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

March 31, 2021

Contents

Contents

Part I

Theory

Chapter 1

Quantum technologies

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 kind of properties a material needs to exhibit to be an eligible host for quantum devices. Finally, we will giving a few specific examples of materials with promising point defects that have been comprehensively researched. Importantly, this will motivate the reasoning for finding new materials that might excel in areas where other materials falls short for utilization in quantum technology.

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*.

3

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   Quantum computing

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 computing 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].

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

Classical Bit   **Quantum Bit**

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.

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].

1.1.1   Quantum computing requirements

As ever-promising the concepts of quantum technology are, the physical real-
izations 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 [13]

DiVincenzo formulated in the year of 2000 seven basic criteria for a phys-
ical 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

These five criteria must be met for a quantum platform to be considered a
quantum computer.

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 [14]. 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 [15] 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) [15, 16]. 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, 16].

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 [17] 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 [18].

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 [17]. 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 [19]. 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 [19].

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 [19].

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 [17, 19], while the best classical sensors scale with $\sqrt{N}$.

1.4   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, 17].

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 [17]. 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 [17]. 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 optical 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.

1.5   Introduction to semiconductor physics

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.
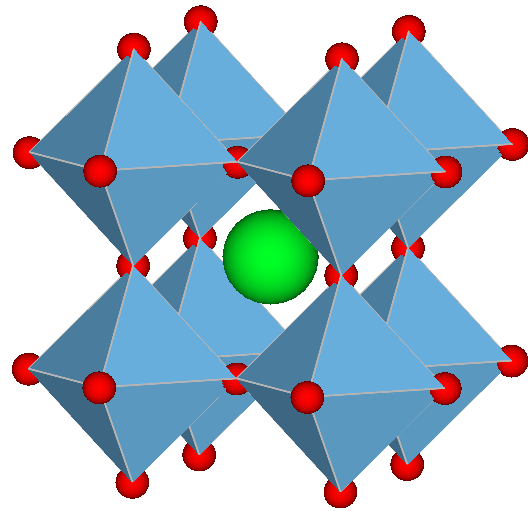
(a)

(b)

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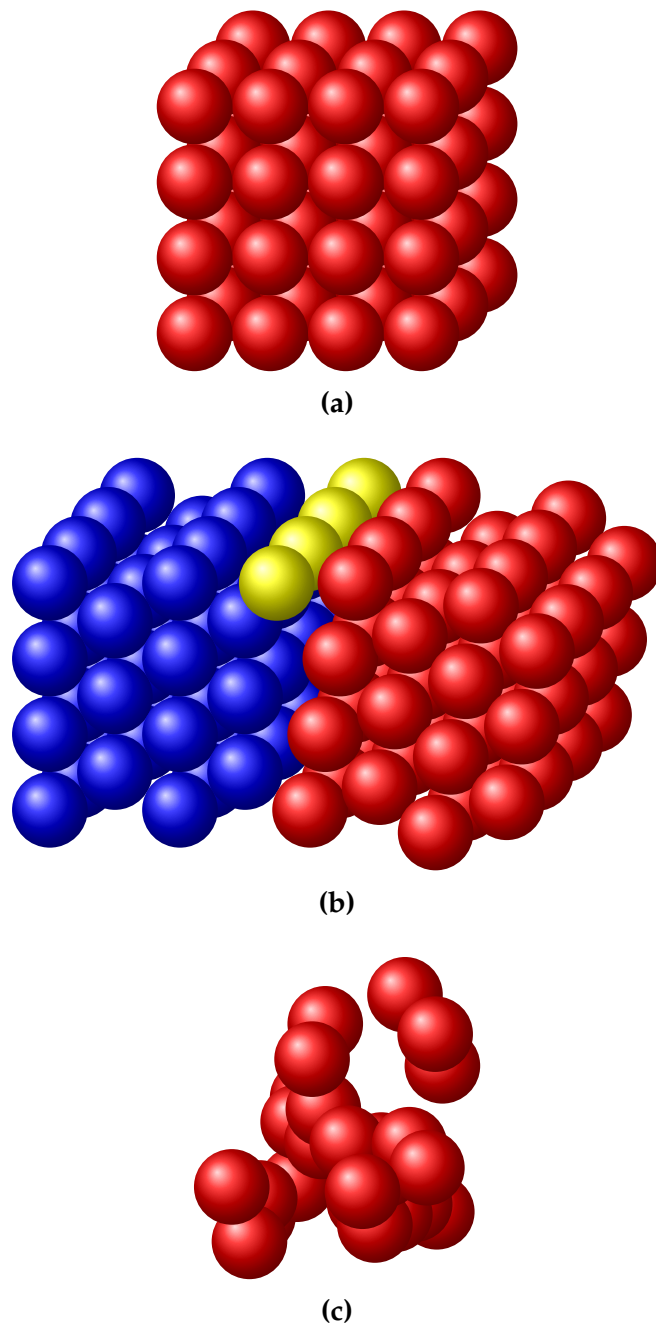
Figure 1.2: Schematic representation of different degrees of ordered structures, where (a) is a crystalline of a simple cubic lattice, (b) is a polycrystalline hexagonal lattice, and (c) is an amorphous lattice.

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.

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. This is examplified as 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].

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

(a)

(b)

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

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

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].

1.5.1 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 dimen-

sions. 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

1.5.2   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.
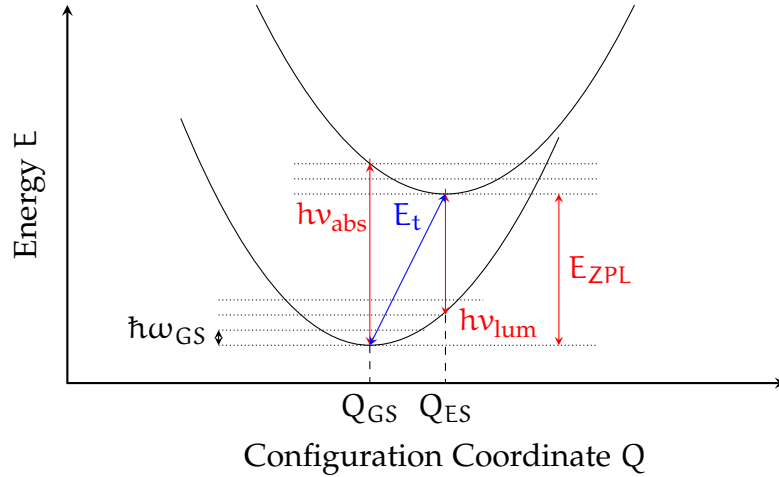
Figure 1.5: A schematic representation of a configuration coordination diagram based on Ref. [31].

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.

# 1.6   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.

In this section we will provide specific examples of a variety of promising candidates, and what properties they possess that makes them auspicious. Additionally, we will briefly mention what the challenges with the candidates are, and why it is important to explore other viable options.

1.6.1   Diamond - the benchmark material for QT

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.

(a)

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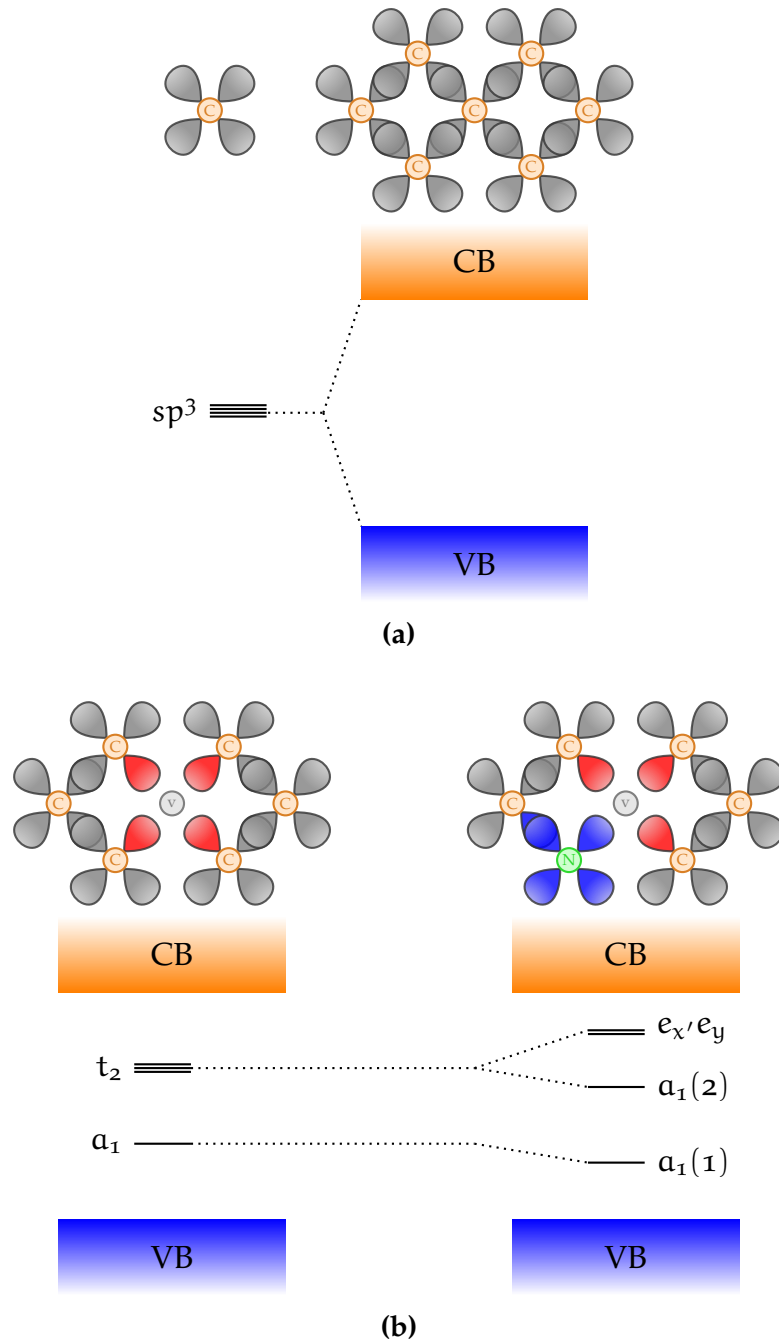
Figure 1.6: A schematic representation of the electronic structure of the NV^{-1} defect in a tetrahedrally coordinated semiconductor, exemplified by diamond. Figure used from Ref. [33].

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.

1.6.2   Qubit material host requirements

Therefore, we turn to the search of other QT compatible hosts 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 point defect. 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.

Material	Band gap E_g (eV)	Spin-orbit splitting Δ_{so} (meV)	Stable spinless nuclear isotopes?
3C-SiC	2.39	10	Yes
4H-SiC	3.26 [37]	6.8	Yes
6H-SiC	3.02	7.1	Yes
AlN	6.13	19 [38]	No
GaN	3.44	17.0	No
AlP	2.45	50 [39]	No
GaP	2.27	80	No
AlAs	2.15	275	No
ZnO	3.44 [40]	-3.5	Yes
ZnS	3.72 [41]	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 tetra-hedrally 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.

