Predicting solid-state qubit material host

by

Oliver Lerstøl Hebnes

Thesis

for the degree of

Master of Science

Faculty of Mathematics and Natural Sciences
University of Oslo

January 25, 2021

Contents

Part I

Theory

Chapter 1

Semiconductor quantum platforms

This chapter will provide a brief overview of the current state-of-the-art 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. Thereafter we will look into how materials are built up, and what the characteristics of a semiconductor is.

1.1 Quantum technologies

Quantum technology (QT) refers to practical applications and devices that utilize the principles of quantum physics as a foundation. Technologies in this spectrum are based on concepts such as *superposition*, *entanglement* and *coherence*, which are all closely related to one another.

A quantum superposition refers to that any two or more quantum eigenstates 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 states. This is according to the wave-particle duality which states that every particle or another quantum entity may be described as either a particle or a wave. When measuring the state of a system residing in a superposition of eigenstates, however, the system falls back to one of the basis states that formed the superposition, destroying the original configuration.

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

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 a particle at the same time. Thus, we often calculate the probability for a particle to be in a state which results in concepts such as an electron sky surrounding an atom core. However, remember that equation (1.1) is an inequality, which means that it is possible to create a state where neither the position nor the momentum is well defined.

1.1.1 Quantum computation

The start of the digital world's computational powers can be credited to Alan Turing. In 1937, Turing [2] 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 have in accordance to Moore's law [3], 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 a doubling of computer performance every 18 months [4]. Alas, miniaturization cannot go on forever as transistors are mass-produced at 5 nm today and are expected to reach a critical limit of 3 nm in the following years [5].

To sustain the digital world's increasing computational demand, other alternatives than the conventional classical computer must be explored. This is where quantum computation comes into the picture. The term quantum computer is a device that exploits quantum properties to solve certain computational problems more efficiently than allowed by Boolean logic [6].

The idea is to pass information in the form of a quantum bit, or *qubit* for short. They are the building blocks of quantum computers, and as opposed to the conventional 0 or 1-bits that classical computers are based on, they can inhabit any superposition of the states 0 or 1. This is illustrated in figure 1.1.

The architecture of a gate-based quantum computer is dependent on a set of quantum logic gates that perform unitary transformations on sets of qubits [7, 8]. 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 the system to the ground state [9].

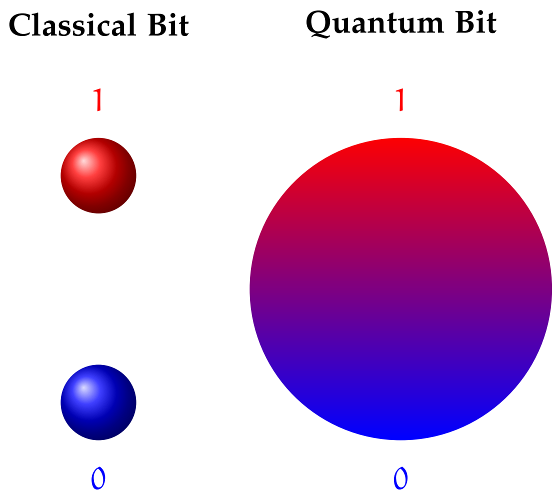


Figure 1.1: Conceptual illustration of the two-level classical bit, which are restricted to the boolean states 1 (true) or 0 (false), and the quantum bit that can be in any superposition of the states 0 or 1.

It has been demonstrated that exponentially complex problems can be reduced to polynomially complex problems for quantum computers [4]. For example, a quantum search algorithm found by Grover [10] offers a quadratic speed-up compared to classical algorithms, while Shor's quantum integer factorization algorithm [11] presents an exponential speed-up. Intriguingly, Google reported in 2019 that they ran a random number generator algorithm on a superconducting processor containing 53 qubits in 200 seconds, which would most likely take several times longer for a classical supercomputer to solve [12]. It is anticipated that quantum computers will excel in exceedingly complex problems, while many simpler tasks may not see any speed-up at all compared to the classical regime. Hence, quantum- and classical computers are envisioned to coexist for each their purpose.

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 [7] and maintain entaglement over both time and distance. Additionally, decoherence and other quantum noise occurs as a result of 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 [4].

1.1.2 Quantum communication

Quantum communication refers to the transfer of a state of one quantum system to another. Since information can be stored in qubits, we picture *flying qubits* that transfer information from one location to another [13]. The benefits of using flying qubits are in particular valued in quantum cryptography, since the quantum nature of qubits can be exploited to add extra layers of security [4].

Consider the example of encrypting a digitally transmitted conversation. It is difficult to avoid someone eavesdropping on a conversation, however, the problem is diminished if the eavesdropper does not speak the language, keeping the information in the conversation safe. This is the original idea of encryption, such that the information has been encrypted into something incomprehensible for any eavesdropper. A common practice is to encrypt information and share a public key, which everyone can read, and a private key, only known for the sender and receiver of information. This should be sufficient to keep the information secure, given that the complexity of the private key is impenetrable.

Importantly, 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, it still presents a challenge, since encrypted information stored today could be deciphered in ten or twenty years' time . Consequently, finding an encryption method that could make information either impossible to eavesdrop on or make the security unbreakable forever is very desirable. This is the ultimate goal of quantum cryptography [4].

Consider the example of information encoded into a qubit as a superposition of two quantum states. 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 [14] which declare that quantum states cannot be copied.

A clever approach to ensure confidentiality is to send the encryption key before sending the actual encrypted information. If the key is received unperturbed, the key remains secret and can be safely employed. If it turns out perturbed, confidentiality is still intact since the key does not contain any information and can be discarded. This approach is termed the *quantum key distribution* (QKD) [14, 15]. It should be noted that this requires both the sender and receiver to have access to methods for sending, receiving and storing qubit states, such as a quantum computer. Additionally, the sender and receiver will need to initially exchange a common secret which is later expanded, making quantum key *expansion* a more exact term for QKD [4, 15].

Most applications and experiments use optical fibers for sending information via photons, with the distance regarded as the main limitation. This is because 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 [16] and are compatible with wavelengths used in telecommunication. Nonetheless, secure QKD up to 400 km has recently been demonstrated using optical fibres in academic prototypes [17].

1.1.3 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 reliable and precise measurements [16]. Thus, improving the accuracy of sensors for every measurement done is desirable. One method to improve measurement accuracy, resolution and sensitivity can be by utilizing quantum sensors. Quantum sensors exploit quantum properties to measure a physical quantity [18]. This is possible because quantum systems are highly susceptible to pertubations to its surroundings, and can be used to detect physical properties such as either temperature or an electrical or magnetic field [18].

For a quantum system to be able to function as a quantum sensors, a few criterias needs to be met. 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 [18].

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 [16, 18], while the best classical sensors scale with $\sqrt{N}$.

1.1.4 Quantum computing requirements

As ever-promising the concepts of quantum technology are, the physical realizations are in the preliminary stage of development. Here we will concretize critical principles for a physical realisation of a quantum platform.

> "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 [19]

DiVincenzo formulated in the year of 2000 seven basic criteria for a physical qubit system with a logic-based architecture [7].

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) must be met for a quantum platform to be considered a quantum computer, while the two last criteria (6-7) were 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 directed to Refs. [8, 16].

Superconducting circuits can be used in quantum computing, since electrons in superconducting materials can form Cooper pairs via an effective electron-electron attraction when the temperature is lower than a critical limit. Below the limit, electrons can move without resistance in the material [20]. 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 [16]. Today, superconducting Josephson junctions are the most widely used quantum platform, but they requires very low temperature (mK) to function, making them costly to use. Additionally, the current devices experience a relatively short coherence time, causing challenges in scaling up.

Single photons is an eligible quantum platform that can be implemented as qubits with one-qubit gates being formed 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 [8].

By fixing the nuclear spin of solid-state systems, it is possible to implement a quantum platform that experience long spin coherence. This enables the manipulation of qubits that utilize electromagnetic fields, making one-qubit gates realizable.

The isolated atom platform 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 achieved through laser-induced spin coupling, however scaling up to large atom numbers induce problems in controlling large systems and cooling of the trapped atoms or ions.

