Predicting solid-state qubit candidates

by

Oliver Lerstøl Hebnes

Thesis

for the degree of

Master of Science

Faculty of Mathematics and Natural Sciences
University of Oslo

December 18, 2020

Contents

Part I

Theory

Chapter 1

Quantum platforms from semiconductors

In this chapter we will give a brief overview of the current domain in quantum technological advances. This will not only give us insights in how the technology is being used today, but also grant us the opportunity to discuss key concepts that are fundamental to understand for this thesis. Importantly, it will motivate the reasoning for finding new materials that can be used in devices for this purpose. Thereafter we will look into how materials are build up, and what the characteristics of a semiconductor is.

1.1 Quantum technologies

Quantum technology (QT) refers to practical applications and devices that utilise the principles of quantum physics as a foundation. Technologies in this specter are based on concepts such as *superposition*, *entanglement* and *coherence*.

A quantum superposition refers to any two or more quantum states can be added together into another valid quantum state, such that every quantum state can be represented as a sum, or a superposition, of two or more distinct state. This is according to the wave-particle duality which phrase that every particle or another quantum entity may be described as either a particle or a wave. When measuring a superposition, however, the system falls back to one of the basis states that formed the superposition, destroying the original configuration.

A quantum entanglement refers to when a two- or many-particle state cannot be expressed independently of the state of the other particles, even when the particles are separated by a significant distance. As a result, the many-particle state is termed an entangled state [1].

Quantum coherence arise if two waves coherently interfere with each other and generate a superposition of the two states. Likewise, loss of coherence is known as *decoherence*.

The terms superposition, entanglement and coherence are closely related. The two latter terms are the primary features of quantum mechanics that are lacking in classical mechanics, giving rise to paradoxes such as Einstein's famous Schrödingers cat, which is both dead and alive at the same time when in its coherent state inside a closed box.

Additionally, another concept that the reader should be familiar with is the famous Heisenberg uncertainty principle. It states that

$$\sigma_x \sigma_p \leqslant \frac{\hbar}{2}, \tag{1.1}$$

where σ_x is the standard deviation for the position and σ_p is the standard deviation in momentum. This means that we cannot accurately predict both the position and momentum of the same particle at the same time. Thus, we often calculate the probability for a particle to be in a state which results in imaginable concepts such as an electron sky surrounding a neutron. However, it is often forgotten that equation (1.1) is an inequality, which means that it is possible to create a state where both the position and momentum is not-well defined, opposed to having both well defined.

1.1.1 Quantum communication

Quantum communication refers to the transfer of a state of one atom to another. This can be done with the use of a photon carrying a quantum state, which is termed a *flying qubit*. A significant portion of quantum communication is also part of other quantum information theories such as quantum cryptography, which comes as a remedy to a rising paranoia concerning security [2, 3].

It is difficult to avoid someone eavesdropping a conversation, however, the problem fades away if the eavesdropper does not speak the language, which keeps the content of the conversation safe. Herein, we will refer to the conversation as a public key, and the content of the conversation as a private key. Albeit, we live in a digital world where most of our actions are increasingly being stored as information, and we could imagine that the eavesdropper in the latter example stored the conversation. Even if the content of the conversation was encrypted with a private key, it still presents a challenge, since encrypted information stored today could be deciphered in ten or twenty years' time[1]. In addition, the digital worlds encryption keys

[1]As an example, in Martin Gardner's *Scientific American* column in 1976 [4], the 129-digit RSA key was thought to be safe for 5000 mips (million instructions per second) years, equal to 4×10^{25} years. Only 17 years later, the factorization was a reality and the public key was

could be broken anytime by an imaginable complex algorithm, which could result in Armageddon. Consequently, finding an encryption method that could make information either impossible to eavesdrop or make the security unbreakable forever is very desirable. This is the ultimate goal of quantum cryptography [3].

The idea is to pass information in the form of a quantum bit, or *qubits* for short. They are the building blocks of quantum computers, and as opposed to a conventional 0- or 1-bits that classical computers are based on, they can inhabit any superposition of the states 0 or 1. This approach is termed the *quantum key distribution* (QKD) [7, 8]. Now, if a wild eavesdropper would try to measure the information, the nature of quantum physics tells us that the original configuration would be destroyed and the receiver would be alerted of the eavesdropper. Furthermore, if the eavesdropper would try to make a copy of the message, the copying itself would be limited of the no-cloning theorem [7] which declare that quantum states cannot be copied.

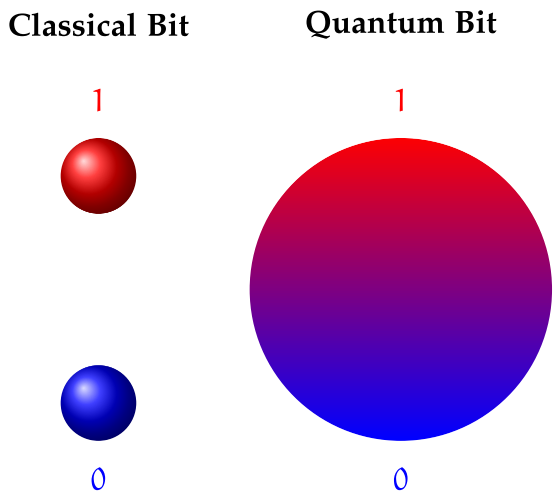


Figure 1.1: Conceptual illustration of the two-level classical bit, which are restricted to the boolean states 1 (true) or 0 (false), and the quantum bit that can be in any superposition of the states 0 or 1.

A clever approach to ensure confidentiality is to send the encryption key of the information before sending the actual encrypted information. If the key is received unperturbed then no one knows the key and it can be safely used. If it turns out perturbed, confidentiality is still intact since the key does not contain any information and can be discarded [7]. It should be noted that this requires both the sender and receiver to not only have each their quantum computer to store qubits in its memory, but also that they will need to initially exchange a common secret which is later expanded, making quantum key *expansion* a more exact term for QKD [3, 8].

Most applications and experiments use optical fibers for sending information via photons, with the distance regarded as the main limitation. This is

revealed to be "The Magic Words are Squeamish Ossifrage" [5]. To compare to todays fastest supercomputer Fugaku [6] by making two *very* rough estimations that flops and mips are approximately the same in addition to solely basing the calculation on computing power, the 129-digit RSA private key would be found in less than half a second using Fugaku.

reasoned by that classical repeaters are unable to enhance quantum information because of the no-cloning theorem, making photon loss in optical fiber cables inevitable. Thus, quantum communication must reinvent the repeater concept, using hardware that preserves the quantum nature [9]. Nonetheless, secure QKD up to 400km has recently been demonstrated using optical fibres in academic prototypes [10].

1.1.2 Quantum sensing

Measurements are part of our digital world today to a great extent. There would be no way to exchange goods, services or information without some extent of reliable and precise measurements [9]. Thus, improving the accuracy of sensors for every measurement done is desirable. One method to improve measurements can be by utilising quantum sensors, which is the use of a quantum property, such as a quantum object or quantum coherence, to measure a physical quantity [11]. This is possible because quantum systems are easily receptive to pertubations of its surroundings due to decoherence, and can be used to detect physical properties such as time, rotations, temperature and pressure [11].

For a quantum system to be able to function as a quantum sensors, a few of the criterias formulated by DiVincenzo has to be met [12]. Firstly, the quantum system needs to have discrete and resolvable energy levels. The quantum system also needs to be controllably initialised into a state that can be identified and coherently manipulated by time-dependent fields. Lastly, the quantum system needs to be able to interact with the physical property one wants to measure through a coupling parameter [11].

It is also possible to also exploit quantum entanglement to improve the precision of a measurement. This gain of precision is used to reach what is called the Heisenberg-limit, which states that the precision scales as the number of particles N in an idealized quantum system [9, 11], while the best classical sensors scale with $\sqrt{N}$.

1.1.3 Quantum computation

The start of the digital world's computational powers can be credited to Alan Turing. In 1937, Turing [13] published a paper where he described the *Turing machine*, which is regarded as the foundation of computation and computer science. It states that only the simplest form of calculus, such as boolean Algebra (1 for true and 0 for false), is actually computable. This required developing hardware that could handle classical logic operations, and was the basis of transistors that are either in the state ON or OFF depending on the electrical signal. Equipped with a circuit consisting of wires and transistors, commonly known as a computer, we could develop software to solve all kinds of possible applications.

Driven by the development of software, conventional computers has, in accordance to Moore's law [14], doubled the amount of transistors on integrated circuit chips every two years as a result of smaller transistors. Furthermore, the clock frequency has enhanced with time, resulting in doubling of computer performance every 18 months [3]. Alas, miniaturization cannot go on forever as transistors are mass-produced at 5nm today and are expected to reach a critical limit of 3nm in the following years [15].

