechanics. Moreover it is "dirty" compared with the extremely elegant Keldysh theory of the NEGF [244]. Nevertheless the drawbacks of such methods are far beyond the "elegance". There are mainly two extremely limiting disadvantages in the using the Master Equation. Firstly, since it involves transition probabilities among all the possible states, its complexity increases very fast with the number of electrons composing the multi-electron system. In particular with N_e electrons the Master Equation requires working with matrices of dimension $2^{N_e}\text{x}2^{N_e}$. Second, there is no way of representing or introducing the broadening within the Master Equation framework. Thus it is of course not suitable for the strong coupling case, but neither for all intermediate cases in which broadening is present, even if small. The above presented approaches instead have no limitations in these terms. For this reason in literature the Master Equation is usually employed for modeling simple SETs with few electron states and in strong Coulomb blockade and extremely weak coupling with contacts. Because of these disadvantages recent literature (last two/three years) seems to be oriented toward the approach (a.) or eventually (b.) for high accuracy.
Another modeling technique, possible for SETs, is the so called Quantum Monte Carlo. Analogously to its "classical" or "standard" counterpart, it is a statistical method in which

random transitions among states are performed (considering quantum mechanical laws) and thanks to simulation of a huge number of transitions, accurate transition rates are found. This method is very effective in considering also a large number of electron states, and for this reason it is often used for circuit-level simulations of SET-based networks. A drawback is that, with this method, the physics is "hidden", and even if the final result can be accurate, the physical insight of the device is lost.
In literature several references are present about these topics, here I have chosen to avoid reporting them for conciseness, since I will never use these last two approaches in the rest of this work. A simplified introduction to Master Equation is moreover present in [44].

A remark on EHT and other approaches

In chapter 18 it was mentioned many times that semi-empirical methods are generally less accurate than *ab initio* ones, but the result might also be accurate. The correct general way of dealing with semi-empirical methods is thus to "validate" them by means of comparisons with experimental data or with *ab initio* calculations. Even in this case the approach is again this one. Indeed with the above mentioned methods and corrections, there is good confidence that calculations performed e.g. with NEGF or DFT + NEGF with the corrections of point (a.) are close to reality, even if there is no comparison with experimental data. Indeed in that cases the theory predicts it, and within a reasonable error range the results can be assumed reliable (with the limits of the considered approximations). With semi-empirical methods, and with EHT in particular, there is no *a priori* reason to assume that the result is wrong or inaccurate. Nevertheless, there is neither a theory behind that ensures the correctness (within certain limits). Thus in general one can use semi-empirical methods or the EHT method, but they should be somehow "validated" as explained above. In literature there are examples [85] [246] in which SETs, operating in the case of Coulomb blockade, are simulated with EHT + NEGF method; but in general no theoretical justification of the reason why the results are matched with experimental data or other methods is provided. Likely because such explanation does not exist, accordingly to what highlighted above. Sometimes in order to emulate weak contact coupling, the molecule is placed a little bit more far away from the contacts, again with no theoretical justification (also because in this case the geometry of the device is modified). Obviously if the molecule is taken away from contacts the coupling diminish, like does the orbital delocalization, thus the desired results can be obtained. In this optics one may think of treating the distance between molecule and contacts as a sort of additional fitting parameter (notice that there are already other fitting parameters in EHT and semi-empirical methods) for emulating the weak coupling with contacts. From this standpoint, if the final results are "validated" and enough accurate (for the purposes of the study) then there is no reason in my opinion of avoiding using such methods, that present the great advantage of being computationally inexpensive when compared with DFT or other approaches, being however aware of the above mentioned limits.
A final remark is provided about another approach, somehow similar in the principle to the one just described. In [247] the authors successfully perform simulation in the case of strong electron-electron interaction by proceeding as follows. They consider separately the up-spin and the down-spin states, and in particular they calculate the transmission spectra for up- and down-spin separately. At the end they sum the two spectra (obtained considering separated the up-spin and the down-spin), and the final result is somehow accurate. As the authors state this procedure allows to get good result but it is just a computational tool, with no theoretical foundation. Indeed they are considering a C_{60} molecule that has no magnetic properties, thus in general there is no reason to consider separated spins since electrons can in principle start at source with a given spin state and reach the drain with another spin state, i.e. spin mixing is possible in principle (the current is not spin-polarized). Nevertheless they neglect the spin mixing due to spin-orbit coupling

(i.e. an electron can change spin state during propagation). Moreover they make another unphysical assumption: they neglect non-collinear spin densities (i.e. they assume that the electron cannot point in any direction but it can be only UP or DOWN belonging to the same plane). All these facts are non-physical, nevertheless the final result matches with other experimental and theoretical works. Notice that the transport regime that is considered is coherent (with separated spins). Since the final result is somehow accurate, this method can be used for simulations in the weak coupling regime; with the awareness of what just mentioned, and with some computational advantages. The reason why the final results is not totally wrong is because by considering separately the UP and the DOWN spin cases they are implicitly assuming that the electrons with different spins do not interact and thus they travel in completely separated channels. This corresponds somehow in emulating an high charging potential such that the presence of an electron in a state prevent another electron to transit in the same level (Coloumb blockade). This argument is not at all convincing, nevertheless such unphysical method works fine in some situations, and, aware of its assumptions, there is no reason to avoid using it.

Bibliography

[1] Chiara Elfi Spano. Molecular nanocomputing: An engineering approach from physics to circuit architectures. Master's thesis, Politecnico di Torino, Torino, 2020.

[2] Fabrizio Mo. Molecular electronic sensors modeling: From theory to applications. Master's thesis, Politecnico di Torino, Torino, 2020.

[3] Yuri Ardesi. Study of molecules for quantum dot cellular automata. Master's thesis, Politecnico di Torino, Torino, 2018.

[4] Giuliana beretta. Study of field-coupled nanocomputing based on molecules for neural systems. Master's thesis, Politecnico di Torino, Torino, 2020.

[5] Fabrizio Riente. Design methods and tools for nanocomputing: from silicon nanoarrays to nano magnetic logic. Master's thesis, Politecnico di Torino, Torino, 2016.

[6] Giovanna Turvani. Out of plane nanomagnetic logic. Master's thesis, Politecnico di Torino, Torino, 2016.

[7] Marco Vacca. Emerging technologies - nanomagnets logic (nml). Master's thesis, Politecnico di Torino, Torino, 2013.

[8] Chiara Cannavo'. Computing architectures based on skyrmions. Master's thesis, Politecnico di Torino, Torino, 2019.

[9] Paolo Pagliarulo. Studio di gate lim programmabili sviluppati su tecnologia racetrack memory. Master's thesis, Politecnico di Torino, Torino, 2019.

[10] Elio Pio Reveruzzi. Challenges of organic quantum dot cellular automata fabrication. Master's thesis, Politecnico di Torino, Torino, 2018.

[11] Giorgio Alemanno. Study of clocking system for molecular fcn. Master's thesis, Politecnico di Torino, Torino, 2018.

[12] Deloitte analysis, 2020. https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/technology-media-telecommunications/deloitte-cn-tmt-semiconductors-the-next-wave-en-190422.pdf.

[13] Samsung makes 1TB flash eUFS module, 2019. https://www.electronicsweekly.com/news/business/samsung-makes-1tb-flash-module-2019-01/.

[14] Cerebras, 2019. https://fuse.wikichip.org/news/3010/a-look-at-cerebras-wafer-scale-engine-half-square-foot-silicon-chip/.

[15] ASML ANNUAL REPORT 2013. https://www.sec.gov/Archives/edgar/data/937966/000119312514046822/d546896d20f.htm.