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 [42]. 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 [42]. In particular, it is the point group that decides which multiplicity a given energy level should have [43]. A higher defect symmetry group generally facilitates degenerate states, which may give rise to high spin states according to Hund's rules [42, 44]. Inversion symmetry in the host crystal can also be beneficial, resulting in reduced inhomogenous broadening and
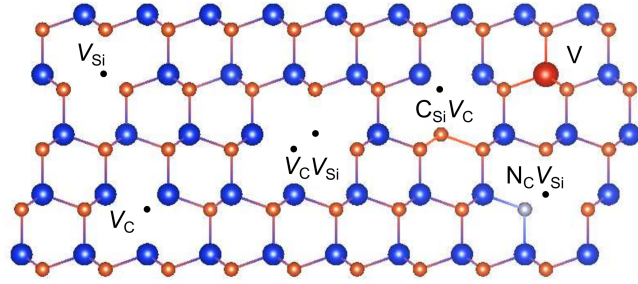
spectral diffusion of optical transitions as a consequence of being generally insensitive to external electric fields [42].

1.6.3　Silicon carbide

Silicone carbide (SiC) is an emerging quantum platform that exists in a wide variety of polytypes, with 3C, 4H and 6H being the most prominent configurations. Several of the polytypes have been demonstrated to host SPEs with a slightly different emitter characteristic, which provides the opportunity to select the desired properties based on the variety of lattice configurations and point defects available [6, 45, 46]. While 3C has a cubic structure, we find 4H in a hexagonal structure with both hexagonal (h) and pseudo-cubic (k) lattice sites. 6H is found in a hexagonal structure with the three orientations that are labelled h, k_1 and k_2. Importantly, SiC in the three varieties experience wide-band gaps, low spin orbit coupling and stable spinless nuclear isotopes [6, 30, 37], as seen from table 1.1. Furthermore, SiC benefits from mature fabrication on the wafer-scale, which checks the last of the four (H1-H4) QT host requirements, marking it as a suitable quantum material platform.

Figure 1.7: Schematic illustration of various point defects in 4H-SiC, where Si atoms are blue while C atoms are orange. The illustration includes the point defects Si vacancy (V_{Si}), C vacancy (V_C), divacancy ($V_{Si}V_C$), carbon antisite-vacancy pair ($C_{Si}V_C$), nitrogen-vacancy ($N_C V_{Si}$) and the vanadium impurity (V).

The most studied emitters in SiC include the carbon antisite-vacancy pair $C_{Si}V_C$ that emits in the red, the silicon vacancy V_{Si} that emits in the near infra-red, and the divacancy ($V_{Si}V_C$) and the nitrogen-vacancy center ($N_C V_{Si}$) that both emit at near-telecom wavelengths. Thus, the two latter emitters could potentially ease the integration with optic fiber technologies as compared to e.g. the NV$^-$. Additionally, the four different point defects have all been identified as room-temperature SPEs with demonstrated coherent spin control [47]. Illustrations of several configurations of emitters in 4H-SiC are included in figure 1.7.

1.6.4   Other promising material hosts

Single photon emitters have been observed in other semiconductor materials, however most of the emitters are yet to be identified or are in an early stage of identification. Therefore, specific details about spin- or emission-related structure are yet to be implemented. In this section we will briefly mention recent promising materials for QT.

One immediate potential candidate is silicon, considering the favorable device fabrication processes that are available. It has demonstrated that phosphorous impurities at Si sites can store a quantum state for over 30 seconds, enabling their use in a potential Kane quantum computer [48]. Unfortunately, the P impurity lack any single photon source capabilities. Recently, however, the G-center arising from the carbon-interstitial carbon-substitutional (C_sC_i) complex was identified as an promising SPE candidate with single photon emissions at telecom wavelength [49].

Other materials that emits individual photons have been detected in other wide-band gap semiconductors, including ZnO, ZnS, GaN and AlN [48]. Unfortunately, challenges due to the specific materials complicate the implementation of defects for QT. ZnO and ZnS experience a broad emission due to a large photon involvement. GaN and AlN, on the other hand, are more susceptible to a more narrow emission, where room-temperature SPE has been demonstrated for both GaN [50] and wurtzite AlN films [51]. The defect levels for AlN films have been tentatively assigned to the nitrogen-vacancy and divacancy complexes, but they tend to occur too close to the band edges for any SPE [48, 52].

Two-dimensional materials such as hexagonal boron nitride (h-BN) are also of interest as quantum platforms [53]. The structure exists in single- or multilayers, and it has been demonstrated a broad range of stable room-temperature single-photon emitters [54, 55]. However, secure identification for the source of the emission is yet to be established [56, 57].

1.6.5   Associated challenges with material host discovery

The idea of finding new potential host candidates to utilise point defects in QT is of a challenging sort. Recall, we have made four criteria that deals with the required (H1) band gaps, (H2) spin-orbit coupling, (H3) availability and (H4) spin-zero isotopes, but we have no knowledge of if there should be more criteria or to what extent a criteron needs to be fulfilled. What we do know is that there are major advantages if materials exhibit properties such as isolation in the lattice and weak electron-photon interaciton, however, the process to provide any quantity of measurements are through approximations and material-specific properties. These approximations does not neccessarily capture quantum properties well.

Furthermore, the identified candidates constitues an immensely selective group of only a handful potential hosts. As an example, most known potential hosts are unary or binary compounds. This is probably due to the increasing complexity dealing with an additional level of interactions in the lattice. Therefore, there are reasons to believe that many potential hosts are yet to be discovered, which serves as a motivation for studies involving exploratory research for new candidates.

Chapter 2

Understanding information from ab-initio calculations

In this work we will be looking into information gained by *ab-initio* calculations, which means "from first principles". Initially, the method of the calculations can be regarded as a black box where one provides the structure of a material as input, which the black box in return feed us the outcome in terms of interesting constants or a different structure. In this chapter we provide the neccessary knowledge to understand what is happening inside the black box, and the quality of the output. The significance of this will be abundantly clear when we later have to interpret information regarding several thousands solid-state systems, and not a single material. In particular, the quality of the output is dependent on being consistent.

The black box is based on a successful theory called density functional theory (DFT), which is an approach for predicting physical properties of solid-state systems. We will ony summarize the neccessary theory behind density functional theory, leaving most of the quantum-mechanical world untouched. We will start with an explanation of why it is difficult to calculate interactions between particles, and then review key approximations and methods regarding the theory. However, even if density functional theory solves some problems, it also introduce new challenges, which will be thoroughly discussed.

2.1   Introduction to density functional theory

To fully understanding what challenges the density functional theory solves, 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 utilize the theory of quantum mechanics.

2.1.1   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 $V_{\text{ext}}(\mathbf{r})$ and the boundary conditions [58].

2.1.2   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

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

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 [59] that can calculate 514 TFlops, and we will assume that we need 2000 Flops to calculate each term inside the sum [58], 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.

2.1.3 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. From here, one can obtain the electronic and nuclear eigenfunction, with the derivation shown in A.1.

2.1.4 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 [58] approximation to this is to assume that electrons can be described independently, suggesting the *ansatz* for a two-electron wavefunction

$$\Psi_\kappa(\mathbf{r}_1, \mathbf{r}_2) = A \cdot \psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2), \tag{2.13}$$

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]:

$$\Psi_\kappa(\mathbf{r_1}, \mathbf{r_2}) = \frac{1}{\sqrt{2}}\Big(\psi_1(\mathbf{r_1})\psi_2(\mathbf{r_2}) - \psi_1(\mathbf{r_2})\psi_2(\mathbf{r_1})\Big). \qquad (2.14)$$

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

2.1.5 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.

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle \leqslant \langle \psi | H | \psi \rangle = E \qquad (2.15)$$

This enables a minimization of energy in terms of wavefunction parameters. A more thorough walk-through of the variational principle is included in appendix A.2.

2.2 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].

Hohenberg and Kohn [60] 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 [61]. 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 [58]. 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

---

[1] not including spin

if the exact functional of the electron density is known. And 57 years after Hohenberg and Kohn published their paper, the exact functional still remains unknown.

We will start this chapter with a brief mention of the Hohenberg-Kohn theorems and its implications, before we delve further into the Kohn-Sham equation. From the

2.2.1 The Hohenberg-Kohn theorems

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

The theorem can be proved by utilising the variational principle for two different external potentials with the same ground state density. The proof is included in appendix A.3.1.

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 $E[n_0] < E[n]$.*

From theorem 1, we know that the external potential is uniquely determined by the density, which in turn uniquely determines the ground state wavefunction. Therefore, all other observables of the system are uniquely determined and we can express the energy as function of the density,

$$E[n] = \overbrace{T[n] + U_{ee}[n]}^{F[n]} + U_{en}[n].$$ (2.16)

where $F[n]$ is an universel functional known as the Hohenberg-Kohn functional. The proof for theorem 2 is found in appendix A.3.2.

2.2.2 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 [62] reformulated the Hohenberg-Kohn theorems by generating the exact ground-state density $n_0(r)$ using a Hartree-like total wavefunction

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, .., \mathbf{r}_{N_e}) = \psi_1^{KS}(\mathbf{r}_2)\psi_2^{KS}(\mathbf{r}_2)...\psi_{N_e}^{KS}(\mathbf{r}_{N_e}),$$ (2.17)

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

$$n(\mathbf{r}) = \sum_{j=1}^{N_e} |\psi_j^{KS}(\mathbf{r})|^2. \qquad (2.18)$$

Recalling that equation 2.16 describes the total energy as a functional of the density,

$$E[n] = T[n] + U_{ee}[n] + U_{en}[n], \qquad (2.19)$$

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*

$$E_{xc}[n] = \Delta T + \Delta U = \big(T[n] - T_s[n]\big) + \big(U_{ee}[n] - U_s[n]\big), \qquad (2.20)$$

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

$$E[n] = \overbrace{\sum_j \int \psi_j^{KS*} \frac{-\hbar^2 \nabla^2}{2m} \psi_j^{KS} d\mathbf{r}}^{T_s[n]} + \overbrace{\frac{1}{2} \int\int q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'}^{U_s[n]}$$
$$+ \underbrace{\int V_{en}(\mathbf{r})n(\mathbf{r})d\mathbf{r}}_{U_{en}[n]} + \underbrace{\big(T[n]-T_s[n]\big) + \big(U_{ee}[n]-U_s[n]\big)}_{E_{xc}[n]}. \qquad (2.21)$$

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,

$$\left\{ -\frac{\hbar^2}{2m_e} \nabla_s^2 + V_H(\mathbf{r}) + V_{j\alpha}(\mathbf{r}) + V_{xc}(\mathbf{r}) \right\} \psi_s^{KS}(\mathbf{r}) = \epsilon_s^{KS} \psi_s^{KS}(\mathbf{r}) \qquad (2.22)$$

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

$$E[n] = \sum_j \epsilon_j^{KS} - \frac{1}{2} \int \int q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[n] - \int V_{xc}(\mathbf{r})n(\mathbf{r}) d\mathbf{r}. \quad (2.23)$$

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.

