

PREDICTIONS  
OF QUBITS-CANDIDATES USING  
MACHINE LEARNING

by

Oliver Lerstøl Hebnes

THESIS  
for the degree of  
MASTER OF SCIENCE



Faculty of Mathematics and Natural Sciences  
University of Oslo

October 9, 2020



# Contents

<b>I</b>	<b>Theory</b>	<b>1</b>
<b>1</b>	<b>Quantum mechanics</b>	<b>3</b>
1.1	The single-electron Schrödinger equation . . . . .	3
1.1.1	Eigenfunctions . . . . .	5
1.2	Variational principle . . . . .	6
1.3	Hartree-Fock approximations . . . . .	7
1.4	The many-particle Schrödinger equation . . . . .	7
1.4.1	Challenges . . . . .	8
1.5	The Born-Oppenheimer approximation . . . . .	9
1.6	The density functional theory . . . . .	10
1.6.1	Hohenberg-Kohn's theorems . . . . .	11
1.6.2	The Kohn-Sham equation . . . . .	13
1.6.3	The exchange-correlation energy . . . . .	14
<b>2</b>	<b>Identifying materials</b>	<b>17</b>
2.1	Crystallography . . . . .	17
2.2	Semiconductors . . . . .	18
<b>II</b>	<b>Material Science Databases</b>	<b>19</b>
<b>III</b>	<b>Implementation</b>	<b>21</b>
<b>IV</b>	<b>Results</b>	<b>23</b>
<b>V</b>	<b>Conclusion</b>	<b>25</b>
<b>VI</b>	<b>Appendices</b>	<b>27</b>



# **Part I**

## **Theory**



# Chapter 1

## Quantum mechanics

In the purpose of fully understanding the underlying physics behind computational material science, we will need to investigate how we can calculate the forces happening inside a crystal. This is a field of expertise where classical models gives inaccurate estimates, thus it is inevitable to start off with some fundamental quantum mechanics.

We will in this thesis only formulate the necessary theory behind density functional theory, leaving most of the quantum-mechanical world untouched. However, the fundamental theory remains the same and we will start our venture with the single-electron Schrödinger equation.

### 1.1 The single-electron Schrödinger equation

As every other introduction to quantum mechanics books, we will start of investigating the Schrödinger equation with only one electron [1]

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V\Psi \quad (1.1)$$

for a convenient external potential  $V_{\text{ext}}(r)$  that is independent of time. We will try to look for solutions for (1.1) by separating the wave function into a space-dependent and time-dependent function

$$\Psi(r, t) = \psi(r)\phi(t). \quad (1.2)$$

By inserting ordinary derivatives and dividing each side with equation (1.2), our Schrödinger equation (1.1) now reads

$$i\hbar \frac{1}{\phi(t)} \frac{d\phi(t)}{dt} = -\frac{\hbar^2}{2m} \frac{1}{\psi(r)} \nabla^2 \psi + V(r) \quad (1.3)$$

Since the potential function  $V(r)$  is independent of time, we observe the time and space dependencies of each side and state the fact that both sides has to be constant. Thus, two intriguing equations unveil themselves;

$$i\hbar \frac{1}{\phi(t)} \frac{d\phi(t)}{dt} = E\phi(t) \quad (1.4)$$

and

$$\frac{\hbar^2}{2m} \frac{1}{\psi(r)} \nabla^2 \psi + V(r) = E\psi(r) \quad (1.5)$$

where the first equation (1.4) has a general solution  $\phi(t) = C \exp(-iEt/\hbar)$  and  $C = 1$  after normalization, and the second equation (1.5) is known as time-independent Schrödinger equation. These two equations are connected through the variable  $\epsilon$ .