To sustain the digital worlds increasing computational demand, other alternatives than the conventional classical computer must be explored. This is where quantum computation comes into the picture, and the term quantum computer, which is a device that would utilise quantum nature to solve certain computational problems more efficiently than allowed by Boolean logic [16]. Since a qubit has a quantum nature and is the counterpart to the classical bit, it naturally follows that it is in the center of attention in quantum computation[2].

The architecture of a quantum computer is dependent on a set of quantum logic gates that perform unitary transformations on sets of qubits [12, 17]. Other implementations of quantum computers exists, such as the adiabatic quantum computer. This approach is not based on gates, but on defining the answer of a problem as the ground state of a complex network of interactions between qubits, and then controlling the interactions to adiabatically evolve to the ground state [18].

It has been demonstrated that exponentially complex problems can be reduced to polynomially complex problems for quantum computers [3]. For example, a quantum search algorithm found by Grover [19] offers a quadratic speed-up compared to classical algorithms, which initally is not a huge speed-up but indeed has a limitless spectrum of applications. Additionally, Shor's quantum integer factorization algorithm [20] present an exponential speed-up, and can be used to break the widely used RSA-scheme in cryptography. Even more impressively, Google reported in 2019 that they ran a random number generator algorithm using 53 qubits on a superconducting processor in 200 seconds which would take 10.000 years for a classical computer [21]. It is expected that quantum computers will excel in exceedingly complex problems, while many simpler tasks may not see any speed-up at all. Hence, it is not expected that quantum computers will run classical computers out of business, but rather that they will coexist for each their purpose. However, it is not to avoid that both quantum hardware and -software is in its preliminary phase of development, and it will be interesting to follow up the progression

[2]There exists other systems such as the quantum d-state system, known as *qudits*, that can also be utilised in quantum computation [17].

the following years.

Quantum computing is a highly sought-after goal, but there are extensive challenges that needs to be adressed. Controlling a complex many-qubit system is difficult, since it is not always possible to establish interactions between qubits [12]. Additionally, decoherence and other quantum noise occurs as a result to the high volatility of quantum states, making quantum state manipulation prone to errors. The *quantum error correction* protocols and the theory of *threshold theorem* deals with this vulnerability, stating that noise most likely does not pose any fundamental barrier to the performance of large-scale computations [3].

1.1.4 Quantum computing requirements

As ever-promising the concepts of quantum technology are, it still remains out of reach for the everyday man. we still have not seen quantum computer spread out commercially available. This is not without reason, since there are several challenges that needs to be overcome before it can be commercially available. A few of them have already been mentioned, but we will here concretize critical principles for a physical implementation of quantum platforms.

> "I always said that in some sense, these criteria are exactly the ones that you would teach to kindergarten children about computers, quantum or otherwise" DiVincenzo [22]

DiVincenzo formulated in the year of 2000 seven basic criteria for the physical realization of a quantum platform [12].

1. A scalable physical system with well characterized qubits
2. The ability to initialize the state of the qubits to a simple initial system
3. Have coherence times that are much longer than the gate operation time
4. Have a universal set of quantum gates
5. Have the ability to perform qubit-specific measurements
6. The ability to convert stationary qubits to flying qubits
7. The ability to faithfully transmit flying qubits between specified locations

The first five criteria (1-5) has to be met for a quantum platform to be considered a quantum computer, while the two last criteria (6-7) was added for quantum communication, since its applications provide a unique advantage compared to its classical counterpart.

1.1.5 Available quantum platforms

Many different quantum platforms have been physically implemented, and this section will serve as a brief overview of the current status. For a more thorough review of qubit-implementations, the reader is recommended to read Ref. [9].

Superconducting circuits can be used in quantum computing, since electrons in superconducting materials can form Cooper pairs through an electron-phonon interaction when the temperature is lower than a critical limit. Lower than the limit, electrons can move without resistance in the material [23]. Exploiting this intrinsic coherence, qubits can be made by forming microwave circuits based on loops of two superconducting elements separated by an insulator, also known as Josephson tunnel junctions [9]. Today, superconducting Josephson junctions are the most promising quantum platform, but it requires very low temperature (mK) to function as a quantum computer, making it costly to use. Additionally, the current devices experience a relatively short coherence time, resulting in a challenging upscaling for further experiments.

Another eligible quantum platfom is single photons that can be implemented as one-qubit gates by rotations of the photon polarization. Its use in fiber optics are less prone to decoherence, but faces challenges since the more complex photon-photon entanglement and control of multi-qubits is strenuous [17].

Another implementation of a quantum platform can be done by fixing the nuclear spin of solid-state systems, resulting in long spin coherence. Thus, one can manipulate qubits utilising electromagnetic fields, making one-qubit gates possible.

An alternative platform is the isolated atom platform which is characterized by its well-defined atom isolation. Here, every qubit is based on energy levels of a trapped ion or atom. Quantum entanglement can be enabled through laser-induced spin coupling, however scaling it up to large numbers induce problems in controlling large systems and cooling of the trapped atoms or ions.

Quantum dots are also in use as quantum platforms, and can be imagined as artifical atoms which are confined in a solid-state host. It possess the similar coherence potential as the isolated atom platform, but without the drawback of confining and cooling of a given atom or ion. One approach to generate a quantum dot is to trap a hole or an electron in a semiconductor's discrete energy level potential. In such a system, the spin degree of freedom is considered auspicious due to its long coherence time [9].

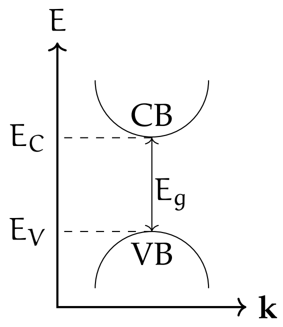
Lastly, we will mention point defects in semiconductors as a physical implementation of a quantum platform. It shares many of the attributes of quantum dots, such as discrete optical transitions and controllable coherent spin states, but it is vulnerable to small changes in the lattice of the semiconductor. However, one can utilise the strength of the solid-state semiconductor host to isolate to some extent the point defect, rewarding extended coherence time. Thus, we will from here on focus our attention on qubit-host qandidates that can fascilitate such defects.

1.2 First principles of semiconductors

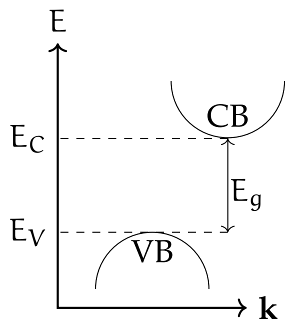
Isolated atoms have distinct energy levels, where the Pauli exlusion principle [24] states for fermions that each energy level can accomodate two electrons of opposite spin. In a solid, the discrete energy levels of the isolated atom spread into continuous energy bands for the solid since the wavefunctions of the electrons in the neighboring atoms overlap. Hence, an electron is not neccessarily localized at a particular atom anymore as it would for a system with an isolated atom. Every material has a unique band structure, similar to every human have their unique fingerprint.

Which energy bands that are occupied by electrons, is the key in understanding the electrical properties of solids. The highest occupied electron band at 0 K is called the valence band (VB), while the lowest unoccupied electon band is called the conduction band (CB). In between the two bands, we find an area that contains no electron energy states which is known as the band gap and its energy is denoted as E_g.

To be able to accelerate electrons in a solid using an electrical field, they must be able to move into new energy states. At 0 K, the entire valence band of a semiconductor is full with electrons and no available new states, making it impossible to flow current through the material. This can be solved by using either thermal or optical energy to excite energies from the valence band to the conduction band, in order to *conduct* electricity. In room temperature, many semiconductors will be able to have excited electrons in the conducting band solely from thermal energy matching the energy band gap [25].



(a)



(b)

Figure 1.2: A schematic drawing of a (a) direct bandgap and an (b) indirect bandgap.

In some scenarios, thermal or optical energy is not sufficient for an excitation since the energy bands are also dependent by a crystal momentum. A difference in the momentum of the minimal-energy state in the conduction band and the maximum-energy state in the valence band, is known to be an *indirect bandgap* as seen in figure 1.2a. If there is no difference at all, the material has a *direct bandgap*, which is visualized in figure 1.2a.

1.3 A brief overview of material science

The interactions in atomic structure and characteristics of matter is the foundation of material science. The applications of material science are unlimited, and if the reader takes a quick look around, hen will observe that every artifical material is made for a purpose, either being a bottle of water or a chair to sit in.

Solid materials, like a plastic bottle, are formed by densely packed atoms. These atoms can randomly occur through the material without any long-range order, which would categorize the material as an *amorphous solid*. Amorphous solids are frequently used in gels, glass and polymers [25].

However, the atoms can also be periodically ordered in small regions of the material, classifying the material as a *polycrystalline solid*. All ceramics are polycrystalline with a broad specter of applications ranging from kitchen-porcelain to orthopedical bio-implants [26].

A third option is to have these atoms arranged with infinite periodicity, making the material a *crystalline solid* or more commonly named a *crystal*. The three options are visualised in figure 1.3.