It is possible to solve the Kohn-Sham equations by applying a self-consistent field method. This is a computational scheme, and for further details one can consult the appendix A.4.

2.2.3   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 [62]:

$$V_{xc}(\mathbf{r}) = V_{xc}^{\text{electron gas}} [n(\mathbf{r})]. \quad (2.24)$$

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 [63]. 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 [64]. Considering the relatively low computational cost and relatively 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

$$V_{xc}^{GGA}(\mathbf{r}) = V_{xc}\left[n(\mathbf{r}), \nabla n(\mathbf{r})\right]. \tag{2.25}$$

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) [65] and Perdew-Burke-Ernzerhof (PBE) [66].

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 [67]. The latter limitation is thought to be due to self-interaction in the Hartree potential in equation 2.22.

To improve the accuracy of excited state properties estimations, other methods have been developed. Tran and Blaha (TB) [68] adapted the exchange potential suggested by Becke-Roussel (BR) [69] that leads to band gaps close to experimental values while still using a cheap semilocal method [70]. The modified TB-mBJ version introduces a parameter relative to the BJ version,

$$V_{xc}^{TB-mBJ}(\mathbf{r}) = c \cdot V_{xc}^{BR}(\mathbf{r}) + (3c - 2)\frac{1}{\pi}\sqrt{\frac{5}{6}}\sqrt{\frac{t(\mathbf{r})}{\rho(\mathbf{r})}}, \tag{2.26}$$

where ρ is the electron density, t is the kinetic-energy density, V_{xc}^{BR} is the Becke-Roussel exchange potential and c is the parameter that changes weights of the BJ potential and is fit to experimental data.

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 [71] introduced a 50% Hartree-Fock exact exchange and 50% LDA energy functional, while Perdew *et al.* [72] 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 [73], 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.* [74] 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 [75]. The functionals are expressed as

$$E_{xc}^{HSE} = a E_x^{HF,SR}(\omega) + (1-a)E_x^{PBE,SR}(\omega) + E_x^{PBE,LR}(\omega) + E_c^{PBE} \qquad (2.27)$$

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.* [75] 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.

Unfortunately, an area where both GGA and hybrid functionals are reportedly inadequate is in calculating the dispersion interactions [76]. Many implementations have been developed to deal with this vulnerability, with one of them being the non-local van der Waals density functional (vdW-DF) [77]

$$E_{xc} = E_x^{GGA} + E_c^{LDA} + E_c^{nl}, \qquad (2.28)$$

where E_x^{GGA} is GGA exchange energy [76].

2.2.4   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. What is common for all approximations is that they are specifically designed to solve one given optimization, therefore it is not neccessarily one functional that is

consider superior in all fields of interest. One could consider that the hybrid functionals due to a high accuracy overall should be dominant, but that is only if one have the computational capacity required. 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 [78]. It is especially regarded as successful in contexts where DFT can make important contributions to scientific questions where it would be essentially impossible to determine through experiments [64]. One can only hope that the future will be as bright as the past, and that this successful theory provides incentives for further growth in the next generation.

Chapter 3

Machine learning

The enourmous amount of data generated in the digital world today is beyond comprehension. In 2019, more than 500 hours of video was uploaded to Youtube every second, totalling to over 82 years of content every day [1]. In addition, more than 1.5 billion web sites exists [2].

However, an increasing amount of data comes hand in hand with an increasing demand of knowledge about the data. If we are unable to extract information from the data, the data serves no intention and exists as an excess. Therefore, we need methods to process and automate data analysis, which is what the promises of *machine learning* covers. Machine learning can reveal patterns in data with ease where a human would face difficulties, and use this information to predict or generate new data. Many tools in machine learning is based on probability theory, which can be applied to problems involving uncertainty. Thus, machine learning is also commonly named as *statistical learning* [79].

There are mainly two types of machine learning, either *supervised* or *unsupervised* learning. In unsupervised learning we are given inputs $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$, where $\mathbf{x}_i$ is a training input that has D-dimensions that describes each entry, where each dimension is known as a *feature*. The features could be examplified as height or weight, or it could be something complex that has no practical meaning (at least not to humans). Since there are no features describing what an entry is, it is up to the tools of machine learning to find patterns in the data, and is the essence of unspervised learning. In the supervised approach, on the other hand, the model tries to learn a mapping from inputs $\mathbf{x}$ to outputs y, given a labeled set of pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$. The set $\mathcal{D}$ is known as the training set, and N is the number of entries. The flexibility of the shape of a feature is also shared with the output. It can in principle

---

[1]Source: https://www.youtube.com/intl/no/about/press extracted 15.02.2021
[2]Source: https://www.statista.com/chart/19058/how-many-websites-are-there extracted 29.03.2021

be anything, but it is mostly assumed that the output is either *categorical* or *nominal* restricted by a finite set $y_i \in \{1, ..., C\}$. The problem is defined as *classification* if the output is categorical, or *regression* if the output is real-valued [79].

3.1   Supervised learning

Supervised learning applied to classification has as goal to learn the target output $y \in \{1, .., C\}$ from the inputs **x**. The number of classes is C, and depicts if the classification is *binary* ($C = 2$), *multiclass* ($C > 2$), or *multi-label* if the class labels are not mutually exclusive (examplified with the weather can be both sunny and cold at the same time). Normally, classification is used when the problem is formulated as a multiclass classification, and hereon we will adapt to the formulation as well [79].

In order to be able to learn from data, we will need to formulate a function approximate. Assume $y = f(x) + \epsilon$ for some unknown function f and a random error term ϵ with mean zero. We can then try to approximate f from a labeled training set, which we can use to make the predictions $\hat{y} = \hat{f}(\mathbf{x})$. With the estimated $\hat{f}$ we can make predictions on unlabeled data and achieve a *generalized model*. The estimated function $\hat{f}$ is often considered as a black box, since we are not neccessarily interested in the exact shape of the function but rather the predictions.

As simple as the idea behind supervised classification appears, a generalized model remains deeply dependent on the available data. Imagine a training set containing two entries. The first entry is a young and tall person labeled healthy. The other entry is an old and short person labeled sick. The pattern in this simple scenario is abundantly clear, but will face a challenge if it were to predict on a test set containing a person who is young and short. Therefore, it is desirable to compute the probability of an entry belonging to one class. The probability distribution is given by $p(y|\mathbf{x}, \mathcal{D})$, where the probability is conditional on the input vector (test set) x and the training set $\mathcal{D}$. If the output is probabilistic, we can compute the estimation to the true label as

$$\hat{y} = \hat{f}(\mathbf{x}) = \arg\max f(x) p(y = 1|\mathbf{x}, \mathcal{D}), \tag{3.1}$$

which represents the most probable class label and is known as the *maximum a posteriori* estimate [79].

3.2   Evaluating accuracy of a model

It would be desirable to find one superior model that we could utilize on all types and sizes of datasets. Unfortunately,

It would be desirable to find one superior model that we could utilize on all types and sizes of datasets. Unfortunately, there is no algorithm that has this property, since one model might be recognized as best on one particular dataset, while others are far better on other datasets. This is known as the *no free lunch theorem* (Wolpert 1996 [80]). The same goes with evaluating the model - there is no metrics that stand alone as the best metric to evaluate a model. Choosing how to actually evaluate a model can be the most challenging part of a statistical learning procedure.

3.2.1 Bias-variance tradeoff

To illustrate a challenge in choosing the correct parameters, we give an example using the mean squared error (MSE) as a *cost function*, which we want to minimize in order to improve the accuracy of the model [79]. Assume that our data can be represented by

$$\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\epsilon},$$

where $f(\mathbf{x})$ is an unknown function and $\boldsymbol{\epsilon}$ is normally distributed with a mean equal to zero and variance equal to σ^2. Furthermore, we also assume that the function $f(\mathbf{x})$ can be approximated to a model $\hat{\mathbf{y}}$, where the model is defined by a design matrix $\mathbf{X}$ and parameters $\boldsymbol{\beta}$,

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}.$$

The parameters $\boldsymbol{\beta}$ are in turn found by optimizing the mean squared error (MSE) via the cost function

$$C(\mathbf{X}, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 = E\left[(\mathbf{y} - \hat{\mathbf{y}})^2\right].$$

The cost function can be rewritten as

$$E\left[(\mathbf{y} - \hat{\mathbf{y}})^2\right] = \frac{1}{n} \sum_i \left(f_i - E\left[\hat{\mathbf{y}}\right]\right)^2 + \frac{1}{n} \sum_i \left(\hat{y}_i - E\left[\hat{\mathbf{y}}\right]\right)^2 + \sigma^2$$

$$= E\left[\left(\mathbf{f} - E\left[\hat{\mathbf{y}}\right]\right)^2\right] + Var\left(\hat{\mathbf{y}}\right) + \sigma_\epsilon^2$$

where $E[\mathbf{y}] = \mathbf{f}$, $E\left[\boldsymbol{\epsilon}\right] = \mathbf{0}$ and $Var\left(\mathbf{y}\right) = Var\left(\boldsymbol{\epsilon}\right) = \sigma_\epsilon^2$.

The first term on the right hand side is the squared bias, the amount by which the average of our estimate differs from the true mean, while the

second term represents the variance of the chosen model. The last term is the variance of the error ϵ, also known as the irreducible error. In general, an estimated function $\hat{f}$ will never be a perfect estimate for f since we can not reduce the error introduced by ϵ. Therefore, any model will always be restricted to an upper bound of accuracy due to the irreducible error.

A model with high variance will typically experience larger fluctuations around the true value, while a model with high bias corresponds to a larger error in the average of estimates. This is schematically visualized as function of model complexity in figure 3.1. If the model is not complex enough due to high bias and low variance, the algorithm can end up not learning the relevant relations between features and output. This is known as *underfitting* [79]. On the other hand, a complex model with low bias and high variance might find trends in random noise from the training data instead of the relevant features, resulting in *overfitting* [79]. An ideal model would be one that simultaneously achieves low variance and low bias Therefore, we have to do a trade-off between how much bias and variance we would like in the model.
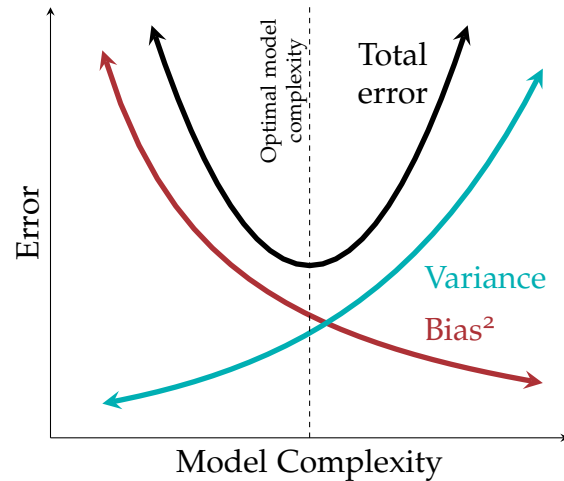
Figure 3.1: A schematic representation of the bias-variance tradeoff as a function of model complexity, adapted from Ref. [81]. The error associated with a model can be decomposed into variance and bias, where a compromise between the lowest bias and variance corresponds to the optimal model complexity.