A quantum dot (QD) can be imagined as an artifical atom which is confined in a solid-state host. As an example, a quantum dot can occur when a hole or an electron is trapped in the localized potential of a semiconductor's nanostructure. QDs exhibit similar coherence potential as the isolated atom platform, but without the drawback of confining and cooling of the given atom or ion [16]. Moreover, it is possible to limit decoherence due to nuclear spins by dynamic decoupling of nuclear spin noise and isotope purification [8].

A QD can normally be defined litographically using metallic gates, or as self-assembled QDs where a growth process creates the potential that traps electrons or holes. The difference between them is a question of controllability and temperature, since the metallic gates is primarily controlled electrically and operate at < 1 K, while self-assembly QDs are primarily controlled optically at ~ 4 K [8]. Despite requiring very low temperatures, QDs have the potential for fast voltage control and opticial initialization. As with trapped ions, electrostatically defined quantum dots experience a short-range exchange interaction, imposing a limitation for quantum computing and quantum error correction protocols. A potential solution could include photonic connections between quantum dots. On the contrary, self-assembled quantum dots couple strongly to photons due to their large size in comparison to single atoms. However, the size and shapes of self-assembled quantum dots are decided randomly during the growth process, causing an unfavourable large range of optical absorption and emission energies [8].

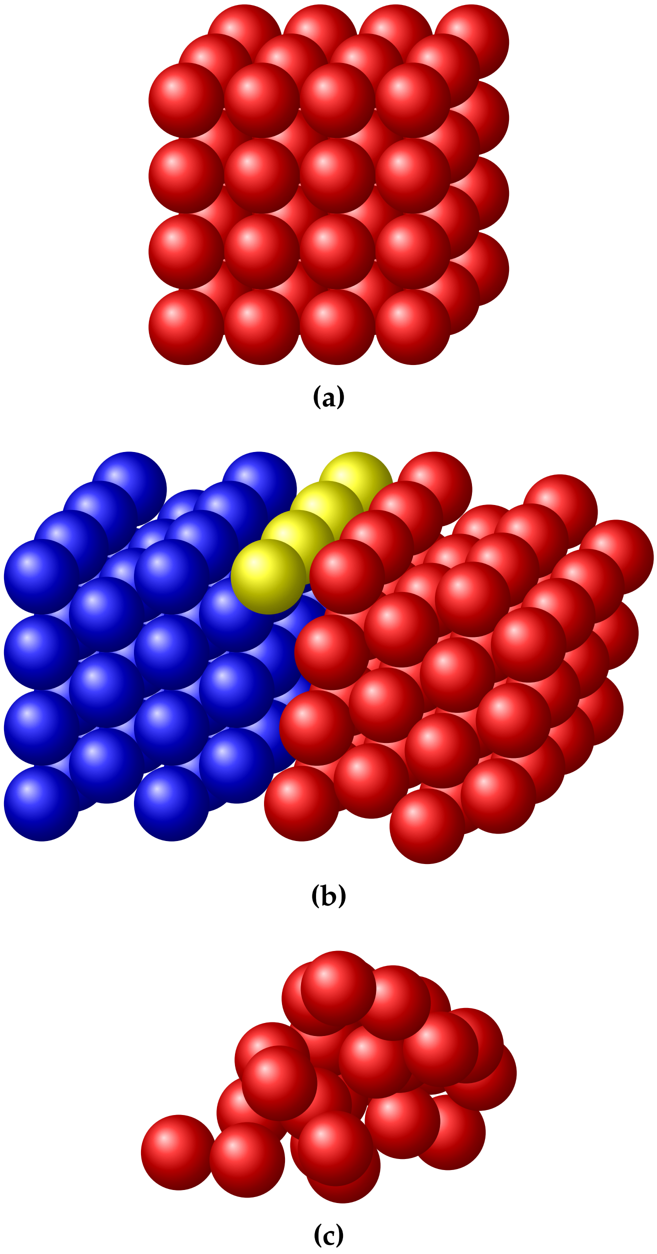
Lastly, we will turn towards point defects in bulk semiconductors as a physical implementation of a quantum platform. Point defects shares many of the attributes of quantum dots, such as discrete optical transitions and controllable coherent spin states, but are vulnerable to small changes in the lattice of the semiconductor. Thus, it can be difficult to isolate a point defect from the surrounding environment. However, one can utilize the strength of the solid-state semiconductor host to isolate to some extent the point defect, yielding extended coherence times and greater optical homogeneity than other quantum dot systems. Before we dwell into the intricacies of point defect qubits as a building block for QT, we will provide the neccessary background for the crystal- and electronic structure of semiconductors.

1.2 A brief overview of materials science

The interactions between atoms and characteristics of matter form the foundation of materials science. The applications of materials science are extensive, with examples such as a bottle of water or to a chair to sit in.

Solid materials, like plastic bottles, 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 [21]. 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 [22]. 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 1.2. Hereon, we will focus on crystalline solids.

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. Every crystal has periodically repeated building blocks called *cells* representing 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.



(a)

(b)

(c)

Figure 1.2: Schematic representation of different degrees of ordered structures, where (a) is a crystalline of a simple cubic lattice, (b) is a polycrystalline hexagonal lattice, and (c) is an amorphous lattice.

One example of a crystal structure is the perovskite structure. Compounds with this structure are characterized by having an ABX_3 stoichiometry whose symmetri belong to one of 15 space groups identified by Lufaso & Woodward [23], such as the cubic, orthorombic and tetragonal. For our purpose, we will be looking into when the X atom is oxygen, and 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 as visualized in figure 1.3.

The motivation behind the research on perovskites is related to the large amount of available ABO_3 chemistries, where a significant portion of these take the perovskite structure. Perovskites have a broad specter of applications, ranging from high-temperature superconductors [24] and ionic conductors [25] to multiferroic materials [26]. Additionally, adding a perovskite-type compound to solar cells has reportedly resulted in higher performance efficiencies while being cheap to produce and simple to manufacture [27, 28]. However, this includes the use of hybrid organic-inorganic compounds and excludes the use of oxygen.

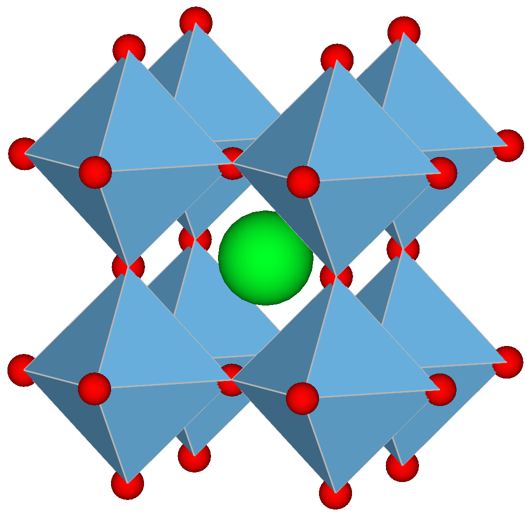


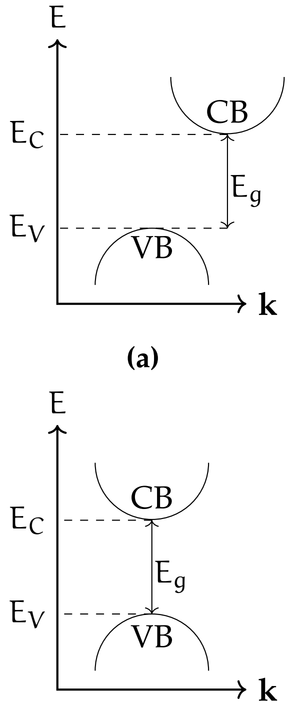
Figure 1.3: A crystal structure of $SrTiO_3$ which is a cubic perovskite. The red atoms are oxygen, whereas the green atom is strontium, and inside every corner-sharing BO_6 octahedral unit is a titanium atom.

1.3 Introduction to semiconductor physics

Isolated atoms have distinct energy levels, where the Pauli exlusion principle [29] states for fermions that each energy level can at most accomodate two electrons of opposite spin. In a solid, the discrete energy levels of the isolated atom spread into continuous energy bands since the wavefunctions of the electrons in the neighboring atoms overlap. Hence, an electron is not neccessarily localized at a particular atom anymore. Every material has a unique band structure, similar to every human having their unique fingerprint.