The periodicity in a crystal is defined in terms of a symmetric array of points in space called the *lattice*, which can be simplified as either a one-dimensional array, a two-dimensional matrix or a three dimensional vector space, depending on the material. At each lattice point we can add an atom to make an arrangement called a *basis*. The basis can be one atom or a cluster of atoms having the same spatial arrangement. For every crystal, there exists periodically repeated building blocks called *cells* which represents the entire crystal. The smallest cell possible is called a *primitive cell*, but such a cell only allows lattice points at its corners and it is often quite rigid to work with when the structure becomes complex. As a solution, we will consider the *unit cell*, which allows lattice points on face centers and body centers.

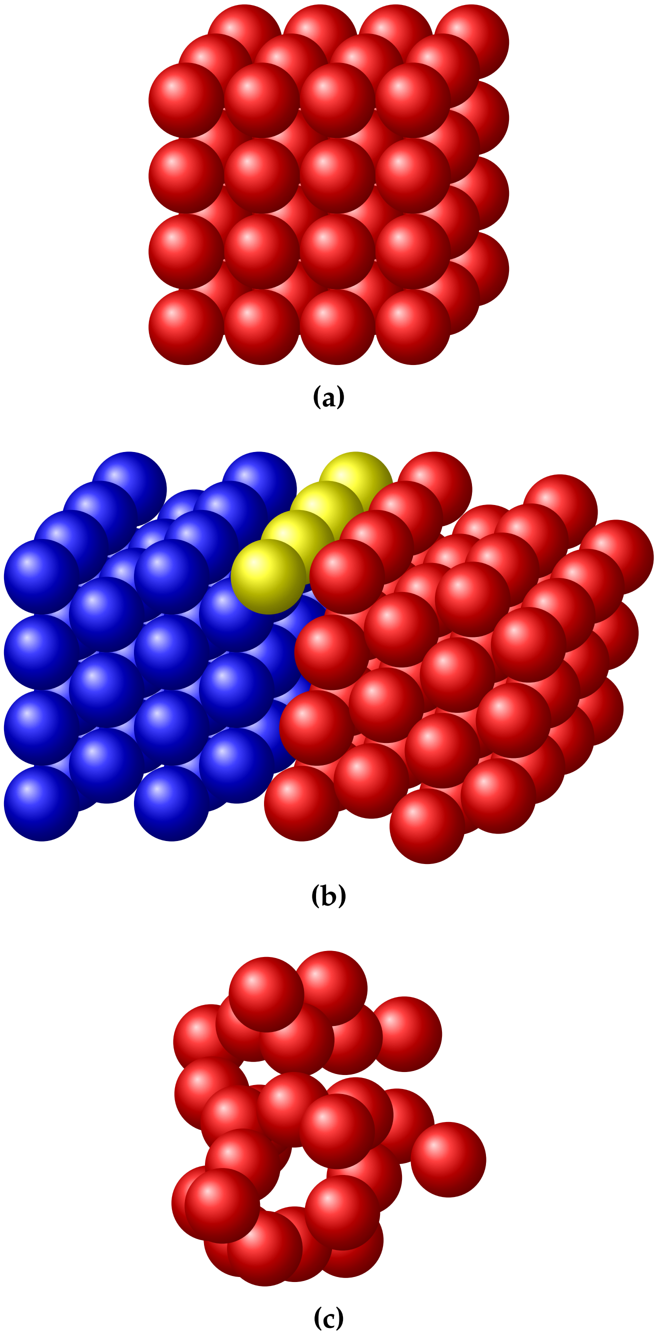


Figure 1.3: Different degrees of ordered structures, where (a) is a crystalline of a simple cubic lattice, (b) is a polycrystalline of a hexagonal lattice, and (c) is an amorphous.

One known crystal structure is the perovskite structure. Compounds with this structure are characterized having an ABX_3 stoichiometry whose symmetri belong to one of 15 space groups identified by Lufaso & Woodward [27], such as the cubic, orthorombic and tetragonal, to name a few. For our purpose, we will be looking into when the C atom is oxygen, herein we will refer to the oxygen-perovskite ABO_3. The A atom is nine- to 12-fold coordinated by oxygen, while the B atom is sixfold coordinated by oxygen, and the BO_6 octahedra are connected to the corners in all three directions.

The motivation behind the research of perovskites is reasoned by that there are a large amount of possible ABO_3 chemistries, whereas a significant portion of these that takes the perovskite structure. They have a broad specter of applications, ranging from high-temperature superconductors [28], ionic conductors [29], and multiferroic materials [30]. Additionally, adding a perovskite structured compound to solar cells has reportedly resulted in higher performance efficiency while being cheap to produce and simple to manufacture [31, 32], however, that includes the use of hybrid organic-inorganic compounds and excludes the use of oxygen.

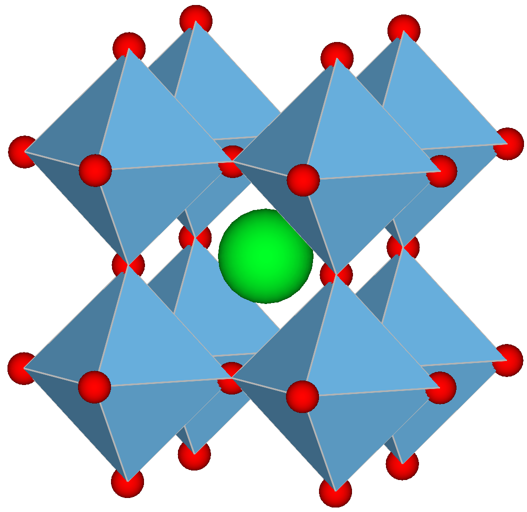


Figure 1.4: A crystal structure of $SrTiO_3$ which is a cubic perovskite. The red atoms are oxygen, whereas the green atom is strontium, and inside every corner-sharing BO_6 octahedral unit is a titanium atom.

1.4 Qubit host qandidates

Weber *et al.* [16] proposed in 2010 four criteria that should be met for a solid-state semiconductor material hosting a qubit defect, in particular the diamond nitrogen-vacancy (NV^{-1}). An ideal crystalline host should have [16]

(H1) A wide band gap to accomodate a deep center.

(H2) Small spin-orbit coupling in order to avoid unwanted spin flips in the defect bound states.

(H3) Availability as high-quality, bulk, or thin-film single crystals.

(H4) Constituent elements with naturally occuring isotopes of zero nuclear spin.

Host candidates that satisfy criteria (H1) and (H2) can be found by studying the electronic structure of the compounds. What tends to happen for a defect in semiconductors is that the electronic states associated with the defect have energies that lie within the (forbidden) band gap of the semiconductor. This is schematically drawn and explained for diamond in figure (1.5). A large band gap can accomodate multiple highly isolated states and make them confined and isolated from interactions, which is favorable for longer coherence times. However, the desired band gap depends on the application, since the Fermi level of large band gap materials is challenging to control [33].

Table (1.1) (adapted from [16]) lists several host candidates that has a promising band gap that can accomodate a deep center defect. The spin-orbit splitting is an indication of the strength of the spin-orbit interaction, and is taken at the Γ point from the valence-band splitting. A smaller value indicates less susceptibility to decoherence.

Criteria (H3) is important for scalability and further potential for a large-scale fabrication. The given qandidate hosts in table (1.1) are all available and can be grown as single crystals, but with varying quality and size.

Normally, nuclear spin is a major source of decoherence for all semiconductor-based quantum technologies. This would exclude the use of all elements in odd groups in the periodic table, since these elements exhibit nonzero nuclear spin. This could help to increase the spin-coherence time of a paramgnetic deep center [16]. However, nuclear spin can also induce additional quantum degrees of freedom for applications in the right configuration [33]. Therefore, criteria 4 is not a strict requirement but is a general recommendation for reducing decoherence time.