3.2.2   Accuracy, precision and recall

Given a model that has dealt with the intricacy of increasing complexity, we would like to evaluate the model's output quality. For a binary supervised classification problem we can measure the accuracy by finding how many correct predictions have been made. Prediction accuracy can provide a fine initial analysis, but it has some significant drawbacks seen in unbalanced datasets. This can be easily explained with a dataset consisting of 99 : 1 ratio of class, since just guessing the majority class will result in a very high 99% accuracy. Perhaps it is the 1% that is the most important class, thus the accuracy score severely lacks information for the model.

Therefore, we turn to other evaluation metrics such as a *a confusion matrix*.

A confusion matrix is a method for measuring the performance of classifiers [79]. It is set up as a table with 4 different categories, where two of the categories are the predicted outcomes of the classifier and the two final categories are the true outcomes. An example of a confusion matrix for a binary classifier is shown in table 3.1.

Table 3.1: A confusion matrix for a binary classifier. The entries on the diagonal of the matrix are correctly predictions, while the rest are wrong predictions.

Predicted label

	1	**0**	**total**
1	True Positive	False Negative	P$'$
0	False Positive	True Negative	N$'$
total	P	N	

Actual label

For the binary confusion matrix there are two possible predicted outcomes, either positive or negative. This gives rise to some terminology.

- True Positive (TP): The classifier correctly predicts a positive event.

- True Negative (TN): The classifier correctly predicts a negative event.

- False Positive (FP): The classifier incorrectly predicts a positive event when the true event was negative.

- False Negative (FN): The classifier incorrectly predicts a negative event when the true event was positive.

From the confusion matrix one can then start estimating the performance of the model, by calculating different factors, such as

- **Recall** is a measure of how many truly relevant results are predicted. A high recall relates to a low false negative rate, and is defined as

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{3.2}$$

- **Precision** is a measure of result relevancy. A high precision relates to a low false positive rate, and is defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3.3}$$

Similar to the bias-variance tradeoff, it is common to compare the recall with the precision to identify the tradeoff for different thresholds. High scores for both reveals that a classifier returns accurate results combined with returning a majority of all positive results. Therefore, an ideal classifier will return many correctly predicted results.

Sometimes a classifier can have drastically different values for the precision and recall. This leads to another estimator for the performance of a classifier, which is known as the F1-score. The F1-score is defined as the harmonic mean of precision and recall,

$$\text{F1-score} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}},$$

and can be used to find a good tradeoff between recall and precision. The highest value of F1-score is 1 and is considered an ideal classifier, while the lowest is 0.

3.2.3   Cross-validation

When evaluating different parameters for models, commonly done in a grid-search scheme, there is an abundant risk of performing an overfit to the test set since we can tweak the parameters to a model so it can perform optimally. To solve this problem, we can exclude a part of the dataset as a validation set (in addition to a test set). Therefore, we can train a model on the training set, and evaluate the parameters on the validation set. After a lot of trial and error and the experiment seems successful, we can do one final evaluation on the test set.

Unfortunately, this reduces the number of samples that can be used for training drastically. A fix for this is to apply a cross-validation. Cross-validation is a technique used to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it.

It is common to apply cross validation into fold, yielding the name of k-fold cross-validation. In k-fold cross-validation, the sample is partioned into k equal sized subsamples, as visualized in figure 3.2. Of the k samples, a single sample is used as validation set while the remaining k-1 samples are used as training data. The process is then repeated k-times, such that each of
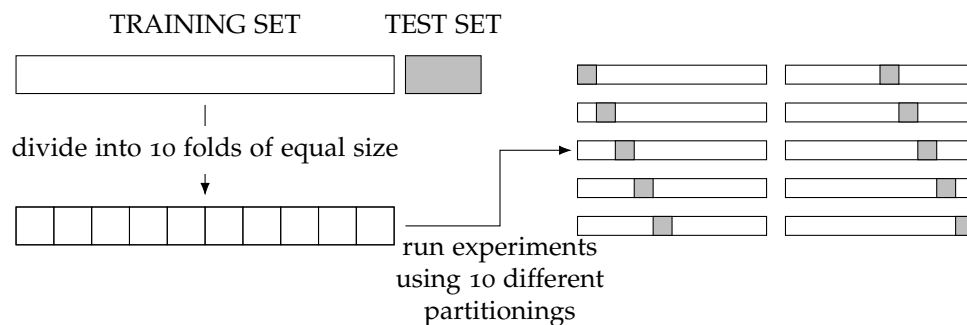
Figure 3.2: A schematic representation of a 10-fold cross-validation scheme.

the k-th subsample is used as validation set exactly one time. Therefore, all observations are used for both training and validation, and each observation is used for validation exactly once. The k results from the folds can then be averaged to produce an estimate. These train and test samples are allowed to have an imbalanced dataset, such as each class is not neccessarily represented equally in each fold. Since supervised algorithms tend to weight each instance equally, it may result in overrepresented classes getting too much weight. Even worse could be the result of an fold where one class is not represented at all, resulting in a model that can't predict a class.

To deal with vulnerability, one can employ a stratified k-fold cross-validation. Stratification is a process that seeks to ensure that each fold is representative of all strata in the data, making each fold having approximately equal class-representation.

3.3   Decision trees

Classification and regression trees (CART), also called decision trees, is one of the more basic supervised algorithms, and can be used for both regression and classification tasks, as the name suggests. The strength of decision trees lays within the simplicity that allows to build more complex networks. We will in this section provide special emphasis to classification trees, but with some remarks to the regression trees to provide a brief perspective of distinctions.

The idea behind decision trees is to find the features that contain the most information regarding the target feature, and then split up the dataset along the values of these features. This feature selection enables the target feature values for the resulting underlying dataset to be as *pure* as possible. The features that can reproduce the best target features are normally said to be the most informative features.

A decision tree can be divided a *root node*, *interior nodes*, and the final *leaf nodes*. Each entity are connected by *branches*. The decision tree is able to learn an underlying structure of the training data and can, given some assumptions, make predictions on unseen query instances. It is the leaf nodes that accomodates the predictions we will make for new entries that is presented to our trained model.

The process behind a decision tree can be seen as a top-down approach. First, we make a leaf provide the classification of a given instance. Then, a node specifies a test of some attribute of the instance, while a branch corresponds to a possible value of an attribute. This allows an entry to be classified by starting at the root node of the tree with corresponding testing of the attribute specified by this node. Subsequently, the instance move down the tree branch corresponding to the value of the attribute. Then the steps can be repeated for a new subtree rooted at the new node.

A classification tree mainly differs to a regression tree by the response of the prediction, since it produces a qualitative response rather than a quantitative one. For a regression tree, the response is given by the mean response of the training observations that belong to the same terminal node. For a classification tree, on the contrary, the response is given by the most commonly occuring class of training observations in the region which it belongs. Thus, the interpretation process includes both the class prediction corresponding to a particular terminal node region, but also in the class proportions among the training observations that fall into that region.

3.3.1   Growing a classification tree

In growing a classification tree, a process called recursive binary splitting is applied which involves mainly two steps.

1. Split the set of possible values $(x_1, x_2, ..., x_p)$ into J distrinct an non-overlapping regions $R_1, R_2, ..., R_J$.

2. If an observation falls within the region R_J, we make the prediction given by the most commonly occuring class of training observations in R_J.

The computational aspect of recursively doing this for every possible combination of features does not defend its use, and therefore the common strategy is to use a top-down approach. Thus, it begins at the top of the tree and consecutively splits the predictor space. This is indicated by two new branches further down the tree. It should be noted that the top-down approach is a greedy approach, since the best split is made at each step of the

tree-growing process, instead of trying to pick a split that will lead to a better tree in a future step.

We can define a *probability density function* (PDF) p_{mk} that represents the number of observations k in a region R_m with N_m observations. This likelihood function can be represented in terms of the proportion $I(y_i = k)$ of observations of this class in region R_m as

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k) \tag{3.4}$$

Hitherto, the splitting of the nodes have been decided by the misclassification error

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i \neq k) = 1 - p_{mk}. \tag{3.5}$$

where the *indicator* function I equals one if we misclassify and equals zero if we classify correctly. However, other methods exists such as the Gini index

$$g = \sum_{k=1}^{K} p_{mk}(1 - p_{mk}) \tag{3.6}$$

and the information entropy

$$s = -\sum_{k=1}^{K} p_{mk} \log p_{mk}. \tag{3.7}$$

The two latter approaches are more sensitive to node purity than the misclassification error, i.e. only containing one class, and are in general preferred [79].

3.3.2   Classification algorithm

The CART algorithm splits the data set in two subsets using a single feature k and a threshold t_k. The pair of quantities (k, t_k) that constitute the purest subset using the gini factor G results in the cost function

$$C(k, t_k) = \frac{m_{\text{left}}}{m} G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}}, \tag{3.8}$$

where $G_{\text{left/right}}$ measures the impurity of left or right subset and $m_{\text{left/right}}$ is the number of instances on either the left or the right subset. The algorithm tries to minimize the cost function to find the pair by splitting the training set in two, and then following the same logic for the next subsets. It will continue to do this recursively until it reaches the maximum depth hyperparameter, or if the next split does not reduce impurity.

3.3.3   Pruning a tree

A decision tree has the ability to turn into a very complex model, making it significantly prone to overfitting. Therefore, techniques that deals with this vulnerability must be implemented.

Pre-puning is a method that stops the growing of a tree if the decrease in error is not sufficient to justify an increasing complex model by adding an extra subtree. However, this implementation has the liability for features that have small predictive power as this might cause a model without any splits at all.

Post-pruning, or just pruning, is the standard method which involves growing the tree to full size, and then prune it. To determine how far to prune it, we can use a cross-validated scheme to evaluate the amount of terminal nodes that has the lowest error.

3.3.4   Pros and cons for decision trees

Decision trees have several clear advantages compared to other algorithms. They are easy to understand, also known as a *white box*, and can be visualised effortlessly for small trees. The algorithm is completely invariant to scaling of the data since each feature is processed separately. Additionally, decision trees can handle both continuous and categorical data and can model interactions between different descriptive features.

As auspicious the advantages of decision trees seems, they are inevitable prone to overfitting and hence does not generalize the data well. Even with pre-puning, post-pruning and setting a maximum depth of terminal nodes, the algorithm is still prone for overfitting [81]. Another important issue concerns training on unbalanced datasets where one class occurs more frequently than other classes, since this will lead to biased trees because the algorithm will favor the more occuring class. Furthermore, small changes in the data may lead to a completely different tree.

Many of these issues can be adressed by using ensemble methods such as either bagging, random forest, or boosting, and can result in a solid improvement of the predictive perfomance of trees.

3.4   Ensemble methods

By using a single decision tree, we often end up with an overfitted model that possess a high variance. Luckily, we can apply methods that aggregate different machine learning algorithms to reduce variance. If each of the algorithms get slightly different results, as they learn different part of the data, we can combine the results into something that is better than any one algorithm

alone. These approaches falls under the categorory of ensemble methods, and will be elaborated further in this section.