Knowing which energy bands 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 electron band is called the conduction band (CB). The energy gap of forbidden energy levels between the maximum VB and the minimum CB is known as the band gap, and its energy is denoted as E_g. If a material can be classified as a semiconductor depends on the band gap and the electrical conductivity. As an example, Silicon is commonly thought of as a semiconductor, and has a band gap of about 1.12 eV at 275 K [30].

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 there are no available states nearby, making it impossible for current to flow through the material. This can be solved by using either thermal or optical energy to excite electrons from the valence band to the conduction band, in order to *conduct* electricity. At room temperature, some semiconductors will have electrons excited to the conduction band solely from thermal energy matching the energy band gap [21].



(a)

(b)

Figure 1.4: A schematic drawing of an (a) indirect- and a (b) direct bandgap.

In some scenarios, thermal or optical energy is not sufficient for an excitation since the energy bands are also dependent on the 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 results in an *indirect bandgap* as seen in figure 1.4a. If there is no difference at all, the material has a *direct bandgap*, which is visualized in figure 1.4b.

Electrons in semiconductor materials can be described according to the Fermi-Dirac distribution

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}},$$

where k is Boltzmann's constant, T is temperature, E is the energy and E_F is the Fermi level. The Fermi-Dirac distribution gives the probability that a state will be occupied by an electron, and at $T = 0$ K, every energy state lower than E_F is occupied by electrons while the opposite is true for energy states above E_F [21].

1.3.1 Point defects in semiconductors

In real life, a perfect crystal without any symmetry-breaking flaw does not exist. These flaws are known as defects and can occur up to three dimensions. An example one-dimensional defect is known as a *line defect*, while two dimensional defects can be *planar defects*, and in three dimensions we have *volume defects*. Lastly, defects can also occur in zero dimensions and are then termed *point defects*. Point defects normally occur as either vacancies, interstitial placement inbetween lattice sites or as substitution of another existing atom in the lattice.

Defects can greatly influence both the electronic and optical propertires of a material. A substitional defect can at first be regarded as an impurity or an antisite, but they can also be intentionally inserted, an approach known as *doping*. Doping can result in an excess of electrons or holes, making the semiconductor either an n- or p-type, respectively. Consequently, the semiconductor will have energy levels in the (forbidden) band gap that originates from the defects. If the energy levels introduced are closer than ~ 0.2 eV to the band egdes, they are termed *shallow* defects.

Shallow defects can contribute with either excess electrons to the conduction band, or excess holes to the valence band. However, the induced charge carriers (electrons or holes) interact strongly with the band egdes, resulting in a delocalized wavefunction regarding the position in the lattice.

For the opposite case, if the energy levels rests closer to the middle of the semiconductor's gap, the introduced defects are known as *deep level* defects. Deep levels normally occur due to either dangling bonds or impurities, and have highly localized electron wavefunctions. This might assure the isolation required for long coherence times, which is an appealing promise in quantum technological advances.

Deep levels can be unfortunate in semiconductors since they can interact with the charge carriers, potentially destroying the desired electronic or optical property of the material. Deep level defects can function as electron-hole recombination centers, or to trap charge carriers, yielding the commonly used name deep level *traps*. Both of the given situations results in a lower concentration of charge carriers, which showcase why deep levels can be unwanted in semiconductor devices. However, deep level defects show extraordinary properties in Q

1.3.2 Optical defect transitions

Optical transitions refers to excitation of charge carriers due to either emission or absorption of electromagnetic radiation, and can be done with a laser light or electron beam. Figure 1.5 represents a configuration coordinate (CC) diagram of a defect transition. The y-axis is a function of the energy E, while the x-axis is a function of the configuration coordination Q. The lowest point in the lower parabola is known as the ground state (GS) configuration Q_{GS}, which is the most stable atomic position, while for the upper parabola it is known as the excited state configuration Q_{ES}. The dotted lines represent vibronic excitations to the energy of the ground state Q_{GS} for the lower parabola, while it represents Q_{ES} for the higher parabola.

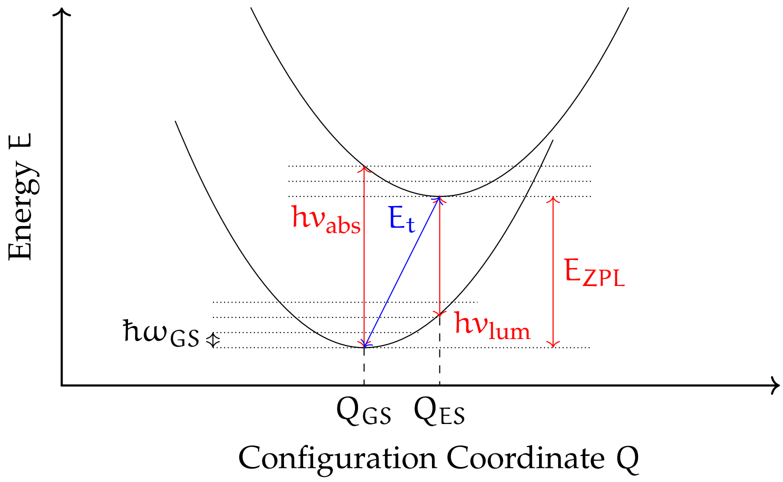


Figure 1.5: A schematic representation of a configuration coordination diagram based on Ref. [31].

The optical transitions in figure 1.5 are marked with red arrows. During slow transitions, such as during thermodynamic defect transitions, the original configuration have time to rearrange due to phonon vibrations. This is schematically drawn as the blue arrow, where the energy E_t equals the ionization energy or the position of the defect level. Optical transitions, on the other hand, are marked in red and occur in a short time range such that the original configuration does not change. They can appear in the exchange of charge carriers with the band egdes, and in a defect's internal excited state, with the latter scenario being most relevant for this thesis.

Consider a defect that rests in the ground state configuration Q_{GS}. Suddenly, it absorbs a photon with energy $h\nu_{abs}$ and occupies an excited vibronic state of the upper parabola after a vertical transition. Through lattice reconfigurations, the defect will move towards the bottom of the upper parabola, also known as Q_{ES}. Eventually, it will relax to the lower parabola by emitting a photon with energy $h\nu_{lum}$, also known as a zero-phonon line (ZPL) of energy E_{ZPL}. On the other hand, any transitions between vibronic excitation levels are phonon-related. How strong the electron-phonon interaction is can be quantified by the Huang-Rhys factor S [32]. If the two parabolas in figure 1.5 have the same configuration of Q, emission into the ZPL is enabled and $S \sim 0$. The stronger the coupling, the smaller amount of emission in the ZPL.

The optical properties of a host material can be greatly influenced by defects, in particular the ES to GS transition that can occur in a defect, as discussed for figure 1.5. If the defect were to fascilitate the emission of single photons with a detectable time inbetween together with a distinguishable ZPL, the defect would be referred to as a single photon source (SPS). The criteria for SPS are not met in many materials, since charge-state transitions often comprise interactions with either the VB or the CB. Thus, most SPSs' GS and ES levels are situated within the band gap of a host material. Consequently, mostly wide-band gap semiconductors are used as host materials for SPSs.

1.4 Semiconductor candidates for quantum technology

The properties of point defects are promising in a quantum technological perspective. We have seen that point defects fasciliate deep energy levels within the band gap of the semiconductor, and provide isolation in the solid-state matrix as a result from a high degree of localization of the defect orbitals. If the host material have a small spin-orbit coupling, it could provide long coherence times for a deep level trap in localized and high-spin states. Additionally, point defects have the potential to be single-photon sources, giving rise to sharp and distinguishable optical transitions, where a significant amount of the emission can be of the energy E_{ZPL}. This is in particular seen in wide bandgap semiconductors, and combined with a weak electron-phonon interaction, can have the capacity to be fabricated as a high-fidelity SPS with a significant ZPL part.

The most studied point defect system is the nitrogen-vacancy (NV^{-1}) in diamond. Figure 1.6 schematically shows the different stages of constructing the negative charge state. Panel 1.6a 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 that generates valence and conduction bands. In panel 1.6b, 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 due to dangling bonds. Substituting a carbon atom with a nitrogen atom further splits the t_2-states into two new states. The states $a(1)$ and e_x/e_y are of importance, as they are the GS and the ES of the qubit defects, respectively. Here, an optical spin-conserving transition can occur due to a laser light of correct wavelength [33], as exemplified from the discussion from the last section.