[16] International Roadmap for Devices and Systems (IRDS) 2020 Edition. https://irds.ieee.org/editions/2020.

[17]

[18] International roadmap for devices and systems - executive report, 2017.

[19] International roadmap for devices and systems - executive report, 2018.

[20] International roadmap for devices and systems - executive report, 2020.

[21] Silvia Karthäuser. Control of molecule-based transport for future molecular devices. *Journal of Physics: Condensed Matter*, 23(1):013001, nov 2010.

[22] Hyungsub Choi and Cyrus C.M. Mody. The long history of molecular electronics: Microelectronics origins of nanotechnology. *Social Studies of Science*, 39(1):11–50, 2009.

[23] G. Cuniberti, G. Fagas, and K. Richter.

[24] John Markoff. Chip industry sets a plan for life after silicon. *NewYork Times*, 29 December.

[25] Paven Thomas Mathew and Fengzhou Fang. Advances in molecular electronics: A brief review. *Engineering*, 4(6):760 – 771, 2018.

[26] James M. Tour. Molecular electronics. synthesis and testing of components. *Accounts of Chemical Research*, 33(11):791–804, 2000. PMID: 11087316.

[27] Michael C. Petty, Takashi Nagase, Hitoshi Suzuki, and Hiroyoshi Naito. *Molecular Electronics*, pages 1–1. Springer International Publishing, Cham, 2017.

[28] Lanlan Sun, Yuri A. Diaz-Fernandez, Tina A. Gschneidtner, Fredrik Westerlund, Samuel Lara-Avila, and Kasper Moth-Poulsen. Single-molecule electronics: from chemical design to functional devices. *Chem. Soc. Rev.*, 43:7378–7411, 2014.

[29] James M. Tour.

[30] Arieh Aviram and Mark A. Ratner. Molecular rectifiers. *Chemical Physics Letters*, 29:277–283, November 1974.

[31] Katharina Kaiser, Lorel M. Scriven, Fabian Schulz, Przemyslaw Gawel, Leo Gross, and Harry L. Anderson. An sp-hybridized molecular carbon allotrope, cyclo[18]carbon. *Science*, 365(6459):1299–1301, 2019.

[32] Qizhi Xu, Giovanni Scuri, Carly Mathewson, Philip Kim, Colin Nuckolls, and Delphine Bouilly. Single electron transistor with single aromatic ring molecule covalently connected to graphene nanogaps. *Nano Letters*, 17(9):5335–5341, 2017. PMID: 28792226.

[33] Chuancheng Jia, Bangjun Ma, Na Xin, and Xuefeng Guo. Carbon electrode– molecule junctions: A reliable platform for molecular electronics. *Accounts of Chemical Research*, 48(9):2565–2575, 2015. PMID: 26190024.

[34] Huimin Wen, Wengang Li, Jiewei Chen, Gen He, Longhua Li, Mark A. Olson, Andrew C. H. Sue, J. Fraser Stoddart, and Xuefeng Guo. Complex formation dynamics in a single-molecule electronic device. *Science Advances*, 2(11), 2016.

[35] F. Zahid, M. Paulsson, E. Polizzi, Ghosh A.E., Siddiqui L., and S. Datta. A self-consistent transport model for molecular conduction based on extended Hückel theory with full three-dimensional electrostatics. *The Journal Of Chemical Physics*, (123), 2005.

[36] Chuancheng Jia, Marjan Famili, Marco Carlotti, Yuan Liu, Peiqi Wang, Iain M. Grace, Ziying Feng, Yiliu Wang, Zipeng Zhao, Mengning Ding, Xiang Xu, Chen Wang, Sung-Joon Lee, Yu Huang, Ryan C. Chiechi, Colin J. Lambert, and Xiangfeng Duan. Quantum interference mediated vertical molecular tunneling transistors. *Science Advances*, 4(10), 2018.

[37] Angelika Balliou, Jiri Pfleger, George Skoulatakis, Samrana Kazim, Jan Rakusan, Stella Kennou, and Nikos Glezos. Programmable molecular-nanoparticle multi-junction networks for logic operations. In *Proceedings of the 14th IEEE/ACM International Symposium on Nanoscale Architectures*, NANOARCH '18, page 37–43, New York, NY, USA, 2018. Association for Computing Machinery.

[38] Yu Li, Chen Yang, and Xuefeng Guo. Single-molecule electrical detection: A promising route toward the fundamental limits of chemistry and life science. *Accounts of Chemical Research*, 53(1):159–169, 2020. PMID: 31545589.

[39] Yu Li, Lihua Zhao, Yuan Yao, and Xuefeng Guo. Single-molecule nanotechnologies: An evolution in biological dynamics detection. *ACS Applied Bio Materials*, 3(1):68–85, 2020.

[40] Wenwen Hu, Liangtian Wan, Yingying Jian, Cong Ren, Ke Jin, Xinghua Su, Xiaoxia Bai, Hossam Haick, Mingshui Yao, and Weiwei Wu. Electronic noses: From advanced materials to sensors aided with data processing. *Advanced Materials Technologies*, 4(2):1800488, 2019.

[41] Matti Kaisti. Detection principles of biological and chemical fet sensors. *Biosensors and Bioelectronics*, 98:437 – 448, 2017.

[42] Yoav Y. Broza, Xi Zhou, Miaomiao Yuan, Danyao Qu, Youbing Zheng, Rotem Vishinkin, Muhammad Khatib, Weiwei Wu, and Hossam Haick. Disease detection with molecular biomarkers: From chemistry of body fluids to nature-inspired chemical sensors. *Chemical Reviews*, 119(22):11761–11817, 2019. PMID: 31729868.

[43] S. J. Ray. Humidity sensor using a single molecular transistor. *Journal of Applied Physics*, 118.

[44] Supriyo Datta. *Quantum transport: atom to transistor*. Cambridge University Press, 2005.

[45] Marc Baldo. *Introduction to Nanoelectronics*. MIT OpenCourseWare Publication, 2011.

[46] Ismael Rattalino, Paolo Motto, Gianluca Piccinini, and Danilo Demarchi. A new validation method for modeling nanogap fabrication by electromigration, based on the resistance–voltage (r–v) curve analysis. *Physics Letters A*, 376(30):2134 – 2140, 2012.

[47] Bingqian Xu and Nongjian J. Tao. Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science*, 301(5637):1221–1223, 2003.

[48] Valentin Dubois, Shyamprasad N. Raja, Pascal Gehring, Sabina Caneva, Herre S. J. van der Zant, Frank Niklaus, and Göran Stemme. Massively parallel fabrication of crack-defined gold break junctions featuring sub-3 nm gaps for molecular devices. *Nature Communications*, 9(1):3433, Aug 2018.

[49] M. Kiguchi, R. Stadler, I. S. Kristensen, D. Djukic, and J. M. van Ruitenbeek. Evidence for a single hydrogen molecule connected by an atomic chain. *Phys. Rev. Lett.*, 98:146802, Apr 2007.

[50] Linda Zotti, Beatrice Bednarz, Juan Hurtado-Gallego, Damien Cabosart, Gabino Rubio-Bollinger, Nicolas Agrait, and Herre Zant. Can one define the conductance of amino acids? *Biomolecules*, 9:580, 10 2019.

[51] Takanori Harashima, Yusuke Hasegawa, Satoshi Kaneko, Manabu Kiguchi, Tomoya Ono, and Tomoaki Nishino. Highly reproducible formation of polymer single-molecule junction for well-defined current signal. *Angewandte Chemie International Edition*, 58, 04 2019.

[52] Elad Mentovich, Bogdan Belgorodsky, and Shachar Richter. Resolving the mystery of the elusive peak: Negative differential resistance in redox proteins. *The Journal of Physical Chemistry Letters*, 2:1125–1128, 04 2011.