3.4.1 Bagging

Bootstrap aggregation, or just *bagging*, is an ensemble method that involves averaging many estimates. If we have M trained trees on different subsamples of the data, chosen randomly, we can compute the ensemble

$$f(\mathbf{x}) = \sum_{b=1}^{M} \frac{1}{B} f_b(\mathbf{x}) \tag{3.9}$$

where f_b is the b'th tree.

Simply re-running the same algorithm on different subsamples can result in a small variance reduction compared to a single tree due to highly correlated predictors, which showcase the need for better approaches.

Random forests provides an improvement of normal bagged trees by choosing a random sample of m predictors as split candidates from the full set of p predictors. The split is restricted in choosing only one of the m predictors. Normally, the value of m is chosen as

$$m \approx \sqrt{p}, \tag{3.10}$$

which means that at each split in a tree, the algorithm is restricted to a very small portion of the available predictors. The specific about the algorithm can be found in Algorithm 1.

Algorithm 1: Random forest algorithm.

---
for *For* $b = 1 : B$ **do**
    Draw a bootstrap sample from the training data;
    Select a tree T_b to grow based on the bootstrap data;
    while *node size smaller than maximum node size* **do**
        Select $m \leqslant p$ variables at random from p predictors;
        Pick the best split point among the m features using CART
          algorithm and create a new node;
        Split the node into daughter nodes;
    end
end
Output the ensemble of trees $\{T_b\}_1^B$ and make predictions

---

By inducing randomness into the model, we arrive at a suprisingly capable model that has a high predictive accuracy [82]. This can be examplified by supposing that there is one strong predictor in a dataset, together with several

other fairly strong predictors. Most of the trees will use this strong predictor at the top split, which means that the bagged trees will look quite similar to each other and will have highly correlated predictions.

However, even with higher prediction accuracy, it comes as a compromise since we loose the easy ability of model interpretation. A single tree can be easy to understand, but interpretation of a huge jungle of trees does not necessarily seem appealing for even an experienced data scientist. Furthermore, a random forest does not substantially reduce the variance as averaging many uncorrelated trees would do, as we will soon find out.

3.4.2 Boosting

Boosting is an ensemble method that fits an additive expansion in a set of elementary basis functions. The basic idea is to combine several weak classifiers, that are only just better than a random guess, in order to create a good classifier. This can be done in an iterative approach were we apply a weak classifier to modify the data. For each iteration, we make sure to weight the observations that are misclassified with a factor. The method is known as adaptive, since the algorithm is able to adapt during the learning process.

In *forward stagewise additive modeling* we want to find an adaptive model

$$f_M(\mathbf{x}) = \sum_{i=1}^{M} \beta_m b(\mathbf{x}; \gamma_m) \tag{3.11}$$

where β_m are expansion parameters that will be determined in a minimization process, and $b(\mathbf{x}; \gamma_m)$ are functions of the multivariable parameter $\mathbf{x}$ that are described by the parameters γ_m. We will in this example consider a binary classification problem with the outcomes $\gamma_i \in \{-1, 1\}$ where $i = 0, 1, 2, ..., n-1$ as the set of observables. The predictions are produced by the classification function $G(\mathbf{x})$.

Then, the error rate of the training sample is given as

$$\overline{err} = \frac{1}{n} \sum_{i=0}^{n-1} I(\hat{y}_i \neq G(\mathbf{x}_i)). \tag{3.12}$$

After defining a weak classifier, we can apply it iteratively to repeatedly modified versions of the data which produce a sequence of different weak classifiers $G_m(\mathbf{x})$. The function $f_M(\mathbf{x})$ will be expressed in terms of

$$G_M(\mathbf{x}) = \text{sign} \sum_{i=1}^{M} \alpha_m G_m(\mathbf{x}). \tag{3.13}$$

The iterative procedure can be defined as

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \beta_m G_m(\mathbf{x}). \tag{3.14}$$

The cost function that leads to the *discrete AdaBoost* algorithm [83] is the exponential cost function

$$C(\mathbf{y}, \mathbf{f}) = \sum_{i=0}^{n-1} \exp\left(-\hat{y}_i(f_{m-1}(\mathbf{x}_i) + \beta G(\mathbf{x}_i))\right), \tag{3.15}$$

or with the weight $w_i^m = \exp\left(-\hat{y}_i f_{m-1}(\mathbf{x}_i)\right)$ we can rewrite to

$$C(\mathbf{y}, \mathbf{f}) = \sum_{i=0}^{n-1} w_i^m \exp\left(-\hat{y}_i \beta G(\mathbf{x}_i)\right). \tag{3.16}$$

We can optimize G for any $\beta > 0$ with

$$G_m(\mathbf{x}) = \text{sign} \sum_{i=0}^{n-1} w_i^m I(\hat{y}_i \neq G(\mathbf{x}_i)). \tag{3.17}$$

This is the classifier that minimize the weighted error rate in predicting y. Furthermore, we can rewrite the cost function to

$$C = \exp(-\beta) \sum_{\hat{y}_i = G(\mathbf{x}_i)} w_i^m + \exp(\beta) \sum_{\hat{y}_i \neq G(\mathbf{x}_i)} w_i^m \tag{3.18}$$

$$= (\exp(\beta) - \exp(-\beta) \sum_{i=0}^{n-1} w_i^m I(\hat{y}_i \neq G(\mathbf{x}_i)) + \exp(-\beta) \sum_{i=0}^{n-1} w_i^m. \tag{3.19}$$

Substituting G_m into C and solving for β, we obtain

$$\beta_m = \frac{1}{2} \log \frac{1 - \overline{\text{err}}}{\text{err}} \tag{3.20}$$

with the error redefined as

$$\overline{\text{err}} = \frac{1}{n} \frac{\sum_{i=0}^{n-1} w_i^m I(\hat{y}_i \neq G_m(\mathbf{x}_i))}{\sum_{i=0}^{n-1} w_i^m}. \tag{3.21}$$

Finally, this leads to an update of

$$f_m(x) = f_{m-1}(\mathbf{x}) + \beta_m G_m(\mathbf{x}) \tag{3.22}$$

and the weights at the next iteration becomes

$$w_i^{m+1} = w_i^m \exp\left(-\hat{y}_i \beta_m G_m(\mathbf{x}_i)\right). \tag{3.23}$$

With the above definitions, we can define the discrete Adaboost algorithm in Algorithm 2.

Algorithm 2: Discrete Adaboost algorithm.

Initialize weights $w_i = 1/n, \quad i = 0, ..., n-1$, such that $\sum_{i=0}^{n-1} w_i = 1$;

for $m = 1 : M$ **do**

　Fit the classifier $f_m(\mathbf{x}) \in \{-1, 1\}$ using weights w_i on the training data;

　Compute the error $\overline{err} = \frac{1}{n} \frac{\sum_{i=0}^{n-1} w_i^m I(\hat{y}_i \neq G_m(\mathbf{x}_i))}{\sum_{i=0}^{n-1} w_i^m}$;

　Define a quantity $\alpha_m = \log \left[(1 - \overline{err}_m)/\overline{err}_m \right]$;

　Set new weights to $w_i \leftarrow w_i \exp\left(\alpha_m I(y_i \neq G(\mathbf{x}_i))\right)$;

end

Compute the new classifier $G(\mathbf{x}) = \sum_{i=0}^{n-1} \alpha_m I(y_i \neq G(\mathbf{x}_i))$;

It is possible to apply different cost functions resulting in a variety of boosting algorithms, which AdaBoost is an example with the cost function being the exponential cost function. But instead of deriving new versions of boosting based on different cost functions, we can find one generic method. The approach is known as *gradient boosting* [84].

Initially, we want to minimize

$$\hat{\mathbf{f}} = \mathrm{argmin}_f L(\mathbf{f}), \tag{3.24}$$

where $\mathbf{f} = \left(f(\mathbf{x}_1, ..., f(\mathbf{x}_N))\right)$ are the parameters of the models, and L is a chosen loss function.

This can be solved stagewise, using an approach named *gradient descent*. At step m, let $\mathbf{g}_m$ be the gradient evaluated at $f(x_i) = f_{m-1}$:

$$\mathbf{g}_m(\mathbf{x}_i) = \left[\frac{\partial L(y_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)} \right]_{f(\mathbf{x}_i) = f_{m-1}(\mathbf{x}_i)}. \tag{3.25}$$

Then we can update

$$\mathbf{f}_m = \mathbf{f}_{m-1} - \rho_m \mathbf{g}_m, \tag{3.26}$$

where rho_m is the step length and can be find by approximating the real function

$$h_m(\mathbf{x}) = -\rho \mathbf{g}_m(\mathbf{x}). \tag{3.27}$$

So far, this only optimize f at a fixed set of points, but we can modify it by fitting a weak classifier to approximate the negative gradient. The resulting

algorithm is shown in Algorithm 3 as the gradient boost algorithm.

Algorithm 3: Gradient boost algorithm.

Initialize the estimate $f_0(\mathbf{x})$;
for $m = 1 : M$ **do**
  Compute the negative gradient vector $\mathbf{u}_m = -\partial C(\mathbf{y}, \mathbf{f})/\partial \mathbf{f}(\mathbf{x})$ at
    $\mathbf{f}(\mathbf{x}) = \mathbf{f}_{m-1}$;
  Fit the base learner to the negative gradient $h_m(\mathbf{u}_m, \mathbf{x})$;
  Update the estimate $f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \nu h_m(\mathbf{u}_m, \mathbf{x})$;
end
Output the final estimation $f_M(\mathbf{x}) = \sum_{m=1}^{M} \nu h_m(\mathbf{u}_m, \mathbf{x})$

3.5 Dimensionality reduction

Supervised learning introduces models that are easy to understand, visualize and has well-defined tools and models. There are several different methods to evaluate a model with many different types of scores such as accuracy, precision and f1-scores. Unfortunately, this does not transfer to unsupervised learning. In unsupervised learning, there is no simple goal for the analysis and the evaluation tends to of a subjective matter. Therefore, unsupervised learning is often used as an *exploratory data analysis* [85]. For data consisting of hundreds or thousands of features, it is possible to apply unsupervised learning to find correlated features and reduce dimensionality of the data, potentially reducing computational effort and time usage drastically. This is the idea of *principal component analysis* (PCA).

3.5.1 Principal component analysis

Principal component analysis is an unsupervised algorithm that tries to find a low-dimension representation of a dataset that contains as much of the variance in the data as possible. Each of the dimensions found by PCA are a linear combination of the features in the dataset, and are known as *principal components*.