In fact, all of the criteria could be regarded as recommendations

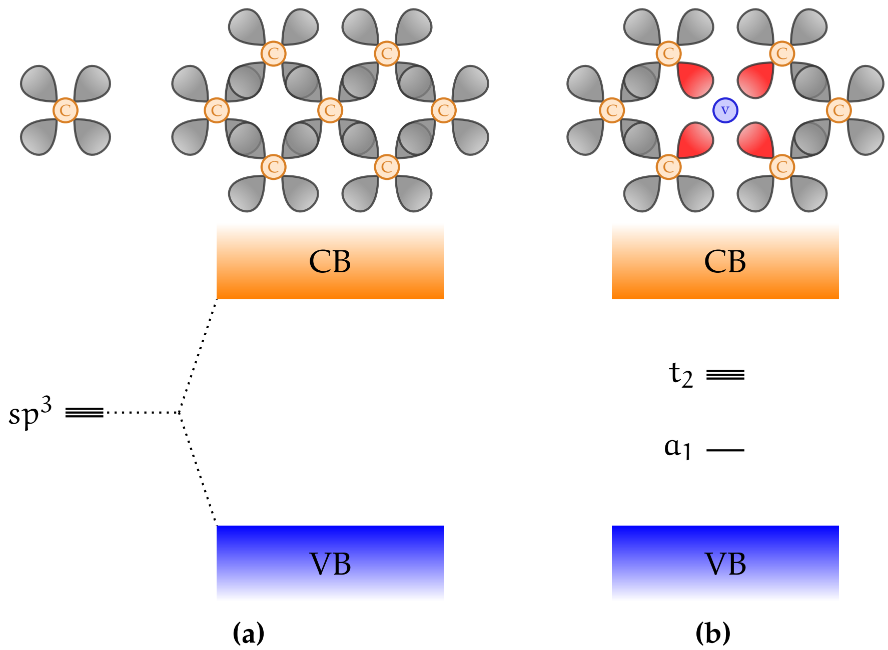


Figure 1.5: A schematic representation of the electronic structure of a point defect in a tetrahedrally coordinated semiconductor, which is in this case diamond. Figure (a) shows the electronic states that correspond to the difference for an isolated atom and a lattice of atoms, as a superposition of sp^3 orbitals generates valence and conduction bands. In figure (b) a vacancy has been created by removing a carbon atom, and the four orbitals interact with each other resulting in two new states with a_1 and t_2 symmetry. The carbon atoms interact stronger with each other than the remaining four orbitals, which is why the two new states rests within the semiconductor's band gap. Figure adapted from ref. [34].

Material	Band gap E_g (eV)	Spin-orbit splitting Δ_{so} (MeV)	Stable spinless nuclear isotopes?
3C-SiC	2.39	10	Yes
4C-SiC	3.26	6.8	Yes
6C-SiC	3.02	7.1	Yes
AlN	6.13	19	No
GaN	3.44	17.0	No
AlP	2.45	50	No
GaP	2.27	80	No
AlAs	2.15	275	No
ZnO	3.44	-3.5	Yes
ZnS	3.72	64	Yes
ZnSe	2.82	420	Yes
ZnTe	2.25	970	Yes
CdS	2.48	67	Yes
C (Diamond)	5.5	6	Yes
Si	1.12	44	Yes
GaAs	1.42	346	No

Table 1.1: Table taken from Weber *et al.* [16] that lists a number of tetrahedrally coordinated hosts whose band gaps are larger than 2.0 (eV), and compares it to diamond, Si and GaAs. For materials where more than one structure is stable in room temperature, the most dominant room-temperature phase's band parameters are chosen.

Chapter 2

Introduction to density functional theory

To fully understanding the underlying physics behind computational material science, we will need to investigate how we can calculate the forces acting inside a crystal. Since these forces are happening on a microscopic scale, we will need to utilise the theory of quantum mechanics.

In this chapter, we will ony summarize the neccessary theory behind density functional theory, leaving most of the quantum-mechanical world untouched. However, the fundamental theory remains the same and we will start our venture with the Schrödinger equation.

2.1 The Schrödinger equation

In principle, we can describe all physical phenomenas of a system with the wavefunction $\Psi(\mathbf{r}, t)$ and the Hamiltonian $\hat{H}(\mathbf{r}, t)$, where $\mathbf{r}$ is the spatial position and t is the time. Unfortunately, analytical solutions for the the time-dependent Schrödinger equation,

$$i\hbar \frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = \hat{H}(\mathbf{r}, t)\Psi(\mathbf{r}, t), \tag{2.1}$$

are extremely rare. More conveniently, we can generate a general wavefunction by a summation of eigenfunctions,

$$\Psi(\mathbf{r}, t) = \sum_{\kappa} c_\kappa \psi_\kappa(\mathbf{r}, t), \tag{2.2}$$

where c_κ is a constant and ψ_κ is the κ-th eigenfunction. A general wavefunction does not neccessarily describe stationary states, and consequently does not have distrinct energies but is rather represented statistically from the expectation value

$$E = \sum_{\kappa} |c_\kappa| E_\kappa. \tag{2.3}$$

Solving the Schrödinger equation for a general wavefunction is rather troublesome, but luckily we can use the eigenfunctions instead, transforming equation 2.1 into the time-independent Schrödinger equation for eigenfunctions

$$\hat{H}\psi_\kappa(\mathbf{r}) = E_\kappa \psi_k(\mathbf{r}), \tag{2.4}$$

where E_κ is the eigenvalue of the κ-th eigenstate $\psi_\kappa(\mathbf{r})$. The eigenfunctions have distinct energies, and the state with the lowest energy is called the ground state. They have the attribute that they are orthogonal and normalized with respect to

$$\left\langle \psi_\kappa\left(\mathbf{r}\right) | \psi_{\kappa'}\left(\mathbf{r}\right) \right\rangle = \delta_{\kappa\kappa'}. \tag{2.5}$$

The symmetry of an eigenfunction depends on the symmetry of the potential $V_{ext}(\mathbf{r})$ and the boundary conditions [35].

2.2 The many-particle Schrödinger equation

As we extend the theory to include many-particle systems, we will gradually explain and add the different contributions that make up the many-body Hamiltonian. During this process, we will neglect any external potential applied to the system.

If we place a simple electron with mass m_e in its own system, it will be in possession of kinetic energy. Instead of just one electron, we can place N_e electrons, and they will together have the total kinetic energy

$$T_e = -\sum_{j=1}^{N_e} \frac{\hbar^2 \nabla_j}{2m_e}. \tag{2.6}$$

All the electrons are negatively charged, causing repulsive Coulomb interactions between each and every electron, totalling to

$$U_{ee} = \sum_{j=1}^{N_e} \sum_{j'<j} \frac{q^2}{|r_j - r_{j'}|}. \tag{2.7}$$

The summation voids counting each interaction more than once. Simultaneously, we can place N_n nuclei with mass m_n in the same system, accumulating the kinetic energy

$$T_n = -\sum_{a=1}^{N_n} \frac{\hbar^2 \nabla_a}{2m_n}. \tag{2.8}$$

As in the example with electrons, the nuclei are also experiencing repulsive interactions between every single nucleus, adding up the total interactions as

$$U_{nn} = \sum_{a=1}^{N_n} \sum_{a'<a} \frac{q^2 Z_a Z_{a'}}{|R_a - R_{a'}|}. \tag{2.9}$$

where Z_a is the atom number of nuclei number a.

The system now contains N_e electrons and N_n nuclei, thus we need to include the attractive interactions between the them,

$$U_{en} = -\sum_{j=1}^{N_e} \sum_{a=1}^{N_n} \frac{q^2 Z_a}{|r_j - R_a|}. \tag{2.10}$$

Together, these equations comprise the time-independent many-particle Hamiltonian

$$\begin{aligned} \hat{H} = &-\sum_{j=1}^{N_e} \frac{\hbar^2 \nabla_j}{2m_e} - \sum_{a=1}^{N_n} \frac{\hbar^2 \nabla_a}{2m_n} + \sum_{j=1}^{N_e} \sum_{j'<j} \frac{q^2}{|r_j - r_{j'}|} \\ &+ \sum_{a=1}^{N_n} \sum_{a'<a} \frac{q^2 Z_a Z_{a'}}{|R_a - R_{a'}|} - \sum_{j=1}^{N_e} \sum_{a=1}^{N_n} \frac{q^2 Z_a}{|r_j - R_a|}. \end{aligned} \tag{2.11}$$

A few problems arise when trying to solve the many-particle Schrödinger equation. Firstly, the amount of atoms in a crystal is very, very massive. As an example, we can numerically try to calculate the equation 2.7 for a 1mm^3 silicon-crystal that contains $7 \cdot 10^{20}$ electrons. For this particular problem, we will pretend to use the current fastest supercomputer Fugaku [6] that can calculate 514 TFlops, and we will assume that we need 2000 Flops to calculate each term inside the sum [35], and we need to calculate it $N_e \cdot N_e/2$ times for the (tiny) crystal. The entire electron-electron interaction calculation would take $2.46 \cdot 10^{19}$ years to finish for a tiny crystal. Thus, the large amount of particles translates into a challenging numerical problem.

Secondly, the many-particle Hamiltonian contains operators that has to be applied to single-particle wavefunctions, and we have no prior knowledge of how Ψ depends on the single-particle wavefunctions ψ_κ.