[53] Limin Xiang, Julio L. Palma, Yueqi Li, Vladimiro Mujica, Mark A. Ratner, and Nongjian Tao. Gate-controlled conductance switching in dna. *Nature Communications*, 8(1):14471, Feb 2017.

[54] Paven Thomas Mathew and Fengzhou Fang. Advances in molecular electronics: A brief review. *Engineering*, 4(6):760 – 771, 2018.

[55] Wenjun Xu, Edmund Leary, Songjun Hou, Sara Sangtarash, Teresa González, Gabino Rubio-Bollinger, Qingqing Wu, Hatef Sadeghi, Lara Tejerina, Kirsten Christensen, Nicolás Agraït, Simon Higgins, Colin Lambert, Richard Nichols, and Harry Anderson. Unusual length-dependence of conductance in cumulene molecular wire. *Angewandte Chemie*, 04 2019.

[56] Yanxi Zhang, Saurabh Soni, Thijs Krijger, Pavlo Gordiichuk, Xinkai Qiu, Gang Ye, H. Jonkman, Andreas Herrmann, Karin Zojer, Egbert Zojer, and Ryan Chiechi.

Tunneling probability increases with distance in junctions comprising self-assembled monolayers of oligothiophenes. *Journal of the American Chemical Society*, 140, 10 2018.

[57] Thijs Stuyver, Tao Zeng, Yuta Tsuji, Paul Geerlings, and Frank De Proft. Diradical character as a guiding principle for the insightful design of molecular nanowires with an increasing conductance with length. *Nano Letters*, 18(11):7298–7304, Nov 2018.

[58] A. Zahir, S. A. A. Zaidi, A. Pulimeno, M. Graziano, D. Demarchi, G. Masera, and G. Piccinini. Molecular transistor circuits: From device model to circuit simulation. In *2014 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pages 129–134, 2014.

[59] Wanda Andreoni, Alessandro Curioni, and Henrik Grönbeck. Density functional theory approach to thiols and disulfides on gold: Au(111) surface and clusters. *International Journal of Quantum Chemistry*, 80(4-5):598–608, 2000.

[60] N. J. Tao. Electron transport in molecular junctions. *Nature Nanotechnology*, 1(3):173–181, Dec 2006.

[61] David L. Cooper, Joseph Gerratt, and Mario Raimondi. The electronic structure of the benzene molecule. *Nature*, 323(6090):699–701, Oct 1986.

[62] Katharine E. Linton, Mark A. Fox, Lars-Olof Pålsson, and Martin R. Bryce. Oligo(p-phenyleneethynylene) (ope) molecular wires: Synthesis and length dependence of photoinduced charge transfer in opes with triarylamine and diaryloxadiazole end groups. *Chemistry – A European Journal*, 21(10):3997–4007, 2015.

[63] Mariagrazia Graziano, A. Zahir, Ahmed Mahmoud, Azzurra Pulimeno, Gianluca Piccinini, and Paolo Lugli. Hierarchical modeling of opv-based crossbar architectures. 08 2014.

[64] Cornelius Thiele, Lukas Gerhard, Thomas Eaton, David Torres, Marcel Mayor, Wulf Wulfhekel, Hilbert von Löhneysen, and Maya Lukas. Stm study of oligo(phenyleneethynylene)s. *New Journal of Physics*, 17, 05 2015.

[65] Chuancheng Jia, Marjan Famili, Marco Carlotti, Yuan Liu, Peiqi Wang, Iain M. Grace, Ziying Feng, Yiliu Wang, Zipeng Zhao, Mengning Ding, Xiang Xu, Chen Wang, Sung-Joon Lee, Yu Huang, Ryan C. Chiechi, Colin J. Lambert, and Xiangfeng Duan. Quantum interference mediated vertical molecular tunneling transistors. *Science Advances*, 4(10), 2018.

[66] Guowen Peng, Mikkel Strange, Kristian S. Thygesen, and Manos Mavrikakis. Conductance of conjugated molecular wires: Length dependence, anchoring groups, and band alignment. *The Journal of Physical Chemistry C*, 113(49):20967–20973, Dec 2009.

[67] Ahmed Mahmoud and Paolo Lugli. First-principles study of a novel molecular rectifier. *Nanotechnology, IEEE Transactions on*, 12:719–724, 09 2013.

[68] Fu Xiaoxiao, Rui Zhang, Guang-Ping Zhang, and Zong-Liang Li. Rectifying properties of oligo(phenylene ethynylene) heterometallic molecular junctions: Molecular length and side group effects. *Scientific reports*, 4:6357, 09 2014.

[69] Linda A. Zotti, Thomas Kirchner, Juan-Carlos Cuevas, Fabian Pauly, Thomas Huhn, Elke Scheer, and Artur Erbe. Revealing the role of anchoring groups in the electrical conduction through single-molecule junctions. *Small*, 6(14):1529–1535, 2010.

[70] QuantumATK version Q-2019.12, Synopsys QuantumATK (www.synopsys.com/silicon/quantumatk.html).

[71] Roar Søndergaard, Sebastian Strobel, Eva Bundgaard, Kion Norrman, Allan Hansen, Edgar Albert, Gyorgy Csaba, Paolo Lugli, Marc Tornow, and Frederik Krebs. Conjugated 12 nm long oligomers as molecular wires in nanoelectronics. *Journal of Materials Chemistry - J MATER CHEM*, 19, 06 2009.

[72] Muhammed Schukfeh, Kristian Storm, Ahmed Mahmoud, Roar Søndergaard, Anna Szwajca, Allan Hansen, Peter Hinze, Thomas Weimann, Sofia Svensson, Achyut Bora,

Kimberly Dick, Claes Thelander, Frederik Krebs, Paolo Lugli, Lars Samuelson, and Marc Tornow. Conductance enhancement of inas/inp heterostructure nanowires by surface functionalization with oligo(phenylene vinylene)s. *ACS nano*, 7, 04 2013.

[73] Stanislav Tsoi, Igor Griva, Scott Trammell, Amy Blum, Joel Schnur, and Nikolai Lebedev. Electrochemically controlled conductance switching in a single molecule: Quinone-modified oligo(phenylene vinylene). *ACS nano*, 2:1289–95, 07 2008.

[74] Ahmed Mahmoud and Paolo Lugli. Atomistic study on dithiolated oligo-phenylenevinylene gated device. *Journal of Applied Physics*, 116:204504, 11 2014.

[75] Mark W. H. Hoorens, Miroslav Medved', Adèle D. Laurent, Mariangela Di Donato, Samuele Fanetti, Laura Slappendel, Michiel Hilbers, Ben L. Feringa, Wybren Jan Buma, and Wiktor Szymanski. Iminothioindoxyl as a molecular photoswitch with 100 nm band separation in the visible range. *Nature Communications*, 10(1):2390, Jun 2019.

[76] Carson Bruns and J. Stoddart. *Molecular Switches and Machines with Mechanical Bonds*, pages 555–733. 10 2016.

[77]

[78] Fei Lu, Qi Qin, Yuan Li, and Jiezhi Chen. Computational design of molecular transistor with van der waals gating. *Applied Physics Express*, 13, 07 2020.

[79] D. Hou and J. H. Wei. The difficulty of gate control in molecular transistors, 2011.

[80] Sujit S. Datta, Douglas R. Strachan, and A. T. Charlie Johnson. Gate coupling to nanoscale electronics. *Phys. Rev. B*, 79:205404, May 2009.

[81] Jonas Fransson, O. Bengone, Andreas Larsson, and Jim Greer. A physical compact model for electron transport across single molecules. *Nanotechnology, IEEE Transactions on*, 5:745 – 749, 12 2006.