By utilizing variable separation to get equation (1.2), we find that the wavefunction is describing a stationary state with probability density

$$\begin{aligned} |\Psi(r, t)|^2 &= \Psi^* \Psi \\ &= \Psi^* e^{iEt/\hbar} \Psi e^{-iEt/\hbar} \\ &= |\Psi(r)|^2 \end{aligned}$$

that is independent of time. Conveniently, this is also true for every expectation value; they are all constant in time. We can also try to express this in classical terms regarding the Hamiltonian, which in this scenario is defined as

$$\hat{H}(r, p) = \frac{p^2}{2m} + V(r) = -\frac{\hbar^2}{2m} \nabla^2 + V(r) \quad (1.6)$$

simplifying equation 1.5 to

$$\hat{H}\psi = E\psi \quad (1.7)$$

and we can find the expectation value of the total energy as

$$\begin{aligned} \langle H \rangle &= \int \psi^* \hat{H} \psi dr \\ &= E \int |\Psi|^2 dr \\ &= E \end{aligned}$$



using the fact that expectation values are constant in time for stationary states. Similarly, we can try to estimate the variance of the Hamiltonian,

$$\begin{aligned}\sigma_H^2 &= \langle H^2 \rangle - \langle H \rangle^2 \\ &= E^2 - E^2 \\ &= 0\end{aligned}$$

which appropriately describes that every measurement of the total energy is certain to return the value  $E$ .

### 1.1.1 Eigenfunctions

So far, we have not given an explanation of what a wavefunction is. As a matter of fact, we have actually found an eigenfunction

$$\psi_\kappa(\mathbf{r}, t) = \psi_\kappa e^{-i\varepsilon_\kappa t/\hbar}$$

where  $\kappa$  denotes the  $k$ -th eigenfunction and  $\varepsilon_\kappa$  is its corresponding energy eigenvalue. The eigenfunctions have distinct energies and have the attribute that they are orthogonal and normalized with respect to

$$\langle \psi_\kappa(\mathbf{r}, t) | \psi_{\kappa'}(\mathbf{r}, t) \rangle = \delta_{\kappa\kappa'}.$$

The state with the lowest energy is called the ground state, and is where it is most likely to find an electron in a single-electron system with no external potential applied.

A general wavefunction can be generated by a summation of eigenfunctions (such as the eigenfunction in the latter case)

$$\Psi(\mathbf{r}, t) = \sum_{\kappa} c_{\kappa} \psi_{\kappa}(\mathbf{r}, t), \quad (1.8)$$

where  $c_{\kappa}$  is a constant. A general wavefunction does not necessarily describe stationary states, and consequently does not have distinct energies but is rather represented statistically from the expectation value

$$E = \sum_{\kappa} |c_{\kappa}|^2 \varepsilon_{\kappa}.$$

Solving Schrödinger equation for a general wavefunction is rather troublesome. Fortunately, we can use the eigenfunctions instead, transforming equation 1.7 into time-independent Schrödinger equation for eigenfunctions

$$\hat{H}\psi_{\kappa}(\mathbf{r}) = \varepsilon_{\kappa}\psi_{\kappa}(\mathbf{r}). \quad (1.9)$$

The shape of an eigenfunction has normally high spatial symmetry that depends on the symmetry of the potential  $V_{\text{ext}}(\mathbf{r})$  and the boundary conditions [2]. The study of how atoms in a crystalline interact with each other is of upmost importance when trying to explain macroscopic consequences.

## 1.2 Variational principle

So far we have tried to solve the time-independent Schrödinger equation with the use of an *ansatz*. Alas, for complex many-body systems, there exists easier<sup>1</sup> approaches that happen to make solid estimates of the ground state energy when a good guess of eigenfunctions are out of range. The variational principle is one of them, and it states that the energy of any trial wavefunction is always an upper bound to the exact ground state energy by definition  $E_0$ .

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle \leq \langle \psi | H | \psi \rangle = E \quad (1.10)$$

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

$$\Psi = \sum_n c_n \psi_n, \quad \text{where} \quad H\psi_n = E_n \psi_n \quad (1.11)$$

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

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

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  $\Psi$  is normalized, thus  $\sum_n |c_n|^2 = 1$ , and the expectation value conveniently is bound to follow equation 1.10. 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;

$$E_0 = \min_{\Psi \rightarrow \Psi_0} \langle \Psi | H | \Psi \rangle. \quad (1.12)$$

This plays a vital role later, as we will see, since the energy is a *functional* of the the wavefunction, denoted as  $E_0[\Psi]$ .