We can write the design matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, with p features and n entries, in terms of its column vectors as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \ldots & \mathbf{x}_{p-1} \end{bmatrix}, \tag{3.28}$$

with a given vector

$$\mathbf{x}_i^{\mathsf{T}} = \begin{bmatrix} x_{0,i} & x_{1,i} & x_{2,i} & \ldots & \ldots x_{n-1,i} \end{bmatrix}. \tag{3.29}$$

Then we can compute the *covariance matrix* of the design matrix $\mathbf{X}$, which is a measurement of the joint variability of the p features in $\mathbf{X}$. The covariance is defined as

$$\text{cov}[\boldsymbol{v}, \boldsymbol{u}] = \frac{1}{n} \sum_{i=0}^{n-1} (v_i - \bar{v})(u_i - \bar{u}), \tag{3.30}$$

where $\boldsymbol{v}$ and $\boldsymbol{u}$ are two vectors with n elements each. The covariance matrix is defined by applying the covariance for every pairwise feauture, resulting in a $p \times p$ matrix. On the diagonal, the covariance of two equal features becomes the variance of one,

$$\text{cov}[\boldsymbol{u}, \boldsymbol{u}] = \frac{1}{n} \sum_{i=0}^{n-1} (u_i - \bar{u})(u_i - \bar{u}) \tag{3.31}$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} (u_i - \bar{u})^2 \tag{3.32}$$

$$= \text{var}[\boldsymbol{u}]. \tag{3.33}$$

The covariance accepts values between 0 and infinity, which rises a computational issue due to loss of numerical precision. Therefore, we scale the covariance matrix using the correlation function

$$\text{corr}[\boldsymbol{u}, \boldsymbol{v}] = \frac{\text{cov}[\boldsymbol{u}, \boldsymbol{v}]}{\sqrt{\text{var}[\boldsymbol{u}]\text{var}[\boldsymbol{v}]}}. \tag{3.34}$$

Since all values are inbetween -1 and 1, we avoid any loss of numerical precision. The resulting covariance matrix $\mathbf{C} \in \mathbb{R}^{p \times p}$ becomes

$$\mathbf{C}[\boldsymbol{x}] = \begin{bmatrix} \text{var}[\boldsymbol{x}_0] & \text{cov}[\boldsymbol{x}_0, \boldsymbol{x}_1] & \text{cov}[\boldsymbol{x}_0, \boldsymbol{x}_2] & \dots & \dots & \text{cov}[\boldsymbol{x}_0, \boldsymbol{x}_{p-1}] \\ \text{cov}[\boldsymbol{x}_1, \boldsymbol{x}_0] & \text{var}[\boldsymbol{x}_1] & \text{cov}[\boldsymbol{x}_1, \boldsymbol{x}_2] & \dots & \dots & \text{cov}[\boldsymbol{x}_1, \boldsymbol{x}_{p-1}] \\ \text{cov}[\boldsymbol{x}_2, \boldsymbol{x}_0] & \text{cov}[\boldsymbol{x}_2, \boldsymbol{x}_1] & \text{var}[\boldsymbol{x}_2] & \dots & \dots & \text{cov}[\boldsymbol{x}_2, \boldsymbol{x}_{p-1}] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \text{cov}[\boldsymbol{x}_{p-1}, \boldsymbol{x}_0] & \text{cov}[\boldsymbol{x}_{p-1}, \boldsymbol{x}_1] & \text{cov}[\boldsymbol{x}_{p-1}, \boldsymbol{x}_2] & \dots & \dots & \text{var}[\boldsymbol{x}_{p-1}] \end{bmatrix},$$

$$\tag{3.35}$$

for all vectors x_i where $i = 0, 1, \ldots, p - 1$. The correlation matrix becomes

$$
K[x] = \begin{bmatrix}
1 & \text{corr}[x_0, x_1] & \text{corr}[x_0, x_2] & \ldots & \ldots & \text{corr}[x_0, x_{p-1}] \\
\text{corr}[x_1, x_0] & 1 & \text{corr}[x_1, x_2] & \ldots & \ldots & \text{corr}[x_1, x_{p-1}] \\
\text{corr}[x_2, x_0] & \text{corr}[x_2, x_1] & 1 & \ldots & \ldots & \text{corr}[x_2, x_{p-1}] \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\text{corr}[x_{p-1}, x_0] & \text{corr}[x_{p-1}, x_1] & \text{corr}[x_{p-1}, x_2] & \ldots & \ldots & 1
\end{bmatrix}.
$$

(3.36)

The covariance matrix can be rewritten as a function of the design matrix,

$$
C[x] = \frac{1}{n} XX^\mathsf{T} = \mathbb{E}[XX^\mathsf{T}],
$$

(3.37)

where $\mathbb{E}[XX^\mathsf{T}]$ is the expectation value.

Further on, we assume that we can do apply a number of orthogonal transformations by some orthogonal matrices $S = [s_0, s_1, \ldots, s_{p-1}] \in \mathbb{R}^{p \times p}$ with the column vectors $s_i \in \mathbb{R}^p$. Additionally, we assume that there is a transformation

$$
C[y] = SC[x]S^\mathsf{T} = \mathbb{E}[SXX^\mathsf{T}S^\mathsf{T}],
$$

(3.38)

such that the new matrix $C[y]$ is diagonal with elements $[\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_{p-1}]$. By multiplying with S^T, we arrive at the given eigenvalue i of the covariance matrix that

$$
S_i^\mathsf{T} \lambda_i = C[x] S_i^\mathsf{T}.
$$

(3.39)

Dimensions with large eigenvalue have a large variation and can therefore be used to find features with useful information since we multiply the eigenvalue with the eigenvectors. When the eigenvalues are small, it means that the eigenvectors shrink accordingly and there is a small variation in these specific features.

So far, we have been leading up to the classical PCA theorem. Assume an orthogonal transformation $W \in \mathbb{R}^{p \times p}$. We can then define the reconstruction error

$$
J(W, Z) = \frac{1}{n} \sum_i (x_i - \bar{x}_i)^2,
$$

(3.40)

with $\bar{x}_i = Wz_i$, where z_i is a row vector with dimension $\mathbb{R}^n$ of the matrix $Z \in \mathbb{R}^{p \times n}$.

The PCA theorem states that minimizing the above reconstruction error corresponds to setting $W = S$, which is the orthogonal matrix which diagonalizes the covariance matrix [79]. The optimal number of features that

corresponds to the encoding is given by the a set of vectors z_i with at most l vectors. This is defined as the orthogonal projection of the data onto the columns by the eigenvectors of the covariance matrix. Instead of using the covariance matrix, it is preferable to use the correlation matrix to keep the numerical precision for any raw implementation. Additionally, it is important to mention that PCA is very sensitive to the data, which is why one should always remember to center the data around before applying PCA. We recommend the reader to read Ref. [79] p. 387 for proof of the classical PCA theorem, as we will not elaborate any further. The algorithm is shown in code-listing 4.

Algorithm 4: Principal component analysis algorithm.

Set up the design matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ with p features and n entries;
   Center the data by subtracting the mean value for each column;
Compute the covariance matrix $\mathbb{E}[\overline{\mathbf{X}}\overline{\mathbf{X}}^\top]$;
Find the eigenpairs of $\mathbf{C}$ with eigenvalues $[\lambda_0, \lambda_1, \ldots, \lambda_{p-1}]$ and
   eigenvectors $[\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{p-1}]$;
Order the eigenvalues, and therefore also the eigenvectors, in
   descending order. Keep only those l eigenvalues larger than a selected
   threshold value.

Instead of choosing an arbitrarily number of dimensions to reduce down to, it is common to choose the number of dimensions that accumulate a sufficiently amount of variance, such as e.g. 95%. However, it remains an subjective analysis in how many principal components one should include as it will depend on both the specific application and specific data set. If it is impossible to give a motivation for reducing a large dataset to just two or three principal components, there might still be a reason for why to apply PCA to a dataset. PCA can be applied as a preprocessing method to reduce the dimensionality of a dataset, and therefore might drastically improve the efficiency of further supervised learning approaches.

Part II

Appendices

Appendix A

Extracting information from ab-initio calculations

A.1 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,

$$\Psi_\kappa^{en}(\mathbf{r}, \mathbf{R}) \approx \Psi_\kappa(\mathbf{r}, \mathbf{R})\Theta_\kappa(\mathbf{R}), \tag{A.1}$$

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:

$$\hat{H}^{en} = \overbrace{T_e + U_{ee} + U_{en}}^{\hat{H}^e} + \overbrace{T_n + U_{nn}}^{\hat{H}^n}. \tag{A.2}$$

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

$$\hat{H^{en}}\Psi_\kappa^{en}(\mathbf{r}, \mathbf{R}) = E_\kappa^{en}\Psi_\kappa^{en}(\mathbf{r}, \mathbf{R}) \quad |\times \int \Psi^*(\mathbf{r}, \mathbf{R})\,d\mathbf{r} \quad \text{(A.3)}$$

$$\int \Psi_\kappa^*(\mathbf{r}, \mathbf{R})(\hat{H}^e + \hat{H}^n)\Psi_\kappa(\mathbf{r}, \mathbf{R})\Theta_\kappa(\mathbf{R})\,d\mathbf{r} = E_\kappa^{en}\underbrace{\int \Psi_\kappa^*(\mathbf{r}, \mathbf{R})\Psi_\kappa(\mathbf{r}, \mathbf{R})\,d\mathbf{r}}_{1}\Theta_\kappa(\mathbf{R}). \quad \text{(A.4)}$$

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.

$$E_\kappa(\mathbf{R})\Theta_k(\mathbf{R}) + \int \Psi_k^*(\mathbf{r}, \mathbf{R})H^n\Psi_k(\mathbf{r}, \mathbf{R})\Theta_k(\mathbf{R})\,d\mathbf{r} = E_k^{en}\Theta_k(\mathbf{R}). \quad \text{(A.5)}$$

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

$$\left(T_n + T_n' + T_n'' + U_{nn} + E_\kappa(\mathbf{R})\right)\Theta_\kappa(\mathbf{R}) = E_\kappa^{en}\Theta_\kappa(\mathbf{R}). \quad \text{(A.6)}$$

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

$$\left(T_e + U_{ee} + U_{en}\right)\Psi_\kappa(\mathbf{r}, \mathbf{R}) = E_\kappa(\mathbf{R})\Psi_\kappa(\mathbf{r}, \mathbf{R}) \quad \text{(A.7)}$$

and the nuclear eigenfunction as

$$\left(T_n + U_{nn} + E_\kappa(\mathbf{R})\right)\Theta_\kappa(\mathbf{R}) = E_\kappa^{en}(\mathbf{R})\Theta_\kappa(\mathbf{r}, \mathbf{R}). \quad \text{(A.8)}$$

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

A.2 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.