[82] Hyunwook Song, Youngsang Kim, Yun Hee Jang, Heejun Jeong, Mark A. Reed, and Takhee Lee. Observation of molecular orbital gating. *Nature*, 462(7276):1039–1043, Dec 2009.

[83] Dong Xiang, Dongku Kim, Hyunhak Jeong, Takhee Lee, Yongjin Cheng, Qingling Wang, and Dirk Mayer. Three-terminal single-molecule junctions formed by mechanically controllable break junctions with side gating. *Nano letters*, 13, 05 2013.

[84] Chuancheng Jia, Agostino Migliore, Na Xin, Shaoyun Huang, Jinying Wang, Qi Yang, Shuopei Wang, Hongliang Chen, Duoming Wang, Boyong Feng, Zhirong Liu, Guangyu Zhang, Da-Hui Qu, He Tian, Mark A. Ratner, H. Q. Xu, Abraham Nitzan, and Xuefeng Guo. Covalently bonded single-molecule junctions with stable and reversible photoswitched conductivity. *Science*, 352(6292):1443–1445, 2016.

[85] A. Nasri, A. Boubaker, B. Hafsi, W. Khaldi, and A. Kalboussi. High-sensitivity sensor using c60-single molecule transistor. *IEEE Sensors Journal*, 18(1):248–254, Jan 2018.

[86] Sattar Arshadi and F. Anisheh. Theoretical study of cr and co- porphyrin-induced c70 fullerene: a request for a novel sensor of sulfur and nitrogen dioxide. *Journal of Sulfur Chemistry*, 38(4):357–371, 2017.

[87] Louis C. Brousseau. Label-free "digital detection" of single-molecule dna hybridization with a single electron transistor. *Journal of the American Chemical Society*, 128(35):11346–11347, 2006. PMID: 16939245.

[88] Yan-Dong Guo, Xiao-Hong Yan, and Yang Xiao. Computational investigation of dna detection using single-electron transistor-based nanopore. *The Journal of Physical Chemistry C*, 116(40):21609–21614, 2012.

[89] S. Pilehvar and K. De Wael. Recent advances in electrochemical biosensors based on fullerene-c60 nano-structured platforms. *Biosensors*, 5.

[90] Ahmed J. Hassan. Computational study of adsorption of some gas molecules on the undoped and n-doped fullerenes c20 bowl as a gas sensor. *Iranian Journal of Organic Chemistry*, 11(3):2659–2665, 2019.

[91] Supriyo Datta. *Electronic transport in mesoscopic systems*. Cambridge University Press, 1995.

[92] Jie Bai, Abdalghani Daaoub, Sara Sangtarash, Xiaohui Li, Yongxiang Tang, Qi Zou, Hatef Sadeghi, Shuai Liu, Xiaojuan Huang, Zhibing Tan, Junyang Liu, Yang Yang, Jia Shi, Gábor Mészáros, Wenbo Chen, Colin Lambert, and Wenjing Hong. Anti-resonance features of destructive quantum interference in single-molecule thiophene junctions achieved by electrochemical gating. *Nature Materials*, 18(4):1476–4660, 2019.

[93] F. Zahid, Ghosh A.E., M. Paulsson, and S. Polizzi, E. Datta. Charging-induced asymmetry in molecular conductors. *Physical review B*, 70(245317), 2004.

[94] Zhi-Hao Zhao, Lin Wang, Shi Li, Wei-Dong Zhang, Gang He, Dong Wang, Shi-Min Hou, and Li-Jun Wan. Single-molecule conductance through an isoelectronic b–n substituted phenanthrene junction. *Journal of the American Chemical Society*, 142(18):8068–8073, 2020. PMID: 32321243.

[95] Alexander V. Rudnev, Veerabhadrarao Kaliginedi, Andrea Droghetti, Hiroaki Ozawa, Akiyoshi Kuzume, Masa-aki Haga, Peter Broekmann, and Ivan Rungger. Stable anchoring chemistry for room temperature charge transport through graphite-molecule contacts. *Science Advances*, 3(6), 2017.

[96] Pascal Gehring, Jakub K. Sowa, Jonathan Cremers, Qingqing Wu, Hatef Sadeghi, Yuewen Sheng, Jamie H. Warner, Colin J. Lambert, G. Andrew D. Briggs, and Jan A. Mol. Distinguishing lead and molecule states in graphene-based single-electron transistors. *ACS Nano*, 11(6):5325–5331, 2017. PMID: 28423272.

[97] Veronika Obersteiner, David A. Egger, and Egbert Zojer. Impact of anchoring groups on ballistic transport: Single molecule vs monolayer junctions. *The Journal of Physical Chemistry C*, 119(36):21198–21208, 2015. PMID: 26401191.

[98] Manabu Kiguchi. Electrical conductance of single c60 and benzene molecules bridging between pt electrode. *Applied Physics Letters*, 95(073301).

[99] K. Oura, V.G. Lifshits, A.A. Saranin, A.V. Zotov, and M. Katayama.

[100] Simulations of graphene nanoribbon field effect transistor for the detection of propane and butane gases: A first principles study. *Nanomaterials*, 10(1:98).

[101] Simulation of graphene nanoribbon based gas sensor. *Journal of Nanoscience and Nanoengineering*, 1(2):66 – 73, 2015.

[102]

[103] David Harvey. DePauw University.

[104] Hongliang Chen, Weining Zhang, Mingliang Li, Gen He, and Xuefeng Guo. Interface engineering in organic field-effect transistors: Principles, applications, and perspectives. *Chemical Reviews*, 120(5):2879–2949, 2020. PMID: 32078296.

[105] Pedro E. Martín Várguez, Frédéric Brunel, and Jean-Manuel Raimundo. Recent electrochemical/electrical microfabricated sensor devices for ionic and polyionic analytes. *ACS Omega*, 5(10):4733–4742, 2020.

[106] M.J. Schoening and A. Poghossian.

[107] Matias Urdampilleta, Cedric Ayela, Pierre-Henri Ducrot, Daniel Rosario-Amorin, Abhishake Mondal, Mathieu Rouzières, Pierre Dechambenoit, Corine Mathonière, Fabrice Mathieu, Isabelle Dufour, and Rodolphe Clérac. Molecule-based microelectromechanical sensors. *Scientific Reports*, 8(1):2045–2322, 2018.

[108] Nishuang Liu, Guojia Fang, Wei Zeng, Hao Long, Longyan Yuan, and Xingzhong Zhao. Novel zno nanorod flexible strain sensor and strain driving transistor with an ultrahigh 107 scale "on" "off" ratio fabricated by a single-step hydrothermal reaction. *The Journal of Physical Chemistry C*, 115(2):570–575, 2011.

[109] Bingfang Wang, Yuanyuan Luo, Bo Liu, and Guotao Duan. Field-effect transistor based on an in situ grown metal–organic framework film as a liquid-gated sens-

ing device. *ACS Applied Materials & Interfaces*, 11(39):35935–35940, 2019. PMID: 31502434.

[110] Óscar Leonardo Camargo Moreira, Wei-Ying Cheng, Huei-Ru Fuh, Wei-Chen Chien, Wenjie Yan, Haifeng Fei, Hongjun Xu, Duan Zhang, Yanhui Chen, Yanfeng Zhao, Yanhui Lv, Gang Wu, Chengzhai Lv, Sunil K. Arora, Cormac Ó Coileáin, Chenglin Heng, Ching-Ray Chang, and Han-Chun Wu. High selectivity gas sensing and charge transfer of snse2. *ACS Sensors*, 4(9):2546–2552, 2019. PMID: 31456397.