---

<sup>1</sup>and powerful

### 1.3 Hartree-Fock approximations

As we venture along from a one-electron system to a two-electron system, we encounter a new wavefunction and Hamiltonian that needs to describe two particles, making the two-electron Schrödinger equation read

$$\left(-\frac{\hbar^2 \nabla_1^2}{2m_e} - \frac{\hbar^2 \nabla_2^2}{2m_e} + \frac{q^2}{|r_1 - r_2|} + V_{\text{ext}}(r)\right) \Psi_{\kappa}(r_1, r_2) = E_{\kappa} \Psi_{\kappa}(r_1, r_2), \quad (1.13)$$

where the two first terms are the kinetic energies of the electrons, while the third term is a potential that describes the repulsive Coulomb interaction between the two electrons. The last term is the external potential, well known from the earlier scenario with only one electron.

The Hartree approximation to the two-electron wavefunction is to make an *ansatz*, a clever guess, for the wavefunction

$$\Psi(r_1, r_2) = A \cdot \psi_1(r_1) \psi_2(r_2). \quad (1.14)$$

The downside with this approach is 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;

$$\Psi(r_1, r_2) = \frac{1}{\sqrt{2}} \left( \psi_1(r_1) \psi_2(r_2) - \psi_1(r_2) \psi_2(r_1) \right) \quad (1.15)$$

### 1.4 The many-particle Schrödinger equation

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

A simple electron with mass  $m_e$  wandering off in its own system, minding its own business, will be in possession of some kinetic energy. If it's a whole legion of electrons, say  $N_e$  electrons, they will have the total kinetic energy

$$T_e = - \sum_{j=1}^{N_e} \frac{\hbar^2 \nabla_j^2}{2m_e}. \quad (1.16)$$

All the electrons are negatively charged, which means that there will exist repulsive Coulomb interactions between each and every electron, totalling to

$$U_{ee} = \sum_{j=1}^{N_e} \sum_{j' < j} \frac{q^2}{|r_j - r_{j'}|}. \quad (1.17)$$

The summation makes sure to not count each interaction more than one time. At the same time, in a different system, a horde of nuclei with each their mass  $m_n$  are roaming their own system. If there are  $N_n$  nuclei there, they would accumulate the kinetic energy

$$T_n = - \sum_{a=1}^{N_n} \frac{\hbar^2 \nabla_a}{2m_n}. \quad (1.18)$$

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

$$U_{nn} = \sum_{a=1}^{N_n} \sum_{a' < a} \frac{q^2 Z_a Z_{a'}}{|R_a - R_{a'}|}. \quad (1.19)$$

where  $Z_a$  is the atom number of the nuclei.

Now, imagine if we brought the two systems together to one system. The equations describing kinetic energy and interactions of each respective particle would still be legit, however, we need to include the attractive interactions between the nuclei and the electrons

$$U_{en} = - \sum_{j=1}^{N_e} \sum_{a=1}^{N_n} \frac{q^2 Z_a}{|r_j - R_a|} \quad (1.20)$$

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

$$\begin{aligned} \hat{H} = & - \sum_{j=1}^{N_e} \frac{\hbar^2 \nabla_j}{2m_e} - \sum_{a=1}^{N_n} \frac{\hbar^2 \nabla_a}{2m_n} + \sum_{j=1}^{N_e} \sum_{j' < j} \frac{q^2}{|r_j - r_{j'}|} \\ & + \sum_{a=1}^{N_n} \sum_{a' < a} \frac{q^2 Z_a Z_{a'}}{|R_a - R_{a'}|} - \sum_{j=1}^{N_e} \sum_{a=1}^{N_n} \frac{q^2 Z_a}{|r_j - R_a|}. \end{aligned} \quad (1.21)$$

### 1.4.1 Challenges

A few problems arise when trying to solve the many-particle Schrödinger equation.

Firstly, the amount of atoms in a crystal is absurd. As an example, we can numerically try to calculate the equation 1.17 for a  $1\text{mm}^3$  silicon-crystal that contains  $7 \cdot 10^{20}$  electrons. For this particular problem, we will pretend to use the current fastest supercomputer Fugaku [3] that can calculate 514 TFlop/s, and we will assume that we need 2000 flops to calculate each term inside the sum [2], 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 amount of particles translates into a numerical problem.

Secondly, while the many-particle Hamiltonian contains operators that has to be applied to single-particle wavefunctions, we do not know how to do it as we have no prior knowledge of how  $\Psi$  depends on the them.