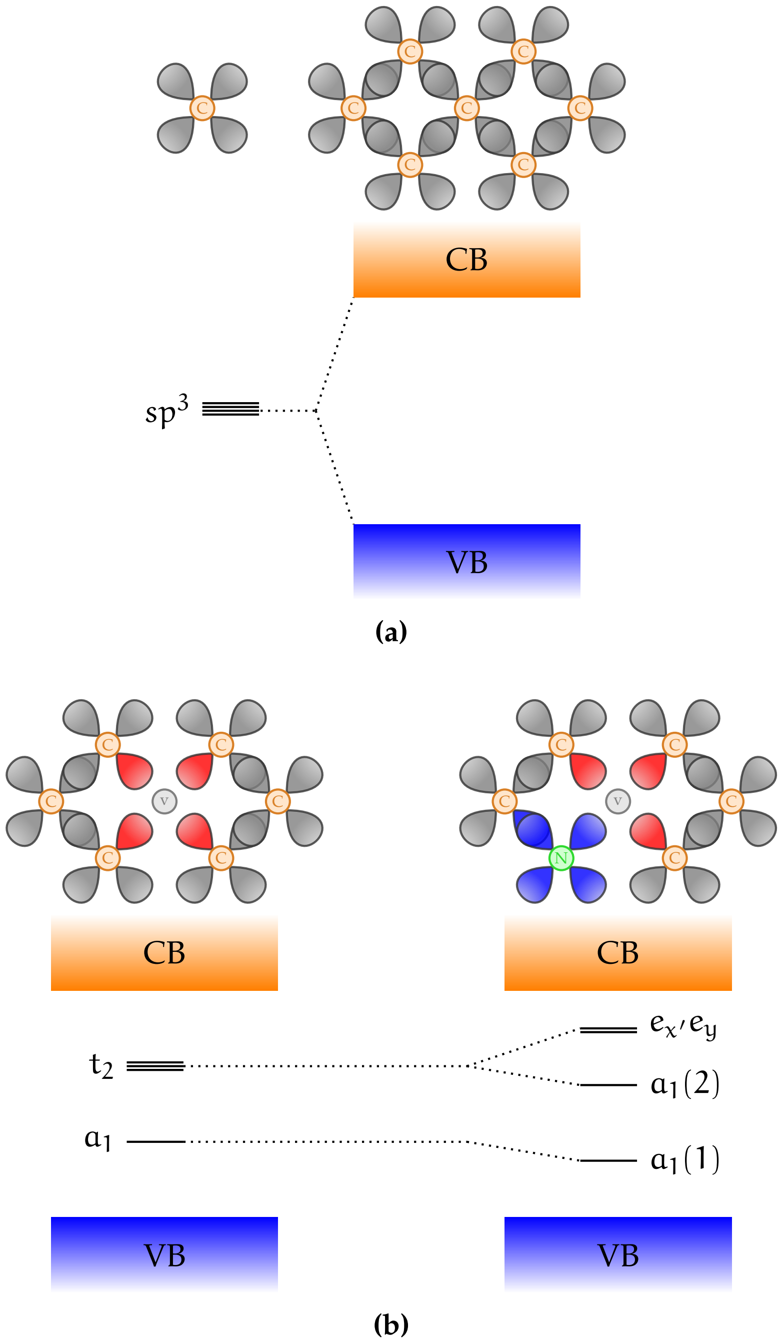


Figure 1.6: A schematic representation of the electronic structure of the NV^{-1} defect in a tetrahedrally coordinated semiconductor, exemplified by diamond. Figure used from Ref. [33].

The nitrogen-vacancy in diamond is a prominent single-photon source up to room temperatures. This involves initializing, manipulating and reading out of the qubit state using optical and electric excitations, and electric and magnetic fields [33]. The potential qubit system have promising applications in quantum- communication and computation, with a demonstrated entanglement between two NV center spins that are separated by 3 m [34]. Nevertheless, perhaps the most propitious application can be seen in quantum sensing as high-sensivity magnetometer with nanoscale resolution [35].

Unfortunately, the NV-center display restricted capababilities for quantum communication and computation. The amount of emission into the zero-phonon line is 4% at 6 K [36], which is low. The emission of the qubit center is not completely compatible with current optical fiber technologies, since the emission is in the red wave-length specter. Additionally, fabricating materials of diamond is far from unchallenging and serves as a signficant incentive to find other promising qubit candidates.

Therefore, we turn to the search of other qubit systems that offers similar capabilities, but that are more user-friendly. In particular, we need to search for new promising materials that can host a potential qubit. Weber *et al.* [6] proposed in 2010 four criteria that should be met for a solid-state semiconductor material hosting a qubit defect, whereas some of the criteria has already been discussed. An ideal crystalline host should have [6]

(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.

Table (1.1) lists several material host candidates that exhibit promising band gap capable of accommodating a deep level 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 may indicate less susceptibility to decoherence.

Criterion (H3) is important for scalability and further potential for a large-scale fabrication. The given candidate hosts provided in table (1.1) can all 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. As a result, the spin-coherence time of a paramagnetic deep center [6] might increase. However, nuclear spin can also induce additional quantum degrees of freedom for applications in the right configuration [37]. Therefore, criterion (H4) is not a strict requirement but is a general recommendation for reducing decoherence time.

Weber *et al* [6] use criteria (H1) – (H4) to specifically find analogies to the NV^{-1} center in other material systems, thus leaving the discussion of other criteria out, such as the choice of crystal system. The atomic configuration and crystal structure of a material strongly influences the properties of a defect, since a defect's orbital and spin structure is dependent on its spatial symmetry [37]. In particular, it is the point group that decides which multiplicity a given energy level should have [38]. A higher defect symmetry group generally facilitates degenerate states, which may give rise to high spin states according to Hund's rules [37, 39]. Inversion symmetry in the host crystal can also be beneficial, resulting in reduced inhomogenous broadening and spectral diffusion of optical transitions as a consequence of being generally insensitive to external electric fields [37].

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 [40]	6.8	Yes
6C-SiC	3.02	7.1	Yes
AlN	6.13	19 [41]	No
GaN	3.44	17.0	No
AlP	2.45	50 [42]	No
GaP	2.27	80	No
AlAs	2.15	275	No
ZnO	3.44 [43]	-3.5	Yes
ZnS	3.72 [44]	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

Table 1.1: Table taken from Gordon *et al.* [33] that lists a number of tetrahedrally coordinated hosts whose band gaps are larger than 2.0 (eV), and compares it to diamond and Si. All experimental values are from Ref. [30], except for where explicity cited otherwise.

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

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

The symmetry of an eigenfunction depends on the symmetry of the potential $V_{ext}(\mathbf{r})$ and the boundary conditions [45].

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 [46] that can calculate 514 TFlops, and we will assume that we need 2000 Flops to calculate each term inside the sum [45], 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.

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

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

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 [45] 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 $n = 1, 2, ...$ 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 [47] 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 [48]. 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 [45]. 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 [49] 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

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

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 [49]:

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

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 [50]. 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 [51]. 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) [52] and Perdew-Burke-Ernzerhof (PBE) [53].

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 [54]. 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 [55] introduced a 50% Hartree-Fock exact exchange and 50% LDA energy functional, while Perdew *et al.* [56] 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 [57], 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.* [58] 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 [59]. 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.* [59] 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.

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 [60]. 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.