[111] Hyunjin Park, Sungmi Yoo, Hyungju Ahn, Joohee Bang, Yuri Jeong, Mihye Yi, Jong Chan Won, Sungjune Jung, and Yun Ho Kim. Low-temperature solution-processed soluble polyimide gate dielectrics: From molecular-level design to electrically stable and flexible organic transistors. *ACS Applied Materials & Interfaces*, 11(49):45949–45958, 2019. PMID: 31738047.

[112] Saravanan Yuvaraja, Sandeep G Surya, Valeriya Chernikova, Mani Teja Vijjapu, Osama Shekhah, Prashant M. Bhatt, Suman Chandra, Mohamed Eddaoudi, and Khaled N. Salama. Realization of an ultrasensitive and highly selective ofet no2 sensor: The synergistic combination of pdvt-10 polymer and porphyrin–mof. *ACS Applied Materials & Interfaces*, 12(16):18748–18760, 2020. PMID: 32281789.

[113] Yunpeng Zhang, Xiaotong Liu, Shi Qiu, Qiuqi Zhang, Wei Tang, Hongtao Liu, Yunlong Guo, Yongqiang Ma, Xiaojun Guo, and Yunqi Liu. A flexible acetylcholinesterase-modified graphene for chiral pesticide sensor. *Journal of the American Chemical Society*, 141(37):14643–14649, 2019. PMID: 31448915.

[114] Yifan Wu, Yin Xiao, Xuepeng Wang, Xiaoxuan Li, and Yong Wang. Chirality discrimination at the single molecule level by using a cationic supermolecule quasi-gated organic field effect transistor. *ACS Sensors*, 4(8):2009–2017, 2019. PMID: 31274289.

[115] Bin Wang, Tan-Phat Huynh, Weiwei Wu, Naseem Hayek, Thu Trang Do, John C. Cancilla, Jose S. Torrecilla, Masrur Morshed Nahid, John M. Colwell, Oz M. Gazit, Sreenivasa Reddy Puniredd, Christopher R. McNeill, Prashant Sonar, and Hossam Haick. A highly sensitive diketopyrrolopyrrole-based ambipolar transistor for selective detection and discrimination of xylene isomers. *Advanced Materials*, 28(21):4012–4018, 2016.

[116] Seohyun Mun, Yoonkyung Park, Yong-Eun Koo Lee, and Myung Mo Sung. Highly sensitive ammonia gas sensor based on single-crystal poly(3-hexylthiophene) (p3ht) organic field effect transistor. *Langmuir*, 33(47):13554–13560, 2017. PMID: 29125766.

[117] World Health Organization. Regional Office for Europe.

[118] A. Bonanno, M. Crepaldi, I. Rattalino, P. Motto, D. Demarchi, and P. Civera. A 0.13 μm cmos operational schmitt trigger r-to-f converter for nanogap-based nanosensors read-out. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(4):975–988, April 2013.

[119] G. Massicotte, S. Carrara, G. Di Micheli, and M. Sawan. A cmos amperometric system for multi-neurotransmitter detection. *IEEE Transactions on Biomedical Circuits and Systems*, 10(3):731–741, June 2016.

[120] S. S. Ghoreishizadeh, I. Taurino, G. De Micheli, S. Carrara, and P. Georgiou. A differential electrochemical readout asic with heterogeneous integration of bio-nano sensors for amperometric sensing. *IEEE Transactions on Biomedical Circuits and Systems*, 11(5):1148–1159, Oct 2017.

[121] S. Naus, I. Tzouvadaki, P. Gaillardon, A. Biscontini, G. De Micheli, and S. Carrara. An efficient electronic measurement interface for memristive biosensors. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4, May 2017.

[122] A. Tuoheti, S. Aiassa, F. Criscuolo, F. Stradolini, I. Tzouvadaki, S. Carrara, and D. Demarchi. New approach for making standard the development of biosensing devices by a modular multi-purpose design. *IEEE Transactions on NanoBioscience*, 19(3):339–346, July 2020.

[123] I. N. Hanitra, F. Criscuolo, N. Pankratova, S. Carrara, and G. D. Micheli. Multi-channel front-end for electrochemical sensing of metabolites, drugs, and electrolytes. *IEEE Sensors Journal*, 20(7):3636–3645, April 2020.

[124] I. N. Hanitra, F. Criscuolo, S. Carrara, and G. De Micheli. Multi-target electrolyte sensing front-end for wearable physical monitoring. In *2019 15th Conference on Ph.D Research in Microelectronics and Electronics (PRIME)*, pages 249–252, July 2019.

[125] Savannah Afsahi, Mitchell B. Lerner, Jason M. Goldstein, Joo Lee, Xiaoling Tang, Dennis A. Bagarozzi, Deng Pan, Lauren Locascio, Amy Walker, Francie Barron, and Brett R. Goldsmith. Novel graphene-based biosensor for early detection of zika virus infection. *Biosensors and Bioelectronics*, 100:85 – 88, 2018.

[126] Giwan Seo, Geonhee Lee, Mi Jeong Kim, Seung-Hwa Baek, Minsuk Choi, Keun Bon Ku, Chang-Seop Lee, Sangmi Jun, Daeui Park, Hong Gi Kim, Seong-Jun Kim, Jeong-O Lee, Bum Tae Kim, Edmond Changkyun Park, and Seung Il Kim. Rapid detection of covid-19 causative virus (sars-cov-2) in human nasopharyngeal swab specimens using field-effect transistor-based biosensor. *ACS Nano*, 14(4):5135–5142, 2020. PMID: 32293168.

[127] Binghe Wang and Eric V. Anslyn. *Chemosensors*. John Wiley & Sons, Ltd, 2011.

[128] Xiaoming Tao, editor. *Wearable Electronics and Photonics*. Woodhead Publishing, 2005.

[129] Jane McCann and David Bryson. *Smart Clothes and Wearable Technology*. Woodhead Publishing, 2009.

[130] I. Jones and G.K. Stylios. *Joining Textiles*. Woodhead Publishing, 2013.

[131] Vinod Kumar Khanna. *Implantable Medical Electronics. Prosthetics, Drug Delivery, and Health Monitoring*. Springer, 2016.

[132] I. N. Hanitra, F. Criscuolo, S. Carrara, and G. De Micheli. Multi-target electrolyte sensing front-end for wearable physical monitoring. In *2019 15th Conference on Ph.D Research in Microelectronics and Electronics (PRIME)*, pages 249–252, July 2019.

[133] M. Padash and S. Carrara. A 3d printed wearable device for sweat analysis. In *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–5, June 2020.

[134] S. Shityakov, N. Roewer, C. Förster, and Croscheit J. A. In silico modeling of indigo and tyrian purple single-electron nano-transistors using density functional theory approach. *Nanoscale Res Lett*, 12(439), 2017.

[135] Namik Akkilic, Stefan Geschwindner, and Fredrik Höök. Single-molecule biosensors: Recent advances and applications. *Biosensors and Bioelectronics*, 151:111944, 2020.

[136] Yanxiao Feng, Yuechuan Zhang, Cuifeng Ying, Deqiang Wang, and Chunlei Du. Nanopore-based fourth-generation dna sequencing technology. *Genomics, Proteomics & Bioinformatics*, 13(1):4 – 16, 2015.

[137] S. Heerema and C. Dekker. Graphene nanodevices for dna sequencing. *Nature Nanotech*, 11.

[138] Henk W. Ch. Postma. Rapid sequencing of individual dna molecules in graphene nanogaps. *Nano Letters*, 10(2):420–425, 2010. PMID: 20044842.

[139] Thomas A. Albright, Jeremy K. Burdett, and Myung-Hwan Whangbo. *Orbital interactions in chemistry*. 2nd edition, 2013.