2.3 The Born-Oppenheimer approximation

The many-particle eigenfunction describes the wavefunction of all the electrons and nuclei and we denote it as Ψ_κ^{en} for electrons (e) and nuclei (n), respectively. The Born-oppenheimer approximation assumes that nuclei, of substantially larger mass than electrons, can be treated as fixed point charges. According to this assumption, we can separate the eigenfunction into an electronic part and a nuclear part,

$$\Psi_\kappa^{en}(\mathbf{r},\mathbf{R}) \approx \Psi_\kappa(\mathbf{r},\mathbf{R})\Theta_\kappa(\mathbf{R}), \tag{2.12}$$

where the electronic part is dependent on the nuclei. This is in accordance with the assumption above, since electrons can respond instantaneously to a new position of the much slower nucleus, but this is not true for the opposite scenario. To our advantage, we already have knowledge of the terms in the many-particle Hamiltionian, and we can begin by separating the Hamiltionian into electronic and nuclear parts:

$$\hat{H}^{en} = \overbrace{T_e + U_{ee} + U_{en}}^{\hat{H}^e} + \overbrace{T_n + U_{nn}}^{\hat{H}^n}. \tag{2.13}$$

Starting from the Schrödinger equation, we can formulate separate expressions for the electronic and the nuclear Schrödinger equations.

$$H\hat{e}n\Psi_\kappa^{en}(\mathbf{r},\mathbf{R}) = E_\kappa^{en}\Psi_\kappa^{en}(\mathbf{r},\mathbf{R}) \quad |\times \int \Psi^*(\mathbf{r},\mathbf{R})d\mathbf{r} \tag{2.14}$$

$$\int \Psi_\kappa^*(\mathbf{r},\mathbf{R})(\hat{H}^e + \hat{H}^n)\Psi_\kappa(\mathbf{r},\mathbf{R})\Theta_\kappa(\mathbf{R})d\mathbf{r} = E_\kappa^{en} \underbrace{\int \Psi_\kappa^*(\mathbf{r},\mathbf{R})\Psi_\kappa(\mathbf{r},\mathbf{R})d\mathbf{r}}_{1}\Theta_\kappa(\mathbf{R}). \tag{2.15}$$

Since $\Theta_\kappa(\mathbf{R})$ is independent of the the spatial coordinates to electrons, we get E_κ as the total energy of the electrons in the state κ.

$$E_\kappa(\mathbf{R})\Theta_k(\mathbf{R}) + \int \Psi_k^*(\mathbf{r},\mathbf{R})H^n\Psi_k(\mathbf{r},\mathbf{R})\Theta_k(\mathbf{R})d\mathbf{r} = E_k^{en}\Theta_k(\mathbf{R}). \tag{2.16}$$

Now, the final integration term can be simplified by using the product rule, which results in

$$\Big(T_n + T_n^{'} + T_n^{''} + U_{nn} + E_\kappa(\mathbf{R})\Big)\Theta_\kappa(\mathbf{R}) = E_\kappa^{en}\Theta_\kappa(\mathbf{R}). \tag{2.17}$$

If we neglect T_n' and T_n'' to lower the computational efforts, we obtain the Born-Oppenheimer approximation with the electronic eigenfunction as

$$(T_e + U_{ee} + U_{en})\,\Psi_\kappa(\mathbf{r},\mathbf{R}) = E_\kappa(\mathbf{R})\Psi_\kappa(\mathbf{r},\mathbf{R}) \tag{2.18}$$

and the nuclear eigenfunction as

$$\big(T_n + U_{nn} + E_\kappa(\mathbf{R})\big)\,\Theta_\kappa(\mathbf{R}) = E_\kappa^{en}(\mathbf{R})\Theta_\kappa(\mathbf{r},\mathbf{R}). \tag{2.19}$$

How are they coupled, you might ask? The total energy in the electronic equation is a potential in the nuclear equation.

2.4 The Hartree and Hartree-Fock approximation

The next question in line is to find a wavefunction $\Psi(\mathbf{r},\mathbf{R})$ that depends on all of the electrons in the system. The Hartre [35] approximation to this is to assume that electrons can be described independently, suggesting the *ansatz* for a two-electron wavefunction

$$\Psi_\kappa(\mathbf{r}_1,\mathbf{r}_2) = A\cdot\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2), \tag{2.20}$$

where A is a normalization constant. This approximation simplifies the many-particle Shrödinger equation a lot, but comes with the downside that the particles are distinguishable and do not obey the Pauli exclusion principle for fermions.

The Hartree-fock approach, however, overcame this challenge and presented an anti-symmetric wavefunction that made the electrons indistinguishable [1]:

$$\Psi_\kappa(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}\Big(\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) - \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)\Big). \tag{2.21}$$

For systems containing more than one particles, the factor $1/\sqrt{2}$ becomes the Slater determinant and is used to normalize the wave function.

2.5 The variational principle

So far, we have tried to make the time-independent Schrödinger equation easier with the use of an *ansatz*, but we do not neccessarily have an adequate guess for the eigenfunctions and the ansatz can only give a rough estimate in most scenarios. Another approach, namely the *variational principle*, states that the energy of any trial wavefunction is always an upper bound to the exact ground state energy by definition E_0.

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle \leqslant \langle \psi | H | \psi \rangle = E \tag{2.22}$$

The eigenfunctions of H form a complete set, which means any normalized Ψ can be expressed in terms of the eigenstates

$$\Psi = \sum_n c_n \psi_n, \quad \text{where} \quad H\psi_n = E_n \psi_n \tag{2.23}$$

for all $n = 1, 2, \ldots$. The expectation value for the energy can be calculated as

$$\begin{aligned} \langle \Psi | H | \Psi \rangle &= \left\langle \sum_n c_n \psi_n \middle| H \middle| \sum_{n'} c_{n'} \psi_{n'} \right\rangle \\ &= \sum_n \sum_{n'} c_n^* c_{n'} \langle \psi_n | H | \psi_{n'} \rangle \\ &= \sum_n \sum_{n'} c_n^* E_n c_{n'} \langle \psi_n | \psi_{n'} \rangle \end{aligned}$$

Here we assume that the eigenfunctions have been orthonormalized and we can utilize $\langle \psi_m | \psi_n \rangle = \delta_{mn}$, resulting in

$$\sum_n c_n^* c_n E_n = \sum_n |c_n|^2 E_n.$$

We have already stated that Ψ is normalized, thus $\sum_n |c_n|^2 = 1$, and the expectation value conveniently is bound to follow equation 2.22. The quest to understand the variational principle can be summarized in a sentence - it is possible to tweak the wavefunction parameters to minimize the energy, or summed up in a mathematical phrase,

$$E_0 = \min_{\Psi \to \Psi_0} \langle \Psi | H | \Psi \rangle . \tag{2.24}$$

2.6 The density functional theory

Hitherto we have tried to solve the Schrödinger equation to get a ground state wave function, and from there we can obtain ground state properties, such as the ground state total energy. One fundamental problem that exists when trying to solve the many-electron Schrödinger equation is that the wavefunction is a complicated function that depends on $3N_e$ variables[1].

[1]not including spin

Hohenberg and Kohn [36] showed in 1964 that the ground-state density $n_0(r) = |\Psi_0(r)|$ determines a general external potential, which includes U_{en}, up to an additive constant, and thus also the Hamiltonian [37]. From another point of view, the theory states that all physical ground-state properties of the many-electron system are unique functionals of the density [35]. A consequence of this is that the number of variables is reduced from $3N_e$ to 3, significantly reducing the computational efforts.

However, the scheme is not without limitations, as the density functional theory (DFT) can only be used to find all the ground-state physical properties if the exact functional of the electron density is known. And 56 years after Hohenberg and Kohn published their paper, the exact functional still remains unknown.

We will start this chapter with a discussion of the Hohenberg-Kohn theorems, before we delve further into the Kohn-Sham equation.

2.6.1 The Hohenberg-Kohn theorems

THEOREM 1. *For any system of interacting particles in an external potential V_{ext}, the density is uniquely determined.*

PROOF. Assume that two external potentials $V_{ext}^{(1)}$ and $V_{ext}^{(2)}$, that differ by more than a constant, have the same ground state density $n_0(r)$. The two different potentials correspond to distinct Hamiltonians $\hat{H}_{ext}^{(1)}$ and $\hat{H}_{ext}^{(2)}$, which again give rise to distinct wavefunctions $\Psi_{ext}^{(1)}$ and $\Psi_{ext}^{(2)}$. Utilizing the variational principle, we find that no wavefunction can give an energy that is less than the energy of $\Psi_{ext}^{(1)}$ for $\hat{H}_{ext}^{(1)}$, that is

$$E^{(1)} = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle < \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle \tag{2.25}$$

and

$$E^{(2)} = \left\langle \Psi^{(2)} \middle| \hat{H}^{(2)} \middle| \Psi^{(2)} \right\rangle < \left\langle \Psi^{(1)} \middle| \hat{H}^{(2)} \middle| \Psi^{(1)} \right\rangle . \tag{2.26}$$