## 1.5 The Born-Oppenheimer approximation

The many-particle eigenfunction describes the wavefunction of all the electrons and nuclei and we denote it as  $\Psi_\kappa^{\text{en}}$  for a nucleus and electron part. The Born-oppenheimer approximation assumes that nuclei, that have a substantial larger mass than electrons, can be treated as fixed point charges. According to this assumption, it is valid to separate the eigenfunction into an electronic part and a nucleus part

$$\Psi_\kappa(\mathbf{r}, \mathbf{R}) \approx \psi_\kappa(\mathbf{r}, \mathbf{R})\Theta_\kappa(\mathbf{R}). \quad (1.22)$$

As seen, the electronic part is dependent on the nuclei. This is according to the assumption above, since electrons can respond instantaneous to a new position of the much slower nucleus, but this is not true for the opposite scenario. For our advantage, we already have knowledge of the terms in the many-particle Hamiltonian, and we can start separating the Hamiltonian into an electronic and nucleus part.

$$\hat{H}^{\text{en}} = \overbrace{\hat{T}_e + U_{ee} + U_{en}}^{\hat{H}^e} + \overbrace{\hat{T}_n + U_{nn}}^{\hat{H}^n}. \quad (1.23)$$

Starting from the time-independent Schrödinger equation, we can try to make separate expressions for the electronic and the nucleus Schrödinger equation.

$$\hat{H}^{\text{en}}\Psi_\kappa^{\text{en}}(\mathbf{r}, \mathbf{R}) = E_\kappa^{\text{en}}\psi_\kappa^{\text{en}}(\mathbf{r}, \mathbf{R}) \quad | \times \int \Psi^*(\mathbf{r}, \mathbf{R})d\mathbf{r} \quad (1.24)$$

$$\int \psi_\kappa^*(\mathbf{r}, \mathbf{R})(\hat{H}^e + \hat{H}^n)\psi_\kappa(\mathbf{r}, \mathbf{R})\Theta_\kappa(\mathbf{R})d\mathbf{r} = E_\kappa^{\text{en}} \underbrace{\int \psi_\kappa^*(\mathbf{r}, \mathbf{R})\psi_\kappa(\mathbf{r}, \mathbf{R})d\mathbf{r}}_1 \Theta_\kappa(\mathbf{R}). \quad (1.25)$$

Since  $\Theta_\kappa(\mathbf{R})$  is independent of the electronic space, we get  $E_\kappa$  as the total energy of the electrons in the state  $\kappa$ .

$$E_\kappa(\mathbf{R})\Theta_\kappa(\mathbf{R}) + \int \Psi_\kappa^*(\mathbf{r}, \mathbf{R})\hat{H}^n\Psi_\kappa(\mathbf{r}, \mathbf{R})\Theta_\kappa(\mathbf{R})d\mathbf{r} = E_\kappa^{\text{en}}\Theta_\kappa(\mathbf{R}). \quad (1.26)$$

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

$$\left(T_n + T'_n + T''_n + U_{aa} + E_\kappa(R)\right)\Theta_\kappa(R) = E_\kappa^{en}\Theta_\kappa(R). \quad (1.27)$$

If we neglect  $T'_n$  and  $T''_n$  to lower the computational efforts, we obtain the Born-Oppenheimer approximation with the electronic eigenfunction as

$$(T_e + U_{ee} + U_{en})\Psi_\kappa(r, R) = E_\kappa(R)\Psi_\kappa(r, R) \quad (1.28)$$

and the nuclei eigenfunction as

$$(T_n + U_{nn} + E_\kappa(r, R))\Theta_\kappa(R) = E_\kappa^{en}(R)\Psi_\kappa(r, R). \quad (1.29)$$

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

## 1.6 The density functional theory

Up until now 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 electronic density number. One fundamental problem that exists when trying to solve the many-electron Schrödinger equation is that the wavefunction is a bloody complicated function that depends on  $3N_e$  variables<sup>2</sup>, which are the spatial dimensions.

Hohenberg and Kohn [4] 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 [5]. 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 [2]. A consequence of this is that the amount of variables are reduced from  $3N_e$  to 3, significantly reducing the computational efforts.