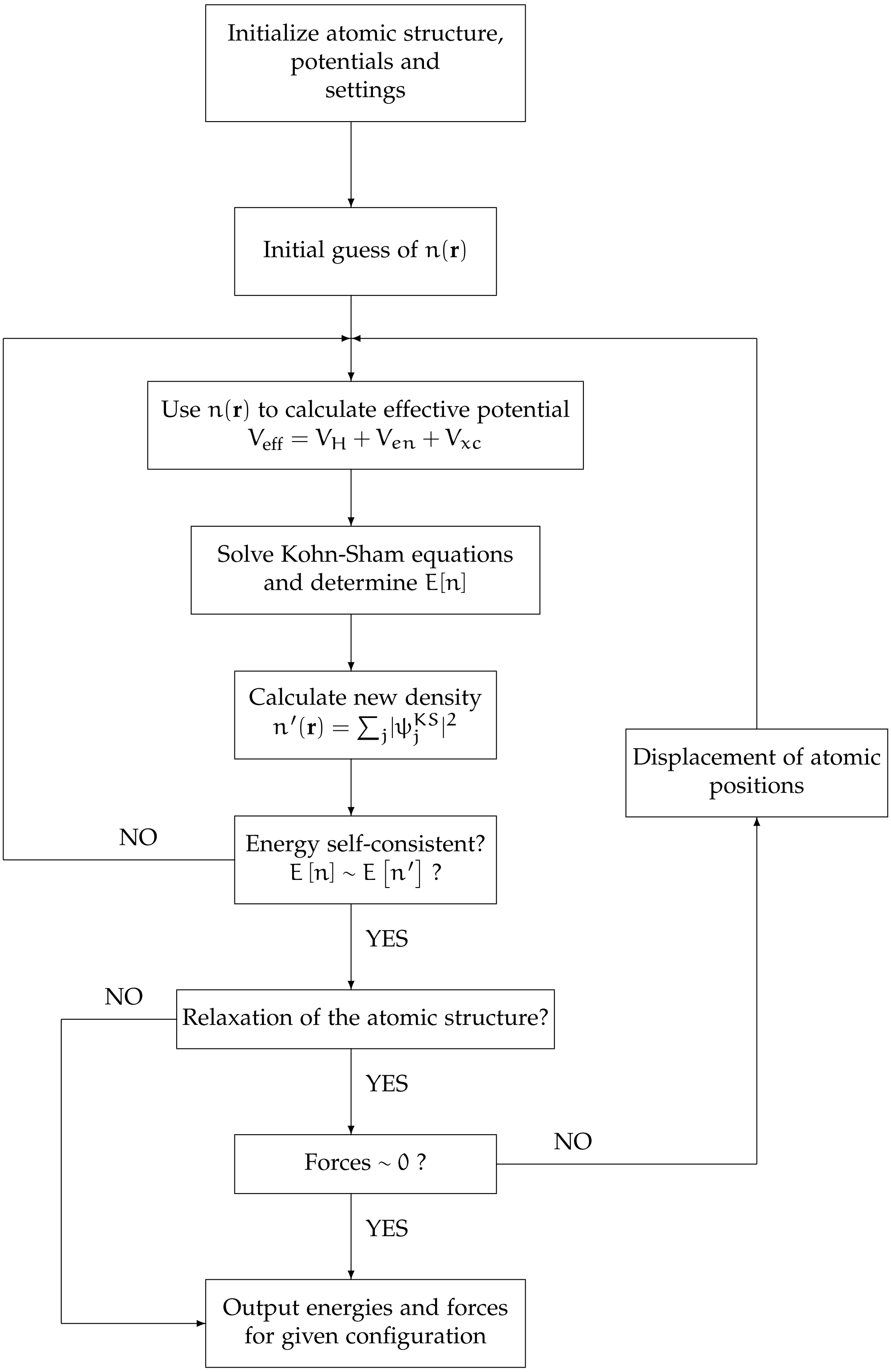


Figure 2.1: A flow chart of the self-consistent field method for DFT.

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 discovery available for every day use, some of them completely open-source while others are commercial. This chapter will give a brief overview of databases available for computational material science, and will serve as a toolbox for how to request information and what kind of python packages exist to process that information.

4.1 Fundamentals of a database

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 that stores ab-initio calculation details and results. 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 experimentally or by first-principles calculations, in particular DFT-calculations using *Vienna ab initio simulation package* (VASP) [61]. VASP is a software for atomic scale materials programming. Table 4.2 and 4.3 shows a selection of databases and cloud services that meets the given criteries, respectively.

4.1.1 API and HTTP requests

To extract information from a database it is convenient to interact through an *API* (Application Programming Interface), which defines important variables such as the kind of requests to be made, how to make them and the data format for transmission. Importantly, this permits communication between different software medias. An API is entirely customizable, and can be made to extend existing functionality or tailormade for specific user-demanding modules.

The APIs that will be encountered is handled by the use of *HTTP* (Hypertext Transfer Protocol), which in its simplest form is a protocol that allows the fetching of resources. The protocol is client-server based, such as the client is requesting information and the server is responding to the request.

The most common HTTP-methods are GET, POST and HEAD, which are used to either retrieve, send, or get information about data, respectively. The latter request is usually done before a GET-method for requests considering large amount of data, since this can be a significant variable for the client's bandwith and load time. Following a request, the server normally responds with one of the status codes in table 4.1.

Status code	Description
2xx	OK - request was successful
3xx	Resource was redirected
4xx	Request failed due to either unsuccessful authentication or client error.
5xx	Request failed due to server error.

Table 4.1: Numeric status code for response. The leftmost digit decide the type of response, while the two follow-up digits depends on the implemented API.

A RESTful (Representational State Transfer) allows users to communicate with a server via a HTTP using a REST Architectural Style [62]. This enables the utilisation of Uniform Resource Identifiers (URI), where each object is represented as a unique resource and can be requested in a uniform manner. Importantly, this allows the use of both URIs and HTTP methods in an API, such that an object is represented by an unique URI whereas a HTTP-method can act on the object. This action will then return either the result of the action, or structured data that represents the object.

4.1.2 Open-source Python libraries for material analysis

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 [69]. It is in particular used together with the cloud service Computational Materials Repository (CMR).

Database	API	Free educational access	Number of entries
AFLOW	REST	True	3.27 M
OQMD [63, 64]	RESTful API (qmpy, matminer)	True	0.82 M
MP [65]	MAPI [66]	True	0.71 M
ICSD [67]	RESTful API	False	0.21 M
Jarvis-DFT	API	True	0.04 M

Table 4.2: 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). The number of entries can give the wrong perception of size of each respective database, as it does not visualise how many calculations have been done for each entry, nor if there might be duplicates.

Cloud service	API/REST	Free educational access
NoMaD	API	True
CMR [68]	ASE	RESTful API
MatNavi	API	True
PRISMS	REST	True
Citrine	API	True
MPDS	API	False
MDF	API	False

Table 4.3: 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).

Another commonly used module is the Python Materials Genomics (pymatgen) [70]. 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 RESTful API.

The Materials Project is also behind a library named matminer [71], 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 4.2 and 4.3, 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

Every database has its own speciality, and no two databases are the same. There exists entries that are fundamentally identical in several databases, but with different properties as a consequence of parameters used, such as the functional utilised in VASP or the relaxation scheme. This section digs up what exactly is each respective database's claim to fame.

4.2.1 Novel Materials Discovery

The Novel Materials Discovery (NOMAD) [72] 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 [72].

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 VASP.

4.2.2 Materials project

Materials project [65] 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 .

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 [73]. 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[74–76] 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 [77]. 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 [78].

AFLOW-ML [79] is an API that uses machine learning to predict thermomechanical 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 [80].

4.2.4 Open Quantum Materials Database

The Open Quantum Materials Database (OQDM) [63, 64] 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 [81].

4.2.5 JARVIS

Joint Automated Repository for Various Integrated Simulations (JARVIS) [82] - 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 [83, 84]. The functional has shown accurate predictions for lattice-parameters and energetics for both vDW and non-vdW bonded materials [85].

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 [86].

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 Practical data extraction with Python-examples

For this section, we will show practical examples of how to extract data that fulfill the criteria for a material to host a qubit candidate given in the theory part. We will begin with the database of Materials Project, and then restrict the query thereafter. This data mining process is reproducable as a jupyter notebook[1] and the databases in question are the ones refered to in the previ-

[1] add and insert DOI for JN data-mining.ipynb

ous section.

Instead of setting up a multiple HTTP-methods, we will here take a look at the easiest method at obtaining data from each database. This includes looking into the APIs that supports data-extraction and that are recommended by each respective database.

The range of data in a database can consist of data from a few entries up to an unlimited amount of entries with even further optional parameters, and has limitless use in applications. However, the amount of data in a database is irrelevant if the data is inaccessible. Therefore, we present an elementary formula which we will use in evaluating if a database is accessible or not. It is defined as

$$\text{Accessibility} = \frac{\text{Extraction speed}}{\text{Amount of data}}. \tag{4.1}$$

A large accessibility term implies an ease in extracting information. This formula does not depend on how accessible a database' user interface is, but a discussion of documentation and user interface will be included in the examples.