Assuming that the ground state is not degenerate, the inequality strictly holds. Since we have identical ground state densities for the two Hamiltonian's, we can rewrite the expectation value for equation 2.25 as

$$\begin{aligned}
E^{(1)} &= \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle \\
&= \left\langle \Psi^{(1)} \middle| T + U_{ee} + U_{ext}^{(1)} \middle| \Psi^{(1)} \right\rangle \\
&= \left\langle \Psi^{(1)} \middle| T + U_{ee} \middle| \Psi^{(1)} \right\rangle + \int \Psi^{*(1)}(\mathbf{r}) V_{ext}^{(1)} \Psi^{(1)}(\mathbf{r}) d\mathbf{r} \\
&= \left\langle \Psi^{(1)} \middle| T + U_{ee} \middle| \Psi^{(1)} \right\rangle + \int V_{ext}^{(1)} n(\mathbf{r}) d\mathbf{r} \\
&< \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle \\
&= \left\langle \Psi^{(2)} \middle| T + U_{ee} + U_{ext}^{(1)} + \overbrace{U_{ext}^{(2)} - U_{ext}^{(2)}}^{0} \middle| \Psi^{(2)} \right\rangle \\
&= \left\langle \Psi^{(2)} \middle| T + U_{ee} + U_{ext}^{(2)} \middle| \Psi^{(1)} \right\rangle + \int \left(V_{ext}^{(1)} - V_{ext}^{(2)} \right) n(\mathbf{r}) d\mathbf{r} \\
&= E^{(2)} + \int \left(V_{ext}^{(1)} - V_{ext}^{(2)} \right) n(\mathbf{r}) d\mathbf{r}.
\end{aligned}$$

Thus,

$$E^{(1)} = E^{(2)} + \int \left(V_{ext}^{(1)} - V_{ext}^{(2)} \right) n(\mathbf{r}) d\mathbf{r} \tag{2.27}$$

A similar procedure can be performed for $E^{(2)}$ in equation 2.26, resulting in

$$E^{(2)} = E^{(1)} + \int \left(V_{ext}^{(2)} - V_{ext}^{(1)} \right) n(\mathbf{r}) d\mathbf{r}. \tag{2.28}$$

If we add these two equations together, we get

$$\begin{aligned}
E^{(1)} + E^{(2)} < E^{(2)} + E^{(1)} &+ \int \left(V_{ext}^{(1)} - V_{ext}^{(2)} n(\mathbf{r}) d\mathbf{r} \right) \\
&+ \int \left(V_{ext}^{(2)} - V_{ext}^{(1)} n(\mathbf{r}) d\mathbf{r} \right) \\
E^{(1)} + E^{(2)} < E^{(2)} + E^{(1)}, &
\end{aligned} \tag{2.29}$$

which is a contradiction. Thus, the two external potentials cannot have the same ground-state density, and $V_{ext}(\mathbf{r})$ is determined uniquely (except for a constant) by $n(\mathbf{r})$. □

THEOREM 2. *There exists a variational principle for the energy density functional such that, if* n *is not the electron density of the ground state, then* $E[n_0] < E[n]$.

Proof. Since the external potential is uniquely determined by the density and since the potential in turn uniquely determines the ground state wavefunction (except in degenerate situations), all the other observables of the system are uniquely determined. Then the energy can be expressed as a functional of the density.

$$E[n] = \overbrace{T[n] + U_{ee}[n]}^{F[n]} + \overbrace{U_{en}[n]}^{\int V_{en} n(r) dr} \tag{2.30}$$

where $F[n]$ is a universal functional because the treatment of the kinetic and internal potential energies are the same for all systems, however, it is most commonly known as the Hohenberg-Kohn functional.

In the ground state, the energy is defined by the unique ground-state density $n_0(r)$,

$$E_0 = E[n_0] = \langle\Psi_0| H |\Psi_0\rangle . \tag{2.31}$$

From the variational principle, a different density $n(r)$ will give a higher energy

$$E_0 = E[n_0] = \langle\Psi_0| H |\Psi_0\rangle < \langle\Psi| H |\Psi\rangle = E[n] \tag{2.32}$$

Thus, the total energy is minimized for n_0, and so has to be the ground-state energy. □

2.6.2 The Kohn-Sham equation

So far, we have tried to make the challenging Schrödinger equation less challenging by simplifying it, with the last attempt containing the Hohenberg-Kohn's theorems where the theory states that the total ground-state energy can, in principle, be determined exactly once we have found the ground-state density.

In 1965, Kohn and Sham [38] reformulated the Hohenberg-Kohn theorems by generating the exact ground-state density $n_0(r)$ using a Hartree-like total wavefunction

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, .., \mathbf{r}_{N_e}) = \psi_1^{KS}(\mathbf{r}_2)\psi_2^{KS}(\mathbf{r}_2)...\psi_{N_e}^{KS}(\mathbf{r}_{N_e}), \tag{2.33}$$

where $\psi_j^{KS}(r_j)$ are some auxiliary independent single-particle wavefunctions. However, the Kohn-Sham wavefunctions cannot be the correct single-particle wavefunctions since our ansatz implies an exact density

$$n(\mathbf{r}) = \sum_{j=1}^{N_e} |\psi_j^{KS}(\mathbf{r})|^2. \tag{2.34}$$

Recalling that equation 2.30 describes the total energy as a functional of the density,

$$E[n] = T[n] + U_{ee}[n] + U_{en}[n], \tag{2.35}$$

we try to modify it to include the kinetic energy $T_s[n]$ and the interaction energy $U_s[n]$ of the auxiliary wavefunction, and the denotation s for single-particle wavefunctions.

$$\begin{aligned} E[n] &= T[n] + U_{ee}[n] + U_{en}[n] + \big(T_s[n] - T_s[n]\big) + \big(U_s[n] - U_s[n]\big) \\ &= T_s[n] + U_s[n] + U_{en}[n] + \underbrace{\big(T[n] - T_s[n]\big) + \big(U_{ee}[n] - U_s[n]\big)}_{E_{xc}[n]} \end{aligned}$$

Here we have our first encounter with the *exchange-correlation energy*

$$E_{xc}[n] = \Delta T + \Delta U = \big(T[n] - T_s[n]\big) + \big(U_{ee}[n] - U_s[n]\big), \tag{2.36}$$

which contains the complex many-electron interaction. For non-interacting system, $E_{xc}[n]$ is conveniently zero, but in interacting systems it most likely is a complex expression. However, one can consider it as our mission to find good approximations to this term, as the better approximations, the closer we get to the exact expression.

The exact total energy functional can now be expressed as

$$\begin{aligned} E[n] = &\overbrace{\sum_j \int \psi_j^{KS*} \frac{-\hbar^2\nabla^2}{2m} \psi_j^{KS} d\mathbf{r}}^{T_s[n]} + \overbrace{\frac{1}{2}\int\int q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'}^{U_s[n]} \\ &+ \underbrace{\int V_{en}(\mathbf{r})n(\mathbf{r})d\mathbf{r}}_{U_{en}[n]} + \underbrace{\big(T[n] - T_s[n]\big) + \big(U_{ee}[n] - U_s[n]\big)}_{E_{xc[n]}}. \end{aligned} \tag{2.37}$$

given that the exchange-correlation functional is described correctly. By utilizing the variational principle, we can now formulate a set of Kohn-Sham single-electron equations,

$$\left\{ -\frac{\hbar^2}{2m_e}\nabla_s^2 + V_H(\mathbf{r}) + V_{j\alpha}(\mathbf{r}) + V_{xc}(\mathbf{r}) \right\} \psi_s^{KS}(\mathbf{r}) = \epsilon_s^{KS}\psi_s^{KS}(\mathbf{r}) \tag{2.38}$$

where $V_{xc}(\mathbf{r}) = \partial E_{xc}[n]/\partial n(\mathbf{r})$ and $V_H(\mathbf{r}) = \int q^2 \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}'$ is the Hartree potential describing the electron-electron interaction. It is worth to notice that $V_H(\mathbf{r})$ allows an electron to interacts with itself, resulting in a self-interaction contribution, however this will be taken care of in V_{xc}.

Finally, we can define the total energy of the system according to Kohn-Sham theory as

$$E[n] = \sum_j \epsilon_j^{KS} - \frac{1}{2}\int\int q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[n] - \int V_{xc}(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \tag{2.39}$$

If V_{xc} is exact, and $E[n]$ gives the true total energy, we still do not know if the energy eigenvalues ϵ_s^{KS} are the true single-electron eigenvalues. However, there exists one exception, which is that the highest occupied eigenvalue of a finite system has to be exact if the density is exact.

The only task that is left for us now is to find the exact expression for $E_{xc}[n]$ as a functional of the density $n(r)$. With that expression, we would be able to calculate the total energies of any material, and most likely solve a few of the biggest puzzles in the history of humankind. Unfortunately, the exchange-correlation potential is unknown for most systems.