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle \leqslant \langle \psi | H | \psi \rangle = E \quad \text{(A.9)}$$

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

$$\Psi = \sum_n c_n \psi_n, \quad \text{where} \quad H\psi_n = E_n\psi_n \tag{A.10}$$

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

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

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

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

We have already stated that Ψ is normalized, thus $\sum_n |c_n|^2 = 1$, and the expectation value conveniently is bound to follow equation A.9. 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,

$$E_0 = \min_{\Psi \to \Psi_0} \langle \Psi| H |\Psi\rangle . \tag{A.11}$$

A.3   The Hohenberg-Kohn theorems

A.3.1   The Hohenberg-Kohn theorem 1

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

$$E^{(1)} = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle < \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle \tag{A.12}$$

and

$$E^{(2)} = \left\langle \Psi^{(2)} \middle| \hat{H}^{(2)} \middle| \Psi^{(2)} \right\rangle < \left\langle \Psi^{(1)} \middle| \hat{H}^{(2)} \middle| \Psi^{(1)} \right\rangle. \tag{A.13}$$

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 A.12 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,

$$E^{(1)} = E^{(2)} + \int \left(V_{ext}^{(1)} - V_{ext}^{(2)} \right) n(\mathbf{r}) \, d\mathbf{r} \tag{A.14}$$

A similar procedure can be performed for $E^{(2)}$ in equation A.13, resulting in

$$E^{(2)} = E^{(1)} + \int \left(V_{ext}^{(2)} - V_{ext}^{(1)} \right) n(\mathbf{r}) \, d\mathbf{r}. \tag{A.15}$$

If we add these two equations together, we get

$$
\begin{aligned}
E^{(1)} + E^{(2)} &< E^{(2)} + E^{(1)} + \int \left(V_{ext}^{(1)} - V_{ext}^{(2)} n(\mathbf{r}) \, d\mathbf{r} \right) \\
&\quad + \int \left(V_{ext}^{(2)} - V_{ext}^{(1)} n(\mathbf{r}) \, d\mathbf{r} \right) \\
E^{(1)} + E^{(2)} &< E^{(2)} + E^{(1)},
\end{aligned}
\tag{A.16}
$$

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 $n(\mathbf{r})$. □

A.3.2   The Hohenberg-Kohn theorem 2

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.

$$E[n] = \overbrace{T[n] + U_{ee}[n]}^{F[n]} + \overbrace{U_{en}[n]}^{\int V_{en}n(r)dr} \tag{A.17}$$

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 $n_0(r)$,

$$E_0 = E[n_0] = \langle \Psi_0 | H | \Psi_0 \rangle . \tag{A.18}$$

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

$$E_0 = E[n_0] = \langle \Psi_0 | H | \Psi_0 \rangle < \langle \Psi | H | \Psi \rangle = E[n] \tag{A.19}$$

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

A.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 A.1.
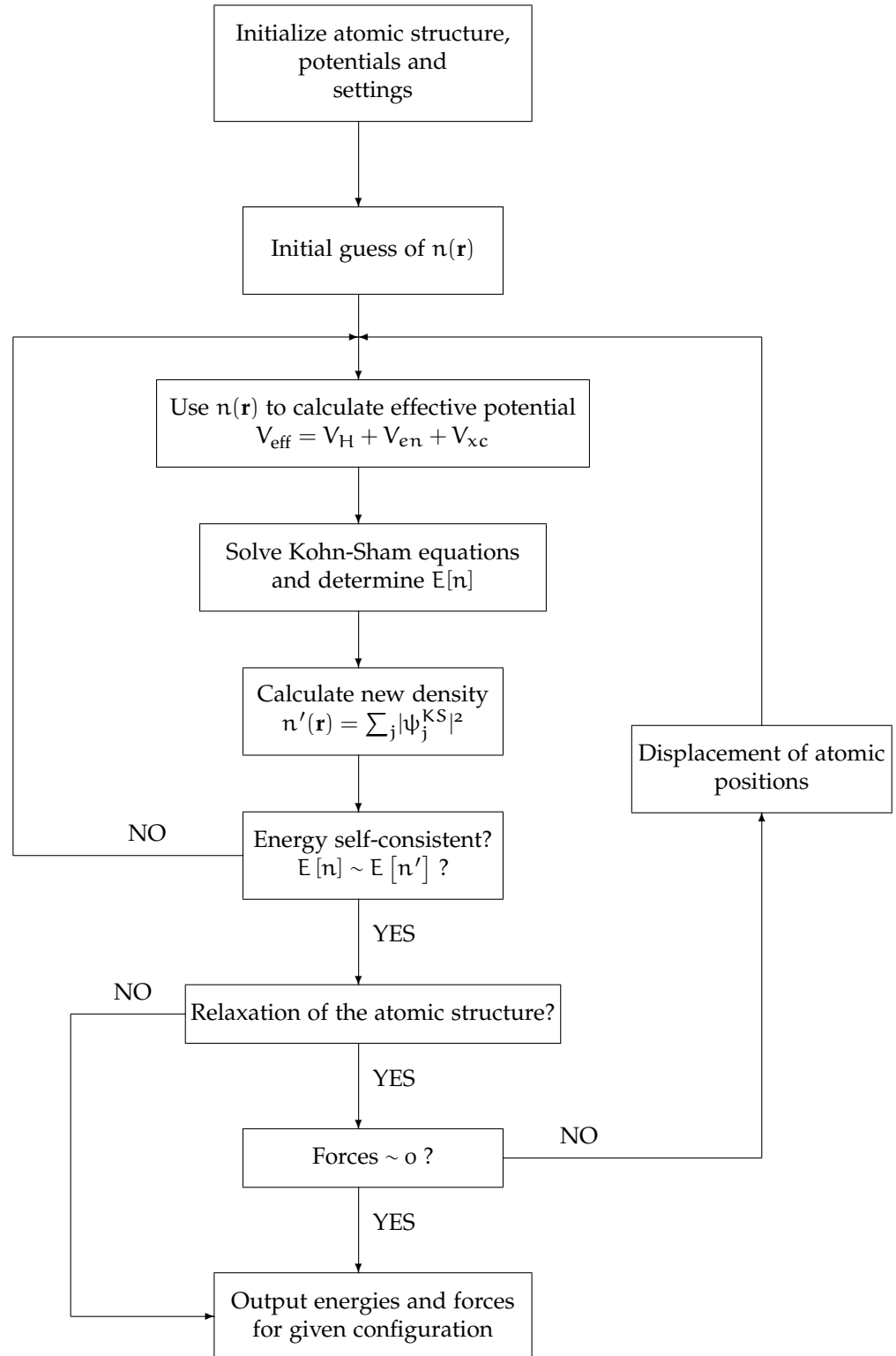
Figure A.1: A flow chart of the self-consistent field method for DFT.

Appendix B

Featurizaton

B.1 Table of featurizers

Table B.1: This thesis' chosen 39 featurizers from matminer. Descriptions are either found from Ref. [86] or from the project's Github page.

Features	Description	Original reference
Composition features		
AtomicOrbitals	Highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO).	[87]
AtomicPacking-Efficiency	Packing efficiency.	[88]
BandCenter	Estimation of absolute position of band center using geometric mean of electronegativity.	[89]
ElementFraction	Fraction of each element in a composition.	-
ElementProperty	Statistics of various element properties.	[90–92]
IonProperty	Maximum and average ionic character.	[91]
Miedema	Formation enthalpies of intermetallic compounds, solid solutions, and amorphous phases using semi-empirical Miedema model.	[93]
	Continued on next page	

63

Table B.1 – continued from previous page

Features	Description	Original reference
Stoichiometry	L^p norm-based stoichiometric attributes.	[91]
TMetalFraction	Fraction of magnetic transition metals.	[92]
ValenceOrbital	Valence orbital attributes such as the mean number of electrons in each shell.	[91]
YangSolid-Solution	Mixing thermochemistry and size mismatch terms.	[94]
Oxid composition features		
Electronegativity-Diff	Statistics on electronegativity difference between anions and cations.	[92]
OxidationStates	Statistics of oxidation states.	[92]
Structure features		
DensityFeatures	Calculate density, volume per atom and packing fraction.	-
GlobalSymmetry-Features	Determines spacegroup number, crystal system (1-7) and inversion symmetry.	-
RadialDistribution-Function	Calculates the radial distribution function of a crystal system.	-
CoulombMatrix	Generate the Coulomb matrix, which is a representation of the nuclear coulombic interaction of the input structure.	[95]
PartialRadial-Distribution-Function	Compute the partial radial distribution function of a crystal structure	[96]
SineCoulomb-Matrix	Computes a variant of the coulomb matrix developed for periodic crystals.	[97]
EwaldEnergy	Computes the energy from Coulombic interactions based on charge states of each site.	[98]

Table B.1 – continued from previous page

Features	Description	Original reference
BondFractions	Compute the fraction of each bond in a structure, based on nearest neighbours.	[99]
Structural-Heterogeneity	Calculates the variance in bond lengths and atomic volumes in a structure.	[100]
MaximumPacking-Efficiency	Calculates the maximum packing efficiency of a structure.	[100]
ChemicalOrdering	Computes how much the ordering of species differs from random in a structure.	[100]
XRDPowder-Pattern	1D array representing normalized powder diffraction of a structure as calculated by pymatgen.	[90]
Site features		
AGNI-Fingerprints	Calculates the product integral of RDF and Gaussian window function	[101]
AverageBond-Angle	Determines the average bond angle of a specific site with its nearest neighbors using pymatgens implementation.	[102]
AverageBond-Length	Determines the average bond length between one specific site and all its nearest neighbors using pymatgens implementation.	[102]
BondOrientational-Paramater	Calculates the averages of spherical harmonics of local neighbors	[103, 104]
ChemEnvSite Fingerprint	Calculates the resemblance of given sites to ideal environment using pymatgens ChemEnv package.	[105, 106]
Coordination-Number	The number of first nearest neighbors of a site	[106]

Table B.1 – continued from previous page

Features	Description	Original reference
CrystalNN-Fingerprint	A local order parameter fingerprint for periodic crystals.	-
GaussianSymm-Func	Calculates the gaussian radial and angular symmetry functions originally suggested for fitting machine learning potentials.	[107, 108]
GeneralizedRadial-Distribution-Function	Computes the general radial distribution function for a site	[103]
LocalProperty-Difference	Computes the difference in elemental properties between a site and its neighboring sites.	[100, 102]
OPSite-Fingerprint	Computes the local structure order parameters from a site's neighbor environment.	[106]
Voronoi-Fingerprint	Calculates the Voronoi tessellation-based features around a target site.	[109, 110]
Density of state features		
DOSFeaturizer	Computes top contributors to the density of states at the valence and conduction band edges. Thus includes chemical specie, orbital character, and orbital location information.	[111]
Band structure features		
BandFeaturizer	Converts a complex electronic band structure into quantities such as band gap and the norm of k point coordinates at which the conduction band minimum and valence band maximum occur.	-

B.2 Erroneous entries

MPID	Full formula	Reference
mp-555563	$PH_6C_2S_2NCl_2O_4$	[112]
mp-583476	$Nb_7S_2I_{19}$	[113]
mp-600205	$H_{10}C_5SeS_2N_3Cl$	-
mp-600217	$H_{80}C_{40}Se_8S_{16}Br_8N_{24}$	-
mp-1195290	$Ga_3Si_5P_{10}H_{36}C_{12}N_4Cl_{11}$	-
mp-1196358	$P_4H_{120}Pt_8C_{40}I_8N_4Cl_8$	-
mp-1196439	$Sn_8P_4H_{128}C_{44}N_{12}Cl_8O_4$	-
mp-1198652	$Te_4H_{72}C_{36}S_{24}N_{12}Cl_4$	-
mp-1198926	$Re_8H_{96}C_{24}S_{24}N_{48}Cl_{48}$	-
mp-1199490	$Mn_4H_{64}C_{16}S_{16}N_{32}Cl_8$	-
mp-1199686	$Mo_4P_{16}H_{152}C_{52}N_{16}Cl_{16}$	-
mp-1203403	$C_{121}S_2Cl_{20}$	-
mp-1204279	$Si_{16}Te_8H_{176}Pd_8C_{64}Cl_{16}$	-
mp-1204629	$P_{16}H_{216}C_{80}N_{32}Cl_8$	-

Table B.2: A table of manually identified entries from Materials Project that experience issues concerning Matminer's featurization tools. These were excluded from the dataset.