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

We will start this section with a discussion of the Hohenberg-Kohn theorems, before we dwell further into the Kohn-sham equation.

---

<sup>2</sup>not including spin

### 1.6.1 Hohenberg-Kohn's theorems

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

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

$$E^{(1)} = \langle \Psi^{(1)} | \hat{H}^{(1)} | \Psi^{(1)} \rangle < \langle \Psi^{(2)} | \hat{H}^{(1)} | \Psi^{(2)} \rangle \quad (1.30)$$

$$E^{(2)} = \langle \Psi^{(2)} | \hat{H}^{(2)} | \Psi^{(2)} \rangle < \langle \Psi^{(1)} | \hat{H}^{(2)} | \Psi^{(1)} \rangle \quad (1.31)$$

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 1.30 as

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

Thus,

$$E^{(1)} = E^{(2)} + \int (V_{\text{ext}}^{(1)} - V_{\text{ext}}^{(2)}) n(r) dr \quad (1.32)$$

Similar procedure be done for  $E^{(2)}$  in equation 1.31, resulting in

$$E^{(2)} = E^{(1)} + \int \left( V_{\text{ext}}^{(2)} - V_{\text{ext}}^{(1)} \right) n(r) dr. \quad (1.33)$$

If we add these two equations together, we get

$$\begin{aligned} E^{(1)} + E^{(2)} &< E^{(2)} + E^{(1)} + \int \left( V_{\text{ext}}^{(1)} - V_{\text{ext}}^{(2)} \right) n(r) dr \\ &\quad + \int \left( V_{\text{ext}}^{(2)} - V_{\text{ext}}^{(1)} \right) n(r) dr \\ E^{(1)} + E^{(2)} &< E^{(2)} + E^{(1)} \end{aligned} \quad (1.34)$$

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

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

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

$$E[n] = \underbrace{T[n] + U_{\text{ee}}[n]}_{F[n]} + \underbrace{\int V_{\text{ext}} n(r) dr}_{U_{\text{ext}}[n]} \quad (1.35)$$

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  $n_0(r)$ ,

$$E_0 = E[n_0] = \langle \Psi_0 | H | \Psi_0 \rangle. \quad (1.36)$$

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

$$E_0 = E[n_0] = \langle \Psi_0 | H | \Psi_0 \rangle < \langle \Psi | H | \Psi \rangle = E[n] \quad (1.37)$$

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



### 1.6.2 The Kohn-Sham equation

So far, we have covered alternative methods to try to solve the many-electron Schrödinger equation, 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. The challenge is that we do not know how to express the energy in terms of the density, so we should strive for other methods in the endeavour of solving the many-particle Schrödinger equation.

In 1965, Kohn and Sham [6] published a paper where they used the Hohenberg-Kohn theorems to derive the Kohn-Sham (KS) equation. They assumed it was possible to generate the exact ground-state density  $n_0(\mathbf{r})$  by using a Hartree-like total wavefunction

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_e}) = \psi_1^{\text{KS}}(\mathbf{r}_1) \psi_2^{\text{KS}}(\mathbf{r}_2) \dots \psi_{N_e}^{\text{KS}}(\mathbf{r}_{N_e}) \quad (1.38)$$

where  $\psi_j^{\text{KS}}(\mathbf{r}_j)$  are some auxiliary independent single-particle wavefunctions. For the observable reader, we can share that the Kohn-Sham wavefunctions cannot be the correct single-particle wavefunctions since our ansatz implies an exact density

$$n(\mathbf{r}) = \sum_{j=1}^{N_e} |\psi_j^{\text{KS}}(\mathbf{r})|^2. \quad (1.39)$$

Recalling equation 1.35 describes the total energy as a functional of the density,

$$E[n] = T[n] + U_{ee}[n] + U_{en}[n], \quad (1.40)$$

we try to modify it to include the kinetic energy  $T_s[n]$  and the interaction energy  $U_s[n]$ , which are known terms.