2.6.3 The exchange-correlation energy

There is one scenario for which we can derive the exact expression of the exchange-correlation functional, namely the *homogeneous electron gas* (HEG). However, this has a natural cause, since by definition $n(\mathbf{r})$ is constant for this situation. Given that it is the variations of electron density that are the foundation of material properties, the usefulness of HEG is limited. The *local density approximation* (LDA) is an approximation based on this approach, where the local density is the only variable used to define the exchange-correlation functional. Specifically, we can set the exchange-correlation potential at each position to be the known exchange-correlation potential from homogeneous electron gas at the electron density observed at that position [38]:

$$V_{xc}(\mathbf{r}) = V_{xc}^{\text{electron gas}} \left[n(\mathbf{r})\right]. \tag{2.40}$$

This is the simplest and most known approximation to the exchange-correlation functional, and accordingly it has a few drawbacks. One of them is the incomplete cancellation of the self-interaction term, which leads to a repulsion that may cause artifical repulsion between electrons, and hence increased electron delocalization [39]. In addition, LDA has proven challenging to use when studying atoms and molecules because of their rapidly varying electron densities, however, the LDA is seen as succesful for bulk materials because of the slowly varying electron density [40]. Considering the relatively low computational cost and high accuracy, the LDA overall makes a good model for estimation of the exchange-correlation functional for bulk-materials.

In the light of the merits of the LDA, an extensive search for new approximations was launched. The *generalized gradient approximation* (GGA) is an extension of the LDA, which includes the gradient of the density

$$V_{xc}^{GGA}(\mathbf{r}) = V_{xc}\left[n(\mathbf{r}), \nabla n(\mathbf{r})\right]. \tag{2.41}$$

The GGA is a good approximation for the cases where the electron density varies slowly, but faces difficulties in many materials with rapidly varying gradients in the density, causing the GGA to fail. Thus, the annotation *generalized* in GGA is set to include the different approaches to deal with this challenge. Two of the most commonly implemented GGA functionals are the non-empirical approaches Perdew-Wang 91 (PW91) [41] and Perdew-Burke-Ernzerhof (PBE) [42].

Both LDA and GGA are commonly known to severely underestimate the band gaps of semiconductor materials, in addition to incorrectly predicting charge localizations originating from narrow bands or associated with local lattice distortions around defects [43]. The latter limitation is thought to be due to self-interaction in the Hartree potential in equation 2.38.

Hybrid functionals intermix exact Hartree-Fock exchange with exchange and correlation from functionals based on the LDA or GGA. Hartree-Fock theory completely ignore correlation effects, but account for self-interaction and treats exchange as exact. Since LDA/GGA and Hartree-Fock supplement each other, they can be used as a combination for hybrid-functionals resulting in some cancellation of the self-interaction error. Becke [44] introduced a 50% Hartree-Fock exact exchange and 50% LDA energy functional, while Perdew *et al.* [45] altered it to $25\% - 75\%$ and favoring PBE-GGA instead of LDA.

The inclusion of Hartree Fock exchange improves the description of localized states, but requires significantly more computational power for large systems. Another method called the *GW* approximation includes screening of the exchange interaction [46], but has a computational price that does not neccessarily defend its use. Thus, the real challenge is to reduce the computational effort while still producing satisfactory results. Heyd *et al.* [47] suggested to separate the non-local Hartree-Fock exchange into a short- and long-range portion, incorporating the exact exchange in the short-range contribution. The separation is controlled by an adjustable parameter ω, which was empirically optimised for molecules to $\omega = 0.15$ and solids to $\omega = 0.11$ and are known as the HSE03 and HSE06 (Heyd-Scuseria-Ernzerhof), respectively [48]. The functionals are expressed as

$$E_{xc}^{HSE} = aE_x^{HF,SR}(\omega) + (1-a)E_x^{PBE,SR}(\omega) + E_x^{PBE,LR}(\omega) + E_c^{PBE} \tag{2.42}$$

where $a = 1/4$ is the Hartree-Fock mixing constant and the abbreviations SR and LR stands for short range and long range, respectively.

Hence, hybrid-functionals are *semi-empirical* functionals that rely on experimental data for accurate results. They give accurate results for several properties, such as energetics, bandgaps and lattice parameters, and can fine-tune parameters fitted to experimental data for even higher accuracy.

Furthermore, the computational effort required for the hybrid-functionals are significantly larger than for non-empirical functionals such as LDA or GGA. Krukau *et al.* [48] reported a substantial increase in computational cost when reducing the parameter ω from 0.20 to 0.11 for 25 solids, and going lower than 0.11 demanded too much to actually defend its use.

Write about TBMBJ functional and OptB88vDW functional (used by JARVIS).

2.6.4 Self-consistent field methods

So, the remaining question is, how do we solve the Kohn-Sham equation? First, we would need to define the Hartree potential, which can be found if we know the electron density. The electron density can be found from the single-electron wave-functions, however, these can only be found from solving the Kohn-Sham equation. This *circle of life* has to start somewhere, but where?

The process can be defined as an iterative method, *a computational scheme*, as visualized in figure 2.1.

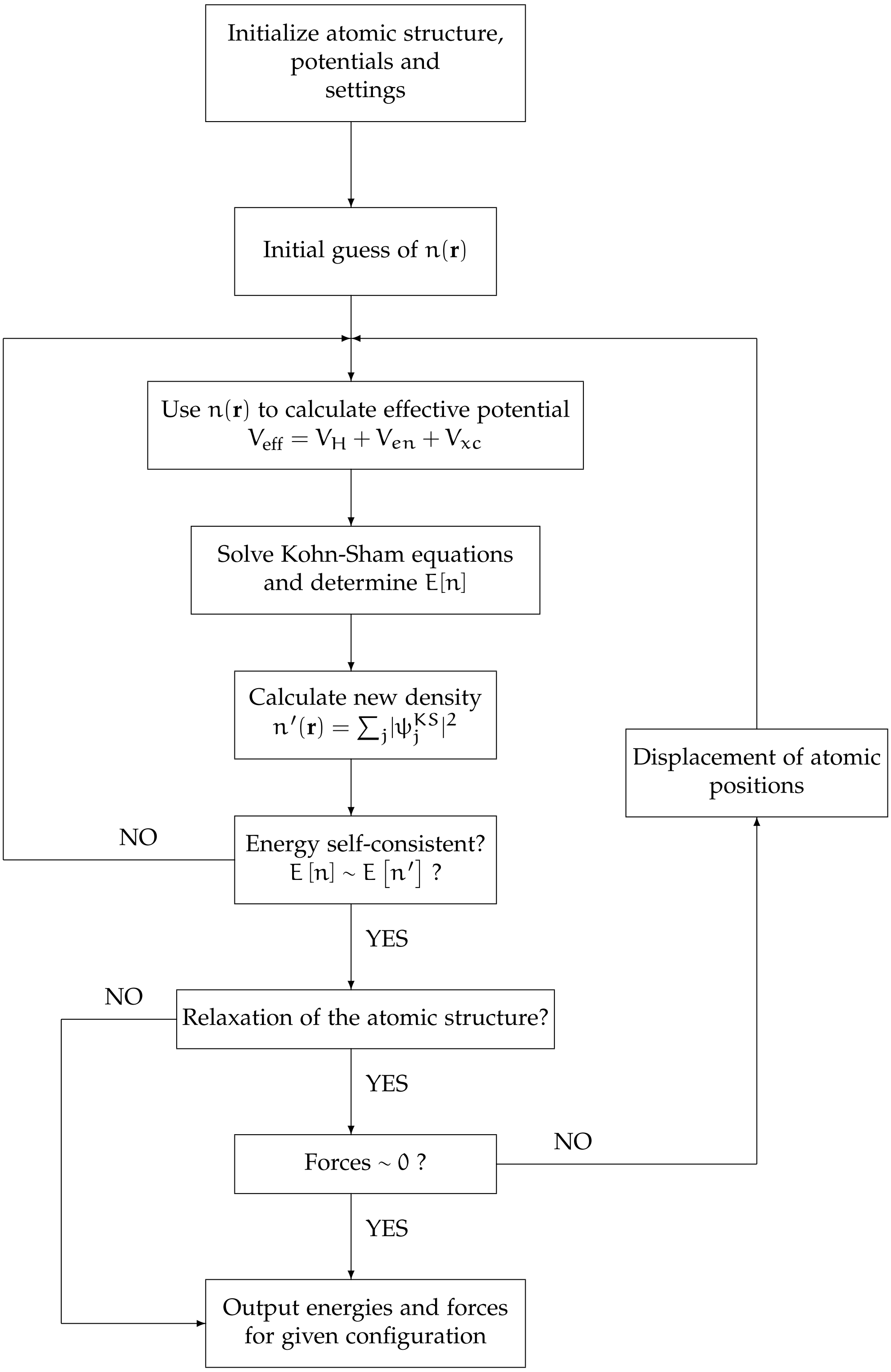


Figure 2.1: A flow chart of the self-consistent field method for DFT.

2.6.5 Limitations of the DFT

If we had known the exact exchange-correlation functional, the density functional theory would yield the exact total energy. Alas, that is not the case and we are bound to use approximations in forms of functionals. The accuracy of calculations is dependent on which functional being used, and normally a higher accuracy means the use of a more complex and computationally demanding computational functional.