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

14. Griffiths, R. B. Nature and location of quantum information. *Physical Review A* **66.** doi:10.1103/physreva.66.012311 (July 2002).

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

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

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

18. 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).

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

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. Neudeck, P. G. Progress in silicon carbide semiconductor electronics technology. *Journal of Electronic Materials* **24,** 283–288 (Apr. 1995).

38. 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).

39. Lawaetz, P. Valence-Band Parameters in Cubic Semiconductors. *Physical Review B* **4,** 3460–3467 (Nov. 1971).

40. Beckers, L. *et al.* Structural and optical characterization of epitaxial waveguiding $BaTiO_3$ thin films on MgO. *Journal of Applied Physics* **83,** 3305–3310 (Mar. 1998).

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

42. Bassett, L. C., Alkauskas, A., Exarhos, A. L. & Fu, K.-M. C. Quantum defects by design. *Nanophotonics* **8,** 1867–1888 (Oct. 2019).

43. James, W. J. Theory of defects in solidsby A. M. Stoneham. *Acta Crystallographica Section A* **32,** 527–527 (May 1976).

44. Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466,** 730–734 (Aug. 2010).

45. Son, N. T. *et al.* Developing silicon carbide for quantum spintronics. *Applied Physics Letters* **116,** 190501 (May 2020).

46. Falk, A. L. *et al.* Polytype control of spin qubits in silicon carbide. *Nature Communications* **4.** doi:10.1038/ncomms2854 (May 2013).

47. Widmann, M. *et al.* Coherent control of single spins in silicon carbide at room temperature. *Nature Materials* **14,** 164–168 (Dec. 2014).

48. Zhang, G., Cheng, Y., Chou, J.-P. & Gali, A. Material platforms for defect qubits and single-photon emitters. *Applied Physics Reviews* **7,** 031308 (Sept. 2020).

49. Redjem, W. *et al.* Single artificial atoms in silicon emitting at telecom wavelengths. *Nature Electronics* **3,** 738–743 (Nov. 2020).

50. Berhane, A. M. *et al.* Photophysics of GaN single-photon emitters in the visible spectral range. *Physical Review B* **97.** doi:10.1103/physrevb.97.165202 (Apr. 2018).

51. Xue, Y. *et al.* Single-Photon Emission from Point Defects in Aluminum Nitride Films. *The Journal of Physical Chemistry Letters* **11,** 2689–2694 (Mar. 2020).

52. Varley, J. B., Janotti, A. & de Walle, C. G. V. Defects in AlN as candidates for solid-state qubits. *Physical Review B* **93.** doi:10.1103/physrevb.93.161201 (Apr. 2016).

53. Toth, M. & Aharonovich, I. Single Photon Sources in Atomically Thin Materials. *Annual Review of Physical Chemistry* **70,** 123–142 (June 2019).

54. Tran, T. T. *et al.* Robust Multicolor Single Photon Emission from Point Defects in Hexagonal Boron Nitride. *ACS Nano* **10,** 7331–7338 (July 2016).

55. Tran, T. T., Bray, K., Ford, M. J., Toth, M. & Aharonovich, I. *Quantum Emission from Hexagonal Boron Nitride Monolayers* in *Conference on Lasers and Electro-Optics* (OSA, 2016). doi:10.1364/cleo_qels.2016.ftu4d.1.

56. Weston, L., Wickramaratne, D., Mackoit, M., Alkauskas, A. & de Walle, C. G. V. Native point defects and impurities in hexagonal boron nitride. *Physical Review B* **97.** doi:10.1103/physrevb.97.214104 (June 2018).

57. Abdi, M., Chou, J.-P., Gali, A. & Plenio, M. B. Color Centers in Hexagonal Boron Nitride Monolayers: A Group Theory and Ab Initio Analysis. *ACS Photonics* **5,** 1967–1976 (Apr. 2018).

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

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

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

61. 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).

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

63. 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).

64. 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).

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

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

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

68. Tran, F. & Blaha, P. Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential. *Physical Review Letters* **102.** doi:10.1103/physrevlett.102.226401 (June 2009).

69. Becke, A. D. & Johnson, E. R. A simple effective potential for exchange. *The Journal of Chemical Physics* **124,** 221101 (June 2006).

70. Koller, D., Tran, F. & Blaha, P. Merits and limits of the modified Becke-Johnson exchange potential. *Physical Review B* **83.** doi:10.1103/physrevb.83.195134 (May 2011).

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

72. 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).

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

74. 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).

75. 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).

76. Klimeš, J., Bowler, D. R. & Michaelides, A. Chemical accuracy for the van der Waals density functional. *Journal of Physics: Condensed Matter* **22,** 022201 (Dec. 2009).

77. Dion, M., Rydberg, H., Schröder, E., Langreth, D. C. & Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Physical Review Letters* **92.** doi:10.1103/physrevlett.92.246401 (June 2004).

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

79. Murphy, K. *Machine learning : a probabilistic perspective* ISBN: 9780262018029 (MIT Press, Cambridge, Mass, 2012).

80. Wolpert, D. & Macready, W. No Free Lunch Theorems for Search (Mar. 1996).

81. Guido, S. *Introduction to Machine Learning with Python* 400 pp. ISBN: 1449369413. <https://www.ebook.de/de/product/23308778/sarah_guido_introduction_to_machine_learning_with_python.html> (O'Reilly UK Ltd., 2016).

82. Caruana, R. & Niculescu-Mizil, A. *An empirical comparison of supervised learning algorithms* in *Proceedings of the 23rd international conference on Machine learning - ICML '06* (ACM Press, 2006). doi:10.1145/1143844.1143865.

83. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics* **28,** 337–407 (Apr. 2000).

84. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29,** 1189–1232 (Oct. 2001).

85. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning* ISBN: 1461471370. <https://www.ebook.de/de/product/20292548/gareth_james_daniela_witten_trevor_hastie_robert_tibshirani_an_introduction_to_statistical_learning.html> (Springer-Verlag GmbH, 2017).

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

87. Kotochigova, S., Levine, Z. H., Shirley, E. L., Stiles, M. D. & Clark, C. W. Local-density-functional calculations of the energy of atoms. *Physical Review A* **55,** 191–199 (Jan. 1997).

88. Laws, K. J., Miracle, D. B. & Ferry, M. A predictive structural model for bulk metallic glasses. *Nature Communications* **6.** doi:10.1038/ncomms9123 (Sept. 2015).

89. Butler, M. A. & Ginley, D. S. Prediction of Flatband Potentials at Semiconductor-Electrolyte Interfaces from Atomic Electronegativities. *Journal of The Electrochemical Society* **125,** 228–232 (Feb. 1978).

90. 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).

91. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials* **2.** doi:10.1038/npjcompumats.2016.28 (Aug. 2016).

92. Deml, A. M., O'Hayre, R., Wolverton, C. & Stevanović, V. Predicting density functional theory total energies and enthalpies of formation of metal-nonmetal compounds by linear regression. *Physical Review B* **93.** doi:10.1103/physrevb.93.085142 (Feb. 2016).

93. Weeber, A. W. Application of the Miedema model to formation enthalpies and crystallisation temperatures of amorphous alloys. *Journal of Physics F: Metal Physics* **17,** 809–813 (Apr. 1987).

94. Yang, X. & Zhang, Y. Prediction of high-entropy stabilized solid-solution in multi-component alloys. *Materials Chemistry and Physics* **132,** 233–238 (Feb. 2012).

95. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Physical Review Letters* **108.** doi:10.1103/physrevlett.108.058301 (Jan. 2012).

96. Schütt, K. T. *et al.* How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Physical Review B* **89.** doi:10.1103/physrevb.89.205118 (May 2014).

97. Faber, F., Lindmaa, A., von Lilienfeld, O. A. & Armiento, R. Crystal structure representations for machine learning models of formation energies. *International Journal of Quantum Chemistry* **115,** 1094–1101 (Apr. 2015).

98. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369,** 253–287 (1921).

99. Hansen, K. *et al.* Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The Journal of Physical Chemistry Letters* **6,** 2326–2331 (June 2015).

100. Ward, L. *et al.* Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Physical Review B* **96.** doi:10.1103/physrevb.96.024104 (July 2017).

101. Botu, V. & Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *International Journal of Quantum Chemistry* **115,** 1074–1083 (Dec. 2014).

102. De Jong, M. *et al.* A Statistical Learning Framework for Materials Science: Application to Elastic Moduli of k-nary Inorganic Polycrystalline Compounds. *Scientific Reports* **6.** doi:10.1038/srep34256 (Oct. 2016).

103. Seko, A., Hayashi, H., Nakayama, K., Takahashi, A. & Tanaka, I. Representation of compounds for machine-learning prediction of physical properties. *Physical Review B* **95.** doi:10.1103/physrevb.95.144110 (Apr. 2017).

104. Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Physical Review B* **28,** 784–805 (July 1983).

105. Waroquiers, D. *et al.* Statistical Analysis of Coordination Environments in Oxides. *Chemistry of Materials* **29,** 8346–8360 (Sept. 2017).

106. Zimmermann, N. E. R., Horton, M. K., Jain, A. & Haranczyk, M. Assessing Local Structure Motifs Using Order Parameters for Motif Recognition, Interstitial Identification, and Diffusion Path Characterization. *Frontiers in Materials* **4.** doi:10.3389/fmats.2017.00034 (Nov. 2017).

107. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of Chemical Physics* **134,** 074106 (Feb. 2011).

108. Khorshidi, A. & Peterson, A. A. Amp: A modular approach to machine learning in atomistic simulations. *Computer Physics Communications* **207,** 310–324 (Oct. 2016).

109. Peng, H. L., Li, M. Z. & Wang, W. H. Structural Signature of Plastic Deformation in Metallic Glasses. *Physical Review Letters* **106.** doi:10.1103/physrevlett.106.135503 (Mar. 2011).

110. Wang, Q. & Jain, A. A transferable machine-learning framework linking interstice distribution and plastic heterogeneity in metallic glasses. *Nature Communications* **10.** doi:10.1038/s41467-019-13511-9 (Dec. 2019).

111. Dylla, M. T., Dunn, A., Anand, S., Jain, A. & Snyder, G. J. Machine Learning Chemical Guidelines for Engineering Electronic Structures in Half-Heusler Thermoelectric Materials. *Research* **2020,** 1–8 (Apr. 2020).

112. None Available. *Materials Data on PH6C2S2N(ClO2)2 by Materials Project* en. 2020. doi:10.17188/1268877.

113. None Available. *Materials Data on Nb7S2I19 by Materials Project* en. 2014. doi:10.17188/1277059.