[140] Ferdinand Huber, Julian Berwanger, Svitlana Polesya, Sergiy Mankovsky, Hubert Ebert, and Franz J. Giessibl. Chemical bond formation showing a transition from physisorption to chemisorption. *Science*, 366(6462):235–238, 2019.

[141] S. J. Ray. Single molecular transistor as a superior gas sensor. *Journal of Applied Physics*, 118(3):034303, 2015.

[142] Barsha Jain, K. Vinod Kumar, B. SanthiBhushan, Kumar Gaurav, Manisha Pattanaik, and Anurag Srivastava. A tetracene-based single-electron transistor as a chlorine sensor. *Journal of Computational Electronics*, 17(4):1515–1520, Dec 2018.

[143] Klaus Sattler. *Handbook of Nanophysics, Clusters and Fullerenes*. 11 2010.

[144] S. M. Sze and M. K. Lee. *Semiconductor Devices: Physics and Technology - 3rd ed.* John Wiley & Sons Inc, 2012.

[145] G. Ghione. *Semiconductor Device for High-Speed Optoelectronics*. Cambridge University Press, 2009.

[146] MT Niemier, GH Bernstein, G Csaba, A Dingler, XS Hu, S Kurtz, S Liu, J Nahas, W Porod, M Siddiq, et al. Nanomagnet logic: progress toward system-level integration. *Journal of Physics: Condensed Matter*, 23(49):493202, 2011.

[147] Gyorgy Csaba and Wolfgang Porod. Behavior of nanomagnet logic in the presence of thermal noise. In *2010 14th International Workshop on Computational Electronics*, 2010.

[148] Marco Vacca, Mariagrazia Graziano, Alessandro Chiolerio, Andrea Lamberti, Marco Laurenti, Davide Balma, Emanuele Enrico, Federica Celegato, Paola Tiberto, Luca Boarino, et al. Electric clock for nanomagnet logic circuits. In *Field-Coupled Nanocomputing*, pages 73–110. Springer, 2014.

[149] M. Cofano, G. Santoro, M. Vacca, D. Pala, G. Causapruno, F. Cairo, F. Riente, G. Turvani, M. R. Roch, M. Graziano, and M. Zamboni. Logic-in-memory: A nano magnet logic implementation. In *2015 IEEE Computer Society Annual Symposium on VLSI*, pages 286–291, July 2015.

[150] J Anthony C Bland and Bretislav Heinrich. *Ultrathin Magnetic Structures I: An Introduction to the Electronic, Magnetic and Structural Properties*, volume 1. Springer Science & Business Media, 2006.

[151] CT Rettner, S Anders, JEE Baglin, T Thomson, and BD Terris. Characterization of the magnetic modification of co/pt multilayer films by he+, ar+, and ga+ ion irradiation. *Applied physics letters*, 80(2):279–281, 2002.

[152] GJ Kusinski and G Thomas. Physical and magnetic modification of co/pt multilayers by ion irradiation. *Microscopy and Microanalysis*, 8(04):319–332, 2002.

[153] Xueming Ju, Stephanie Wartenburg, Jamila Rezgani, Markus Becherer, Josef Kiermaier, Stephan Breitkreutz, Doris Schmitt-Landsiedel, Wolfgang Porod, Paolo Lugli, and Gyorgy Csaba. Nanomagnet logic from partially irradiated co/pt nanomagnets. *Nanotechnology, IEEE Transactions on*, 11(1):97–104, 2012.

[154] Olav Hellwig, Andreas Berger, Jeffrey B Kortright, and Eric E Fullerton. Domain structure and magnetization reversal of antiferromagnetically coupled perpendicular anisotropy films. *Journal of Magnetism and Magnetic Materials*, 319(1):13–55, 2007.

[155] Wei Gong, Hua Li, Zhongren Zhao, and Jinchang Chen. Ultrafine particles of fe, co, and ni ferromagnetic metals. *Journal of Applied Physics*, 69(8):5119–5121, 1991.

[156] Xueming Ju et al. *Micromagnetic Simulation of Field-Coupled Devices from Co/Pt Nanomagnets*. PhD thesis, Technische Universität München, 2012.

[157] A. Kobs A. Vogel S. Wintz M. Im P. Fischer H. P. Oepen U. Merkt G. Meier J. Kimling, T. Gerhardt. Tuning of the nucleation field in nanowires with perpendicular magnetic anisotropy. *Journal of Applied Physics, Vol. 113*, 2013.

[158] S Magdaleno-Adame and JC Olivares-Galvan. Coil systems to generate uniform magnetic field volumes. In *Excerpt from the proceedings of the COMSOL conference*, volume 13, pages 401–411, 2010.

[159] Irina Eichwald, Josef Kiermaier, Stephan Breitkreutz, Junyong Wu, Gyorgy Csaba, Doris Schmitt-Landsiedel, and Markus Becherer. Towards a signal crossing in double-layer nanomagnetic logic. *Magnetics, IEEE Transactions on*, 49(7):4468–4471, 2013.

[160] Xueming Ju, Michael T Niemier, Markus Becherer, Wolfgang Porod, Paolo Lugli, and Gyorgy Csaba. Systolic pattern matching hardware with out-of-plane nanomagnet logic devices. *Nanotechnology, IEEE Transactions on*, 12(3):399–407, 2013.

[161] Thomas Fischbacher, Matteo Franchin, Giuliano Bordignon, and Hans Fangohr. A systematic approach to multiphysics extensions of finite-element-based micromagnetic simulations: Nmag. *Magnetics, IEEE Transactions on*, 43(6):2896–2898, 2007.

[162] Michael Joseph Donahue and Donald Gene Porter. *OOMMF User's guide*. US Department of Commerce, Technology Administration, National Institute of Standards and Technology, 1999.

[163] I. Eichwald C. Hildbrand G. Csaba D. Schmitt-Landsiedel S. Breitkreutz, J. Kiermaier and M. Becherer. Experimental demonstration of a 1-bit full adder in perpendicular nanomagnetic logic. *IEEE transactions on magnetics, Vol. 49, No. 7*, July, 2013.

[164] S. Breitkreutz. *Perpendicular Nanomagnetic Logic: Digital Logic Circuits from Field-coupled Magnets*. PhD thesis, 2015.

[165] J.J.W. Goertz. *Domain wall propagation through notches in perpendicular magnetic anisotropy nanowires*. PhD thesis, 2015.

[166] S. Breitkreutz et al. Influence of the domain wall nucleation time on the reliability of perpendicular nanomagnetic logic. *Proceedings of the 14th IEEE International Conference on Nanotechnology*, August 18-21, 2014.

[167] Stephan Breitkreutz et al. Modelling and simulation of nanomagnetic logic with cadence virtuoso using verilog-a. *45th European Solid State Device Research Conference (ESSDERC)*, 2015.

[168] S. Breitkreutz et al. 1-bit full adder in perpendicular nanomagnetic logic using a novel 5-input majority gate. *EDP Sciences*, 2014.

[169] I. Eichwald X. Ju G. Csaba D. Schmitt-Landsiedel S. Breitkreutz, J. Kiermaier and M. Becherer. Majority gate for nanomagnetic logic with perpendicular magnetic anisotropy. *IEEE transactions on magnetics, Vol. 48, No. 11*, November 2012.

[170] J. Ping Liu, Zhidong Zhang, and Guoping Zhao. *Skyrmions: Topological Structures, Properties, and Applications*, chapter 8, page 213. CRC Press, 2016.

[171] Yan Zhou and Motohiko Ezawa. A reversible conversion between a skyrmion and a domain-wall pair in a junction geometry. *Nature Communications*, 5:4652 EP –, Aug 2014. Article.