4.3.1 Materials Project

The most up-to-date version of Materials Project can be extracted using the python package pymatgen, which is integrated with Materials Project REST API. Other retrievel tools that is dependent on pymatgen includes matminer, with the added functionality of returning a pandas dataframe. Copies of Materials Project are added frequently to cloud services such as Citrine Informatics, but the latest added entries to Materials Project cannot be guaranteed in such a query.

Entries in Materials Project are characterized using more than 60 features[2], 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.

To extract information from the database, we will be utilising the module pymatgen. This query supports MongoDB query and projection operators[3], resulting in an almost instant query. Thus, Materials Project is regarded as a highly accessible database.

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

[3] https://docs.mongodb.com/manual/reference/operator/query/

1. Register for an account[4], and generate a secret API-key.
2. Install pymatgen in the current environment
3. Import pymatgen.
4. Set the required critera.
5. Set the wanted properties.
6. Apply the query.

The code nippet in code listing 4.1 resembles steps $3-6$, with the additional usage of the library Pandas. For this particular query we have set the filter as four items. Firstly, we would like to exclude all spin zero isotopes using the MongoDB operator that matches non of the values specified in the array. Thereafter, we would like to have a compound that is deemed similar to an ICSD entry. All of the resulting entries have be deemed non-magnetic (NM), and lastly, all compounds with polar space groups will be excluded.

```
from pymatgen import MPRester
from pandas import DataFrame

# Insert secret API-key inside MPRester in quotes
with MPRester() as mpr:

    criteria = {"elements":{"$nin":exclude_spin_zero_isotopes},
                "icsd_ids": {'$gte': 0},
                "magnetic_type": {"$eq": "NM"},
                "spacegroup.number": {"$nin":polar_spacegroups}}

    props = ["material_id", "icsd_ids",
             "spacegroup", "band_gap"]

    entries = DataFrame(mpr.query(criteria=criteria,
                                  properties=props))
```

Listing 4.1: Practical example of extracting information from Materials Project using pymatgen, resulting in a Pandas DataFrame named entries that contains the properties given after performing a filter on the database. The criteria is given as a JSON, and supports MongoDB operators.

[4]https://materialsproject.org

4.3.2 Citrine Informatics

Citrine Informatics is a cloud service, which means that the spectrum of stored information varies broadly. We will access research through open access for institutional and educational purposes. Information in Citrine can be stored using a scheme that is broken down into two sections, with private properties for each entry in addition to common fields that are the same for all entries. However, the query happens swiftly and is noted as highly accessible.

In this example, we will gather experimental data using the module matminer. The following steps are required to extract information from Citrine Informatics.

1. Register for an account[5], and generate a secret API-key.
2. Install matminer in the current environment.
3. Import matminer.
4. Set the required critera.
5. Set the wanted properties and common fields.
6. Apply the query.

The code listed in code listing 4.2 gives an easy example to steps $2-4$ with experimental data as filter. The resulting query will be returned as a Pandas DataFrame, but it is not neccessary to include the pandas since it is already implemented in the module matminer.

```
# Insert secret API-key inside CitrineDataRetrieval in quotes
cdr = CitrineDataRetrieval()

criteria = {'data_type': 'EXPERIMENTAL'}
properties = ['Band gap']
common_fields = ['chemicalFormula', "references", "Structure",
    "Crystallinity", "Crystal structure", "uid"]

experimental_entries = cdr.get_dataframe(criteria = criteria,
                             properties = properties,
                             common_fields = common_fields)
```

Listing 4.2: Practical example of extracting information from Citrine Informatics using matminer, resulting in a Pandas DataFrame named experimental_entries that contains the properties given after performing a filter on the database. The criteria is given as a JSON.

[5] https://citrination.com

4.3.3 AFLOW

The query from AFLOW API [74] supports lazy formatting, which means that the query is just a search and does not return values but rather an object. This object is then used in the query when asking for values. For every object it is neccessary to request the desired property, consequently making the query process significantly more time-demanding than similar queries using APIs such as pymatgen or matminer for Citrine Informatics. Hence, the accessibility is strictly limited to either searching for single compounds or if the user possess sufficient time.

Matminer's data retrievel tool for AFLOW is currently an ongoing issue [87], thus we present in code listing 4.3 a function that extracts information from AFLOW and returns a Pandas DataFrame. In contrast to Materials Project and Citrine Informatics, AFLOW does not require an API-key for a query, which reduces the amount of steps to obtain data.

```python
# Tables
import pandas as pd

# Visual representation of the query process
from tqdm import tqdm

# Query library
from aflow import *

import os
if not os.path.exists('data'):
    os.makedirs('data')

def get_dataframe_AFLOW(compound_list, keys, batch_size):
    """
    A function used to make a query to AFLOW.
    ...
    Args
    ----------
    compound_list : list (dim:N)
        A list of strings containing full formula, eg. ["H2O1", "Si1C1"]
    keys : list (dim:M)
        A list containing the features of the compound one wants to extract,
        eg. Egap
    batch_size : int
        Number of data entries to return per HTTP request

    Returns
    -------
    pd.DataFrame (dim:MxN)
        A DataFrame containing the resulting matching queries.
```

```
    This can result
        in several matching compounds
    """
    index = 0
    for compound in tqdm(compound_list):
        print("Current query: {}".format(compound))

        results = search(catalog='icsd', batch_size=batch_size
)\
            .filter(K.compound==compound)

        # If search returns matching query
        if len(results)>0:
            for result in tqdm(results):
                for key in keys:
                    try:
                        aflow_dict[key].append(getattr(result,
key))
                    except:
                        aflow_dict[key].append("None")

                pd.DataFrame.from_dict(aflow_dict)
                  .loc[[index]].to_csv("data/AFLOW_DATA_temp.
csv",
                sep=",",
                index=False,
                header=False,
                mode='a')
                index += 1
        else:
            print("No compound is matching the search")
            continue

    return pd.DataFrame.from_dict(aflow_dict)
```

Listing 4.3: Practical example of extracting information from AFLOW. The function can extract all information in AFLOW for a given list of compounds, however, it is a slow method and requires consistent internet connection.

4.3.4 AFLOW-ML

In this part, we will be using a machine learning algorithm named AFLOW-ML Property Labeled Material Fragments (PLMF) [79] to predict the band gap of structures. This algorithm is compatible with a POSCAR of a compound, which can be generated by the CIF (Crystallographic Information File) that describes a crystal's generic structure. It is possible to download a structure as a poscar by using Materials Project front-end API, but is a cumbersome process to do so individually if the task includes many structures. Extracting the feature of POSCAR is yet to be implemented in the RESful API of pymatgen, thus we demonstrate the versatility of pymatgen with a workaround.

We begin with extracting the desired compounds formula, its material_id for identification, and their respectful structure in CIF-format from Materials Project. In an iterative process, each CIF-structure is parsed to a pymatgen structure, where pymatgen can read and convert the structure to a POSCAR stored as a Python dictionary. Finally, we can use the POSCAR as input to AFLOW-ML, which will return the predicted band gap of the structure. This iterative process parsing and converting, but is an undemanding process. The function that handles this is presented in code listing 4.4.

A significant portion of the process is tied up to obtaining the input-file for AFLOW-ML, and fewer structures will result in an easier process. Nevertheless, we present the following steps in order to receive data from AFLOW-ML.