Nonetheless, density functional theory is considered a very successful approach and Walter Kohn was awarded the Nobel Price in chemistry in 1998 for his development of the density-functional theory [49]. One can only hope that the future will be as bright as the past, and that this successful theory provides incentives for further growth in the next generation.

Chapter 3

Machine learning

3.1 Supervised learning

3.1.1 randomForrest

3.1.2 gradientBoost

3.2 Unsupervised learning

3.2.1 Kmeans

3.2.2 deep belief networks

Part II

Methodology and implementation

Chapter 4

Material Science Databases

There are multiple different databases for material science available for every day use, some of them completely open-source while others commercial. This chapter will give a brief overview of databases available for computational material science, and will serve as a toolbox for the speciality of each respective data base.

A quick search online will reveal the tremendous escalation of effort for big-data driven material science the last few years, resulting in several databases. We will here distinguish between a *cloud service*, which is a place to store independent databases for research and commercial purposes, and a *database*, which is an organized collection of structured information. As an example, a cloud service can store several databases, but a database cannot host a cloud service.

To limit the quest of databases, we have restricted the search for databases and cloud services to include inorganic compounds obtained by first-principles calculations. Table 4.1 and 4.2 shows the databases and cloud services that meets the given criteries, respectively.

Database	API/REST	Free access	Number of compounds
AFLOW	REST	True	3.27 M
MP [50]	MAPI [51]	True	0.66 M
OQMD [52, 53]	RESTful API (qmpy, matminer)	True	0.64 M
ICSD [54]	RESTful API	False	0.21 M
Jarvis-DFT	API	True	0.04 M

Table 4.1: Databases of computational material science sorted after number of compounds. Abbreviations used are Novel Materials Discovery (NOMAD), Automatic-FLOW for Materials Discovery (AFLOW), Materials Project (MP), Inorganic Crystal Structure Database (ICSD) and Open Quantum Materials Database (OQMD).

Cloud service	API/REST	Free access
NoMaD	API	True
CMR [55]	ASE	RESTful API
MatNavi	API	True
PRISMS	REST	True
Citrine	API	False
MPDS	API	False
MDF	API	False

Table 4.2: Cloud services that offers database-storage. Abbreviations used are Computational Materials Repository (CMR), NIMS Materials Database (MatNavi), PRedictive Integrated Structural Materials Science (PRISMS), Materials Platform for Data Science (MPDS) and the Materials Data Fascility (MDF).

4.1 Fundamentals of a database

4.1.1 modules

Many of the databases share convenient modules that are used to adapt, visualize, calculate or predict properties, making it easier for scientists to utilise the databases.

The Atomic Simulation Environment (ASE) is an environment in the Python programming language that includes several tools and modules for setting up, modifying and analyze atomistic simulations [56]. It is particularly used together with the cloud service Computational Materials Repository (CMR).

Another commonly used module is the Python Materials Genomics (pymatgen) [57]. This is a well-documented open module with both introductory and advanced use cases written in Jupyter Notebook for easy reproducibility, and is integrated with the Materials Project REST API.

The Materials Project is also behind a library named matminer [58], which is an open-source software platform written in Python. Matminer provides modules to extract data sets from many cloud-services and databases, with exampels in table 4.1 and 4.2, it can extract features from images (such as the band gap of a compound), and have modules for visualization of properties.

4.2 Databases and cloud services

4.2.1 Novel Materials Discovery

The Novel Materials Discovery (NOMAD) [59] Repository is an open-access platform for sharing and utilizing computational materials science data. NOMAD also consists of several branches such as NOMAD Archieve, which is the representation of the NOMAD repository parsified into a code-independent format, NOMAD Encyclopedia, which is a graphical user interface (GUI) for characterizing materials, and lastly NOMAD Analytics Toolkit, which includes early-development examples of artificial-intelligence tools [59].

Databases that are a part of NOMAD data collection includes Materials Project, the Open Quantum Materials Database and AFLOW. They are all based on the underlying quantum engine Vienna ab initio simulation package (VASP) [60], which is a software based on DFT.

4.2.2 Materials project

Materials project [50] is an open source project that offers a variety of properties of over one hundred thousand of inorganic crystalline materials. It is known as the initiator of materials genomics and has as its mission to accelerate the discovery of new technological materials, with an emphasis on batteries and electrodes, through advanced scientific computic and innovative design .

It is built upon over 60 features[1], some features being irrelevant for some materials while fundamental for others. The data is divided into three different branches, where the first can be described as basic properties of materials including over 30 features, while the second branch describes experimental thermochemical information. The last branch yields information about a particular calculation, in particular information that's relevant for running a DFT script.

Every compound has an initial relaxation of cell and lattice parameters performed using a 1000k-point mesh to ensure that all properties calculated are representative of the idealized unit cell for each respective crystal structure. The functional GGA is used to calculate band structures, while for transition metals it is applied $+U$ correction to correct for correlation effects in d- and f-orbital systems that are not addressed by GGA calculations [61]. The thermodynamic stability for each phase with respect to decomposition, is also calculated. This is denoted as E Above Hull, with a value of zero is defined as the most stable phase at a given composition, while larger positive values indicate increased instability.

Each material contains multiple computations for different purposes, resulting in different 'tasks'. The reason behind this is that each computation has a purpose, such as to calculate the band structure or energy. Therefore, it is possible to receive several tasks for one material which results in more features per material.

4.2.3 AFLOW

The AFLOW[62–64] repository is an automatic software framework for the calculations of a wide range of inorganic material properties. They utilise the GGA-PBE functional within VASP with projector-augmented wavefunction (PAW) potentials to relax twice and optimize the ICSD-sourced structur. They are using a $3000 - 6000$ k-point mesh, indicating a more computationally expensive calculation compared to the Materials Project. Next, the band structure is calculated with an even higher k-point density, in addition to the $+U$ correction term for most occupied d- and f-orbital systems, resulting in a standard band gap [65]. Furthermore, they apply a standard fit gathered from a study of DFT-computed versus experimentally measured band gap widths to the initial calculated value, obtaining a fitted band gap [66].

[1] All features can be viewed in the documentation of the project: github.com/materialsproject/mapidoc/master/materials

AFLOW-ML [67] is an API that uses machine learning to predict thermo-mechanical and electronic properties based on the chemical composition and atomic structure alone, which they denote as *fragment descriptors*. They start with applying a classification model to predict if a compound is either a metal or an insulator, where the latter is confirmed with an additional regression model to predict the band gap width. To be able to predict properties on an independent data set, they utilise a fivefold cross validation process for each model. They report a 93% prediction success rate of their initial binary classification model, whereas the majority of the wrongful predictions are narrow-gap semiconductors. The authors does not compare their predicted band gap to experimental values, but it is found that 93% of the machine-learning-derived values are within 25% of the DFT +U-calculated band gap width [68].

4.2.4 Open Quantum Materials Database

The Open Quantum Materials Database (OQDM) [52, 53] is a free and available database of DFT-calculations. It has included thermodynamic and structural properties of more than 600.000 materials, including all unique entries in the Inorganic Crystal Structure Database (ICSD) consisting of less than 34 atoms.

The DFT calculations are performed with the VASP software whereas the electron exchange and correlation are described with the GGA-PBE, while using the PAW potentials. They relax a structure using $4000 - 8000$ k-point mesh, indicating an even increasing computational expensive calculation than AFLOW again. Several element-specific settings are included such as using the +U extension for various transition metals, lanthanides and actinides. In addition, any calculation containing 3d or actinide elements are spin-polarized with a ferromagnetic alignment of spins to capture possible magnetism. However, the authors note that this approach does not capture complex magnetic, such as antiferromagnetism, which has been found to result in substantial errors for the formation energy [69].

4.2.5 JARVIS

Joint Automated Repository for Various Integrated Simulations (JARVIS) [70] - DFT is an open database based based on the VASP software to perform a variety of material property calculations. It consists of roughly 40.000 3D and 1.000 2D materials using the vdW-DF-OptB88 van der Waals functional, which was originally designed to improve the approximation of properties of two-dimensional van der Waals materials, but has also shown to be effective for bulk materials [71, 72]. The functional has shown accurate predictions for lattice-parameters and energetics for both vDW and non-vdW bonded materials [73].

Structures included in the data set are originally taken from the materials project, and then re-optimized using the OPT-functional. Finally, the combination of the OPT and modified Becke-Johnson (mBJ) functionals are used to obtain a representative band gap of each structure, since both have shown unprecedented accuracy in the calculation of band gap compared to any other DFT-based calculation methods [74].

The JARVIS-DFT database is part of a bigger platform that includes JARVIS-FF, which is the evaluation of classical forcefield with respect to DFT-data, and JARVIS-ML, which consists of 25 machine learning to predict properties of materials. In addition, JARVIS-DFT also includes a data set of 1D-nanowire and 0D-molecular materials, yet not publically distributed.

4.3