[172] Shinichiro Seki and Masahito Mochizuki. *Skyrmions in Magnetic Materials*, chapter 1, page 2. Springer International Publishing, 2016.

[173] W. Kang, Y. Huang, X. Zhang, Y. Zhou, and W. Zhao. Skyrmion-electronics: An overview and outlook. *Proceedings of the IEEE*, 104(10):2040–2061, Oct 2016.

[174] Øyvind Johansen. Electric Control of Skyrmion Dynamics and Spin Torque Oscillators in Magnetic Materials with Inversion Asymmetry. Master's thesis, Norwegian University of Science and Technology, 2016.

[175] Giovanni Finocchio, Felix Büttner, Riccardo Tomasello, Mario Carpentieri, and Mathias Kläui. Magnetic skyrmions: from fundamental to applications. *Journal of Physics D: Applied Physics*, 49(42):423001, sep 2016.

[176] Markus Hoffmann, Bernd Zimmermann, Gideon P. Müller, Daniel Schürhoff, Nikolai S. Kiselev, Christof Melcher, and Stefan Blügel. Antiskyrmions stabilized at interfaces by anisotropic dzyaloshinskii-moriya interactions. *Nature Communications*, 8(1):308, 2017.

[177] Albert Fert, Vincent Cros, and João Sampaio. Skyrmions on the track. *Nature Nanotechnology*, 8:152 EP –, Mar 2013.

[178] Magnetic vortices, skyrmions, etc. Accessed: 2019-07-30.

[179] Christoforos Moutafis, Stavros Komineas, and J A. C. Bland. Dynamics and switching processes for magnetic bubbles in nanoelements. *Phys. Rev. B*, 79, 06 2009.

[180] Hans-Benjamin Braun. Topological effects in nanomagnetism: from superparamagnetism to chiral quantum solitons. *Advances in Physics*, 61(1):1–116, 2012.

[181] J. Ping Liu, Zhidong Zhang, and Guoping Zhao. *Skyrmions: Topological Structures, Properties, and Applications*, chapter 1, page 9. CRC Press, 2016.

[182] Utkan Güngördü, Rabindra Nepal, Oleg A. Tretiakov, Kirill Belashchenko, and Alexey A. Kovalev. Stability of skyrmion lattices and symmetries of quasi-two-dimensional chiral magnets. *Phys. Rev. B*, 93:064428, Feb 2016.

[183] Xichao Zhang, Motohiko Ezawa, and Yan Zhou. Magnetic skyrmion logic gates: conversion, duplication and merging of skyrmions. *Scientific Reports*, 5:9400 EP –, Mar 2015. Article.

[184] Shizeng Lin, Avadh Saxena, and Cristian Batista. Meron crystals in chiral magnets. 06 2014.

[185] Wataru Koshibae and Naoto Nagaosa. Theory of antiskyrmions in magnets. *Nature Communications*, 7:10542 EP –, Jan 2016. Article.

[186] J. Ping Liu, Zhidong Zhang, and Guoping Zhao. *Skyrmions: Topological Structures, Properties, and Applications*, chapter 1, 8, pages 27, 213. CRC Press, 2016.

[187] Lorenzo Camosi, Nicolas Rougemaille, Olivier Fruchart, Jan Vogel, and Stanislas Rohart. Micromagnetics of antiskyrmions in ultrathin films. *Phys. Rev. B*, 97:134404, Apr 2018.

[188] Siying Huang, Chao Zhou, Gong Chen, Hongyi Shen, Andreas K. Schmid, Kai Liu, and Yizheng Wu. Stabilization and current-induced motion of antiskyrmion in the presence of anisotropic dzyaloshinskii-moriya interaction. *Physical Review B*, 96, 09 2017.

[189] J M D Coey. New permanent magnets; manganese compounds. *Journal of Physics: Condensed Matter*, 26(6):064211, jan 2014.

[190] Xichao Zhang, Yan Zhou, Motohiko Ezawa, G. P. Zhao, and Weisheng Zhao. Magnetic skyrmion transistor: skyrmion motion in a voltage-gated nanotrack. *Scientific Reports*, 5:11369 EP –, Jun 2015. Article.

[191] Fereshte Ghahari, Daniel Walkup, Christopher Gutiérrez, Joaquin F. Rodriguez-Nieva, Yue Zhao, Jonathan Wyrick, Fabian D. Natterer, William G. Cullen, Kenji Watanabe, Takashi Taniguchi, Leonid S. Levitov, Nikolai B. Zhitenev, and Joseph A. Stroscio. An on/off berry phase switch in circular graphene resonators. *Science*, 356(6340):845–849, 2017.

[192] Christian Pfleiderer and Achim Rosch. Single skyrmions spotted. *Nature*, 465:880 EP –, Jun 2010.

[193] Mauricio Manfrini. *Spin orbit torques in magnetic materials (Kiran Sethu)*. PhD thesis, 06 2017.

[194] J. Ping Liu, Zhidong Zhang, and Guoping Zhao. *Skyrmions: Topological Structures, Properties, and Applications*, chapter 8, pages 227–228. CRC Press, 2016.

[195] Yi Wang, Praveen Deorani, Xuepeng Qiu, Jae Hyun Kwon, and Hyunsoo Yang. Determination of intrinsic spin hall angle in pt. *Applied Physics Letters*, 105(15):152412, 2014.

[196] Ye-Hua Liu and You-Quan Li. Dynamics of magnetic skyrmions. 24(1):017506, 2015.

[197] Xichao Zhang, Yan Zhou, and Motohiko Ezawa. Magnetic bilayer-skyrmions without skyrmion hall effect. *Nature Communications*, 7:10293 EP –, Jan 2016. Article.

[198] Wanjun Jiang, Xichao Zhang, Guoqiang Yu, Wei Zhang, Xiao Wang, M. Benjamin Jungfleisch, John E. Pearson, Xuemei Cheng, Olle Heinonen, Kang L. Wang, Yan Zhou, Axel Hoffmann, and Suzanne G. E. te Velthuis. Direct observation of the skyrmion hall effect. *Nature Physics*, 13:162 EP –, Sep 2016. Article.

[199] Siying Huang, Chao Zhou, Gong Chen, Hongyi Shen, Andreas K. Schmid, Kai Liu, and Yizheng Wu. Stabilization and current-induced motion of antiskyrmion in the presence of anisotropic dzyaloshinskii-moriya interaction. *Phys. Rev. B*, 96:144412, Oct 2017.

[200] Junichi Iwasaki, Masahito Mochizuki, and Naoto Nagaosa. Universal current-velocity relation of skyrmion motion in chiral magnets. *Nature Communications*, 4:1463 EP –, Feb 2013. Article.

[201] Junichi Iwasaki, Masahito Mochizuki, and Naoto Nagaosa. Current-induced skyrmion dynamics in constricted geometries. *Nature Nanotechnology*, 8:742 EP –, Sep 2013. Article.

[202] Naoto Nagaosa and Yoshinori Tokura. Topological properties and dynamics of magnetic skyrmions. *Nature Nanotechnology*, 8:899 EP –, Dec 2013. Review Article.

[203] H. Fook, C. A. C. Ian, W. Gan, I. Purnama, and W. Lew. Mitigation of magnus force in current-induced skyrmion dynamics. In *2015 IEEE International Magnetics Conference (INTERMAG)*, pages 1–1, May 2015.

[204] J. Sampaio, V. Cros, S. Rohart, A. Thiaville, and A. Fert. Nucleation, stability and current-induced motion of isolated magnetic skyrmions in nanostructures. *Nature Nanotechnology*, 8:839 EP –, Oct 2013. Article.