1. Download AFLOWmlAPI[6].
2. Getting POSCAR from MP.
   (a) Apply the query from Materials Project with "CIF", "material_id" and "full_formula" as properties.
   (b) Insert resulting DataFrame into function defined in code listing 4.4.
3. Insert POSCAR to AFLOW-ML.

```
# Tables
import pandas as pd

# Visual representation of the query process
from tqdm import tqdm

# ML library and structural library
from aflowml.client import AFLOWmlAPI

from pymatgen import Structure
from pymatgen.io.vasp.inputs import Poscar
from pymatgen.io.cif import CifParser

import os
if not os.path.exists('data'):
    os.makedirs('data')

def get_dataframe_AFLOW_ML(entries, pathAndFileName = False):
    """
        A function used to initialise AFLOW-ML with appropiate inputs.
```

[6]http://aflow.org/src/aflow-ml/ to the same directory as code listing 4.4

```
        ...
        Args
        ----------
        entries : Pandas DataFrame
        {
            "cif": {}
                - Materials Project parameter "cif", which is a dict
            "compound": []
                - list of strings
            "material id": []
                - list of strings
        }
        fileName : str
            Path to file, e.g. "data/aflow_ml.csv"
            Writing to a file during iterations. Recommended for large entries.

        Returns
        -------
        Pandas DataFrame
            A DataFrame containing features as compound and material id,
            as well as the keys in the AFLOW-ML algorithm Property
            Labeled Material Fragments.
        """
        def writeToFile(fileName, row):
            pd.DataFrame.from_dict(AFLOW_ML).loc[[index]].to_csv(fileName,
                sep=",",
                index=False,
                header=False,
                mode='a')

        firstIteration = True
        for index, entry in tqdm(entries.iterrows()):

            struc = CifParser.from_string(entry["cif"]).get_structures()[0]

            poscar = Poscar(structure=struc)

            ml = AFLOWmlAPI()
            prediction = ml.get_prediction(poscar, 'plmf')

            if firstIteration:
                AFLOW_ML = {k: [] for k in prediction.keys()}
                AFLOW_ML["full_formula"]   = []
                AFLOW_ML["material_id"] = []
                firstIteration = False
```

```
            for key in prediction.keys():
                AFLOW_ML[key].append(prediction[key])

            AFLOW_ML["full_formula"].append(entry["full_formula"])
            AFLOW_ML["material_id"].append(entry["material_id"])

            if (pathAndFileName):
                writeToFile(pathAndFileName, row=pd.DataFrame.from_dict(AFLOW_ML).loc[[index]])

        return pd.DataFrame.from_dict(AFLOW_ML)
```

Listing 4.4: Practical example of extracting information from AFLOW-ML. The function will convert a CIF-file (from e.g. Materials Project) to a POSCAR, and will use it as input to AFLOW-ML. In return, one will get the structure's predicted band gap. It should be noted that this requires the AFLOW-ML library in the same directory.

4.3.5 JARVIS-DFT

The newest version of the JARVIS-DFT dataset can be obtained by requesting an account at the official webpage, but with the drawback that an administrator has to either accept or deny the request. Thus, the accessibility of the database is dependent on if there is an active administrator paying attention to the requests. Another approach is to download the database through matminer, however with the limitation of not neccessarily having the latest version of the database. The following steps describes the process of extracting JARVIS-DFT using matminer's convenience loader module, and can be regarded as easily accessible with few lines of code and instantanous download.

1. Install matminer in the current environment.
2. Import matminer.
3. Load the dataset using code listing 4.5.

```
from matminer.datasets.convenience_loaders import load_jarvis_dft_3d

JARVIS_entries = load_jarvis_dft_3d(drop_nan_columns=["gap tbmbj"])
```

Listing 4.5: Practical example of extracting information from JARVIS-DFT. For this example, we exclude all metals by removing all non-measured band gaps.

We can observe that there is no advanced search filter when loading the database from matminer. The author of matminer regards this as the user's task, which is easily done through the use of the python library Pandas.

Bibliography

1. Griffiths, D. *Introduction to quantum mechanics* ISBN: 9781107179868 (Cambridge University Press, Cambridge, 2017).
2. Turing, A. M. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London mathematical society* **2,** 230–265 (1937).
3. Moore, G. Cramming More Components Onto Integrated Circuits. *Proceedings of the IEEE* **86,** 82–85 (Jan. 1965).
4. Pavičić, M. *Quantum computation and quantum communication : theory and experiments* ISBN: 9786610743704 (Springer, New York, 2006).
5. Gwennap, L. Apple's 5 Nanometer Chip Is Another Signpost That Moore's Law Is Running Out. *Forbes.* <https://www.forbes.com/sites/linleygwennap/2020/10/12/apple-moores-law-is-running-out/> (Oct. 12, 2020).
6. Weber, J. R. *et al.* Quantum computing with defects. *Proceedings of the National Academy of Sciences* **107,** 8513–8518 (Apr. 2010).
7. DiVincenzo, D. P. The Physical Implementation of Quantum Computation. *Fortschritte der Physik* **48,** 771–783 (Sept. 2000).
8. Ladd, T. D. *et al.* Quantum computers. *Nature* **464,** 45–53 (Mar. 2010).
9. Mizel, A., Lidar, D. A. & Mitchell, M. Simple Proof of Equivalence between Adiabatic Quantum Computation and the Circuit Model. *Physical Review Letters* **99.** doi:10.1103/physrevlett.99.070502 (Aug. 2007).
10. Grover, L. K. A framework for fast quantum mechanical algorithms. arXiv: quant-ph/9711043v2 `[quant-ph]` (Nov. 20, 1997).
11. Shor, P. *Algorithms for quantum computation: discrete logarithms and factoring* in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (IEEE Comput. Soc. Press, 1994). doi:10.1109/sfcs.1994.365700.
12. Martinis, J. M. *et al. Quantum supremacy using a programmable superconducting processor* en. 2019. doi:10.5061/DRYAD.K6T1RJ8.
13. Griffiths, R. B. Nature and location of quantum information. *Physical Review A* **66.** doi:10.1103/physreva.66.012311 (July 2002).

14. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Reviews of Modern Physics* **74,** 145–195 (Mar. 2002).

15. Gisin, N. & Thew, R. Quantum communication. *Nature Photonics* **1,** 165–171 (Mar. 2007).

16. Acín, A. *et al.* The quantum technologies roadmap: a European community view. *New Journal of Physics* **20,** 080201 (Aug. 2018).

17. Boaron, A. *et al.* Secure Quantum Key Distribution over 421 km of Optical Fiber. *Physical Review Letters* **121.** doi:10.1103/physrevlett.121.190502 (Nov. 2018).

18. Degen, C. L., Reinhard, F. & Cappellaro, P. Quantum sensing. *Reviews of Modern Physics* **89.** doi:10.1103/revmodphys.89.035002 (July 2017).

19. Georgescu, I. The DiVincenzo criteria 20 years on. *Nature Reviews Physics* **2,** 666–666 (Nov. 2020).

20. Kristian Fossheim, A. S. *Superconductivity: Physics and Applications* 442 pp. ISBN: 0470844523. <https://www.ebook.de/de/product/3608091/kristian_fossheim_asle_sudboe_superconductivity_physics_and_applications.html> (WILEY, 2004).

21. Ben Streetman, S. B. *Solid State Electronic Devices, Global Edition* 632 pp. ISBN: 1292060557. <https://www.ebook.de/de/product/30394493/ben_streetman_sanjay_banerjee_solid_state_electronic_devices_global_edition.html> (Pearson Education Limited, 2015).

22. Renganathan, G., Tanneru, N. & Madurai, S. L. in *Fundamental Biomaterials: Metals* 211–241 (Elsevier, 2018). doi:10.1016/b978-0-08-102205-4.00010-6.

23. Lufaso, M. W. & Woodward, P. M. Prediction of the crystal structures of perovskites using the software program SPuDS. *Acta Crystallographica Section B Structural Science* **57,** 725–738 (Nov. 2001).

24. Bednorz, J. G. & Müller, K. A. Perovskite-type oxides—The new approach to high-Tcsuperconductivity. *Reviews of Modern Physics* **60,** 585–600 (July 1988).

25. Boivin, J. C. & Mairesse, G. Recent Material Developments in Fast Oxide Ion Conductors. *Chemistry of Materials* **10,** 2870–2888 (Oct. 1998).

26. Cheong, S.-W. & Mostovoy, M. Multiferroics: a magnetic twist for ferroelectricity. *Nature Materials* **6,** 13–20 (Jan. 2007).

27. Ibn-Mohammed, T. *et al.* Perovskite solar cells: An integrated hybrid lifecycle assessment and review in comparison with other photovoltaic technologies. *Renewable and Sustainable Energy Reviews* **80,** 1321–1344 (Dec. 2017).