[205] Riccardo Tomasello, Marco Ricci, Pietro Burrascano, Vito Puliafito, Mario Carpentieri, and Giovanni Finocchio. Electrical detection of single magnetic skyrmion at room temperature. *AIP Advances*, 7(5):056022, 2017.

[206] Spintronics. Accessed: 2019-07-29.

[207] X. S. Wang, H. Y. Yuan, and X. R. Wang. A theory on skyrmion size. *Communications Physics*, 1(1):31, 2018.

[208] Xing Chen, Wang Kang, Daoqian Zhu, Xichao Zhang, Na Lei, Youguang Zhang, Yan Zhou, and Weisheng Zhao. Skyrmion dynamics in width-varying nanotracks and implications for skyrmionic applications. *Applied Physics Letters*, 111(20):202406, 2017.

[209] Maverick Chauwin, Xuan Hu, Felipe Garcia-Sanchez, Neilesh Betrabet, Christoforos Moutafis, and Joseph S. Friedman. Conservative Skyrmion Logic System. *arXiv e-prints*, page arXiv:1806.10337, Jun 2018.

[210] Cesare Rossetti. *Rudimenti di meccanica quantistica*. Levrotto & Bella.

[211] Marcelo Alonso and Edward J. Finn. *Fundamental University Physics - III Quantum and Statistical Physics*. Addison Wesley.

[212] J. J. Sakurai. *Modern quantum mechanics*. Addison Wesley, 1994.

[213] Jun John Sakurai. *Advanced quantum mechanics*. Addison Wesley, 1967.

[214] L. Schiff. *Quantum Mechanics*. Mc Graw-Hill, 4th edition.

[215] Sheldon M. Ross. *Introduction to probability and statistics for engineers and scientists*. Elsevier academic press, 3rd edition.

[216] R. L. Allen and D. W. Mills. *Signal analysis. Time, Frequency, Scale and Structure*. IEEE Press - Wiley Interscience.

[217] L. Lo Presti and F. Neri. *L'analisi dei segnali*. CLUT, 2nd edition.

[218] Omar Manasreh. *Semiconductor Heterojunctions and Nanostructures (Nanoscience and Technology)*. McGraw-Hill.

[219] Frank Jensen. *Introduction to computational chemistry*. John Wiley & sons, 2nd edition.

[220] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[221] Cohen E.R., T. Cvitas, J.G. Frey, B. Holmström, K. Kuchitsu, R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H.L. Strauss, M. Takami, and A.J. Thor. IUPAC & RSC Publishing.

[222] Chao Yang, Juan C. Meza, Byounghak Lee, and Lin-Wang Wang. KSSOLV - a MatLab toolbox for solving the kohn–sham equations. *ACM Transactions on Mathematical Software*, 36(2), 2009.

[223] Jack Simons. *An introduction to theoretical chemistry*. Cambridge University Press.

[224] Lecture III : The Many-Body Hamiltonian and the Functional Derivative. `http://www.physics.metu.edu.tr/~hande/teaching/741-lectures/lecture-03.pdf`.

[225] Ferdows Zahid, Magnus Paulsson, and Supriyo Datta. Electrical conduction through molecules. In *Advanced Semiconductors and Organic NanoTechniques*. Academic Press, 2003.

[226] Kurt Stokbro, Dan Erik Petersen, Søren Smidstrup, Anders Blom, and Mads Ipsen. Semiempirical model for nanoscale device simulations. *Physical review B*, 82(075420), 2010.

[227] Theodore L. Brown, Jr. LeMay Eugene H., Bruce E. Bursten, Catherine J. Murphy, and Patrick M. Woodward. *Chemistry: The Central Science*. Pearson College Div, 14th edition.

[228] Tro Nivaldo. *Chemistry: A Molecular Approach*. Pearson Education, 4th edition.

[229] Ilya G. Kaplan. *Intermolecular Interactions: Physical Picture,Computational Methods and Model Potentials*. 2006.

[230] Mark Waller and Stefan Grimme. *Weak Intermolecular Interactions: A Supermolecular Approach*, pages 1–27. Springer Netherlands, Dordrecht, 2016.

[231] Sujay B. Desai, Hossain M. Fahad, Theodor Lundberg, Gregory Pitner, Hyungjin Kim, Daryl Chrzan, H.-S. Philip Wong, and Ali Javey. Gate quantum capacitance effects in nanoscale transistors. *Nano Letters*, 19(10):7130–7137, 2019. PMID: 31532995.

[232] Michael Galperin, Mark A. Ratner, Abraham Nitzan, and Alessandro Troisi. Nuclear coupling and polarization in molecular transport junctions: Beyond tunneling to function. *Science*, 319(5866):1056–1060, 2008.

[233] William B. Davis, Walter A. Svec, Mark A. Ratner, and Michael R. Wasielewski. Molecular-wire behaviour in p-phenylenevinylene oligomers. *Nature*, 396:6706, 1998.

[234] Seong Ho Choi, BongSoo Kim, and C. Daniel Frisbie. Electrical resistance of long conjugated molecular wires. *Science*, 320(5882):1482–1486, 2008.

[235] C. Toher, A. Filippetti, S. Sanvito, and Kieron Burke. Self-interaction errors in density-functional calculations of electronic transport. *Phys. Rev. Lett.*, 95:146402, Sep 2005.

[236] Stephan Kümmel, Leeor Kronik, and John P. Perdew. Electrical response of molecular chains from density functional theory. *Phys. Rev. Lett.*, 93:213002, Nov 2004.

[237] Søren Smidstrup, Troels Markussen, Pieter Vancraeyveld, Jess Wellendorff, Julian Schneider, Tue Gunst, Brecht Verstichel, Daniele Stradi, Petr A Khomyakov, Ulrik G Vej-Hansen, et al. Quantumatk: An integrated platform of electronic and atomic-scale modelling tools. *J. Phys: Condens. Matter*, 32:015901, 2020.

[238] QuantumATK version Q-2019.12, Synopsys QuantumATK www.synopsys.com/silicon/quantumatk.html.

[239] Luiz G. Ferreira, Marcelo Marques, and Lara K. Teles. Approximation to density functional theory for the calculation of band gaps of semiconductors. *Phys. Rev. B*, 78:125116, Sep 2008.

[240] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton. Electron-energy-loss spectra and the structural stability of nickel oxide: An lsda+u study. *Phys. Rev. B*, 57:1505–1509, Jan 1998.

[241] Matteo Cococcioni and Stefano de Gironcoli. Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. *Phys. Rev. B*, 71:035105, Jan 2005.

[242] Zhizhou Yu, Jian Chen, Lei Zhang, and Jian Wang. First-principles investigation of quantum transport through an endohedral n@c60in the coulomb blockade regime. *Journal of Physics: Condensed Matter*, 25(49):495302, nov 2013.

[243] Antonio Martinez, John R Barker, and Riccardo Di Pietro. Dissipative non-equilibrium green function methodology to treat short range coulomb interaction: current through a 1d nanostructure. *Journal of Physics: Condensed Matter*, 30(29):294003, jun 2018.

[244] Alexander Altland and Ben Simons. *Condensed matter field theory*. Cambridge University Press, 2010.

[245] Henrik Bruus and Karsten Flensberg. *Many-Body Quantum Theory in Condensed Matter Physics: An Introduction*. OUP Oxford, 2004.

[246] A. Nasri, A. Boubaker, W. Khaldi, B. Hafsi, and A. Kalboussi. Transport propreties of organic single electron transistor; dependence on acene length. In *2017 International Conference on Engineering MIS (ICEMIS)*, pages 1–7, 2017.

[247] J. J. Palacios. Coulomb blockade in electron transport through a c_{60} molecule from first principles. *Phys. Rev. B*, 72:125424, Sep 2005.