28. Chen, P.-Y. *et al.* Environmentally responsible fabrication of efficient perovskite solar cells from recycled car batteries. *Energy Environ. Sci.* **7,** 3659–3665 (2014).
29. Pauli, W. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik* **31,** 765–783 (Feb. 1925).
30. Martienssen, W. *Springer handbook of condensed matter and materials data* ISBN: 9786610625949 (Springer, Heidelberg New York, 2005).
31. Pelant, I. *Luminescence spectroscopy of semiconductors* ISBN: 0191738549 (Oxford University Press, Oxford, 2012).
32. Kun Huang, A. R. Theory of light absorption and non-radiative transitions in F -centres. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **204,** 406–423 (Dec. 1950).
33. Gordon, L. *et al.* Quantum computing with defects. *MRS Bulletin* **38,** 802–807 (Oct. 2013).
34. Bernien, H. *et al.* Heralded entanglement between solid-state qubits separated by three metres. *Nature* **497,** 86–90 (Apr. 2013).
35. Taylor, J. M. *et al.* High-sensitivity diamond magnetometer with nanoscale resolution. *Nature Physics* **4,** 810–816 (Sept. 2008).
36. Barclay, P. E., Fu, K.-M. C., Santori, C., Faraon, A. & Beausoleil, R. G. Hybrid Nanocavity Resonant Enhancement of Color Center Emission in Diamond. *Physical Review X* **1.** doi:10.1103/physrevx.1.011007 (Sept. 2011).
37. Bassett, L. C., Alkauskas, A., Exarhos, A. L. & Fu, K.-M. C. Quantum defects by design. *Nanophotonics* **8,** 1867–1888 (Oct. 2019).
38. James, W. J. Theory of defects in solidsby A. M. Stoneham. *Acta Crystallographica Section A* **32,** 527–527 (May 1976).
39. Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466,** 730–734 (Aug. 2010).
40. Neudeck, P. G. Progress in silicon carbide semiconductor electronics technology. *Journal of Electronic Materials* **24,** 283–288 (Apr. 1995).
41. Silveira, E., Freitas, J. A., Glembocki, O. J., Slack, G. A. & Schowalter, L. J. Excitonic structure of bulk AlN from optical reflectivity and cathodoluminescence measurements. *Physical Review B* **71.** doi:10.1103/physrevb.71.041201 (Jan. 2005).
42. Lawaetz, P. Valence-Band Parameters in Cubic Semiconductors. *Physical Review B* **4,** 3460–3467 (Nov. 1971).

43. Beckers, L. *et al.* Structural and optical characterization of epitaxial waveguiding BaTiO3 thin films on MgO. *Journal of Applied Physics* **83,** 3305–3310 (Mar. 1998).

44. Kumbhojkar, N., Nikesh, V. V., Kshirsagar, A. & Mahamuni, S. Photophysical properties of ZnS nanoclusters. *Journal of Applied Physics* **88,** 6260–6264 (Dec. 2000).

45. Persson, C. *Brief Introduction to the density functional theory* 2020.

46. Top500. *SUPERCOMPUTER FUGAKU* June 2020. <https://www.top500.org/system/179807/> (visited on 10/02/2020).

47. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **136,** B864–B871 (Nov. 1964).

48. Toulouse, J. *Introduction to density-functional theory* Sept. 2019. <http://www.lct.jussieu.fr/pagesperso/toulouse/enseignement/introduction_dft.pdf> (visited on 10/25/2020).

49. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140,** A1133–A1138 (Nov. 1965).

50. Allen, J. P. & Watson, G. W. Occupation matrix control of d- and f-electron localisations using DFT+U. *Phys. Chem. Chem. Phys.* **16,** 21016–21031 (2014).

51. David Sholl, J. A. S. *Density Functional Theory: A Practical Introduction* 238 pp. ISBN: 0470373172. <https://www.ebook.de/de/product/7207845/david_sholl_janice_a_steckel_density_functional_theory_a_practical_introduction.html> (WILEY, 2009).

52. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **45,** 13244–13249 (June 1992).

53. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **77,** 3865–3868 (Oct. 1996).

54. Freysoldt, C. *et al.* First-principles calculations for point defects in solids. *Reviews of Modern Physics* **86,** 253–305 (Mar. 2014).

55. Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics* **98,** 1372–1377 (Jan. 1993).

56. Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics* **105,** 9982–9985 (Dec. 1996).

57. Aryasetiawan, F. & Gunnarsson, O. TheGWmethod. *Reports on Progress in Physics* **61,** 237–312 (Mar. 1998).

58. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics* **118,** 8207–8215 (May 2003).

59. Krukau, A. V., Vydrov, O. A., Izmaylov, A. F. & Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *The Journal of Chemical Physics* **125,** 224106 (Dec. 2006).

60. Freitas, L. C. G. Prèmio Nobel de Quìmica em 1998: Walter Kohn e John A. Pople. *Quìmica Nova* **22,** 293–298 (Apr. 1999).

61. Kresse, G. & Furthmüller, J. Efficient iterative schemes forab initiototal-energy calculations using a plane-wave basis set. *Physical Review B* **54,** 11169–11186 (Oct. 1996).

62. Battle, R. & Benson, E. Bridging the semantic Web and Web 2.0 with Representational State Transfer (REST). *Journal of Web Semantics* **6,** 61–69 (Feb. 2008).

63. Kirklin, S. *et al.* The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **1.** doi:10.1038/npjcompumats.2015.10 (Dec. 2015).

64. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **65,** 1501–1509 (Sept. 2013).

65. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1,** 011002 (July 2013).

66. Ong, S. P. *et al.* The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Computational Materials Science* **97,** 209–215 (Feb. 2015).

67. Levin, I. *NIST Inorganic Crystal Structure Database (ICSD)* en. 2020. doi:10.18434/M32147.

68. Landis, D. D. *et al.* The Computational Materials Repository. *Computing in Science & Engineering* **14,** 51–57 (Nov. 2012).

69. Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29,** 273002 (June 2017).

70. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68,** 314–319 (Feb. 2013).

71. Ward, L. *et al.* Matminer: An open source toolkit for materials data mining. *Computational Materials Science* **152,** 60–69 (Sept. 2018).

72. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *Journal of Physics: Materials* **2,** 036001 (May 2019).

73. Wang, L., Maxisch, T. & Ceder, G. Oxidation energies of transition metal oxides within the GGA+U framework. *Physical Review B* **73.** doi:10.1103/physrevb.73.195107 (May 2006).

74. Curtarolo, S. *et al.* AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science* **58,** 227–235 (June 2012).

75. Curtarolo, S. *et al.* AFLOW: An automatic framework for high-throughput materials discovery. *Computational Materials Science* **58,** 218–226 (June 2012).

76. Calderon, C. E. *et al.* The AFLOW standard for high-throughput materials science calculations. *Computational Materials Science* **108,** 233–238 (Oct. 2015).

77. Setyawan, W. & Curtarolo, S. High-throughput electronic band structure calculations: Challenges and tools. *Computational Materials Science* **49,** 299–312 (Aug. 2010).

78. Setyawan, W., Gaume, R. M., Lam, S., Feigelson, R. S. & Curtarolo, S. High-Throughput Combinatorial Database of Electronic Band Structures for Inorganic Scintillator Materials. *ACS Combinatorial Science* **13,** 382–390 (June 2011).

79. Isayev, O. *et al.* Universal fragment descriptors for predicting properties of inorganic crystals. *Nature Communications* **8.** doi:10.1038/ncomms15679 (June 2017).

80. Ferrenti, A. M., de Leon, N. P., Thompson, J. D. & Cava, R. J. Identifying candidate hosts for quantum defects via data mining. *npj Computational Materials* **6.** doi:10.1038/s41524-020-00391-7 (Aug. 2020).

81. Stevanović, V., Lany, S., Zhang, X. & Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Physical Review B* **85.** doi:10.1103/physrevb.85.115104 (Mar. 2012).

82. Choudhary, K. *et al.* JARVIS: An Integrated Infrastructure for Data-driven Materials Design. arXiv: `2007.01831v1 [cond-mat.mtrl-sci]` (July 3, 2020).

83. Thonhauser, T. *et al.* Van der Waals density functional: Self-consistent potential and the nature of the van der Waals bond. *Physical Review B* **76.** doi:10.1103/physrevb.76.125112 (Sept. 2007).

84. Klimeš, J., Bowler, D. R. & Michaelides, A. Van der Waals density functionals applied to solids. *Physical Review B* **83.** doi:10.1103/physrevb.83.195131 (May 2011).

85. Choudhary, K., Cheon, G., Reed, E. & Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Physical Review B* **98.** doi:10.1103/physrevb.98.014107 (July 2018).

86. Choudhary, K. *et al.* Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms. *Scientific Data* **5.** doi:10.1038/sdata.2018.82 (May 2018).

87. Rosenbrock, C. W. A Practical Python API for Querying AFLOWLIB. arXiv: 1710.00813v1 [cs.DB] (Sept. 28, 2017).