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

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

$$E_{xc}[n] = \Delta T + \Delta U = (T[n] - T_s[n]) + (U_{ee}[n] - U_s[n]), \quad (1.41)$$

which contains the complex many-electron interaction in addition to a kinetic energy part. Thus, we have transformed the many-electron problem into an unknown exchange-correlation energy. For non-interacting system,  $E_{xc}[n]$  is

conveniently zero, but in interacting systems it most likely is a very cumbersome 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 can now be expressed as

$$\begin{aligned}
 E[n] = & \underbrace{\sum_j \int \psi_j^{\text{KS}*} \frac{-\hbar^2 \nabla^2}{2m} \psi_j^{\text{KS}} d\mathbf{r}}_{T_s[n]} + \underbrace{\frac{1}{2} \iint q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'}_{U_s[n]} \\
 & + \underbrace{\int V_{\text{en}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}}_{U_{\text{en}}[n]} + \underbrace{(T[n] - T_s[n]) + (U_{\text{ee}}[n] - U_s[n])}_{E_{\text{xc}}[n]}.
 \end{aligned} \tag{1.42}$$

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

$$\left\{ -\frac{\hbar^2}{2m_e} \nabla_s^2 + V_{\text{H}}(\mathbf{r}) + V_{j\alpha}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) \right\} \psi_s^{\text{KS}}(\mathbf{r}) = \epsilon_s^{\text{KS}} \psi_s^{\text{KS}}(\mathbf{r}) \tag{1.43}$$

where  $V_{\text{xc}}(\mathbf{r}) = \partial E_{\text{xc}}[n]/\partial n(\mathbf{r})$  and  $V_{\text{H}}(\mathbf{r}) = \int q^2 \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}'$  is the Hartree potential describing the electron-electron interaction. Note that  $V_{\text{H}}(\mathbf{r})$  contains self-interaction contribution, however this will be taken care of in  $V_{\text{xc}}$ .

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

$$E[n] = \sum_j \epsilon_j^{\text{KS}} - \frac{1}{2} \iint q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{\text{xc}}[n] - \int V_{\text{xc}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \tag{1.44}$$

If  $V_{\text{xc}}$  is exact, and  $E[n]$  gives the true total energy, we still do not know if the energy eigenvalues  $\epsilon_s^{\text{KS}}$  are the true single-electron eigenvalues. However, there exists one exception, which is 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  $E_{\text{xc}}[n]$  as a functional of the density  $n(\mathbf{r})$ . 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. This is, unfortunately, not the case, since it remains unknown.

### 1.6.3 The exchange-correlation energy

The development of

There exists one scenario where we can derive the exact expression of the exchange-correlation functional, namely the *homogeneous electron gas*. However, this has a natural cause, since by definition  $n(\mathbf{r}) = \text{constant}$  for this situation. This is unfortunate, because it is the variations in electron density that are the merits of the calculations. The *local density approximation* (LDA) is an approximation based on this approach, where the local density is the only thing 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 [6]:

$$V_{xc}(\mathbf{r}) = V_{xc}^{\text{electron gas}}[n(\mathbf{r})]. \quad (1.45)$$

This is the simplest and most known approximation of 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 artificial repulsion and increased electron delocalization [7]. In addition, LDA has proven challenging with the use of atoms and molecules given their rapid varying electron densities, however, LDA has been seen as very successful for bulk materials because of the slowly varying electron density [8]. Together with a relative low computational cost, the LDA overall has seen as successful within bulk-materials.



# Chapter 2

## Identifying materials

### 2.1 Crystallography

Solid materials 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 <sup>1</sup>.

However, the atoms can also be found periodic in small regions of the material, classifying the material as a *polycrystalline solid*. All ceramics are polycrystalline with a broad specter of usage ranging from kitchen-porcelain to orthopedical bio-implant [9].

A third option is to have these atoms arranged infinite periodically, making the material a *crystalline solid* or more commonly named a *crystal*.

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, making a *crystal*. For every crystal, there exists periodically repeated building blocks called *cells* which represents the entire crystal. The smallest cell possible is called a *primitive cell*, but such a cell only allows lattice points at its corners and it is often quite rigid to work with when the structure becomes complex. As a solution, we will consider the *unit cell*, which allows lattice points on face centers and body centers.

---

<sup>1</sup>Need source

## 2.2 Semiconductors

Define semiconductors

A material that conducts electrical current is, by definition, a metal. On the other hand, a material that does not conduct electrical current is an insulator. A semiconductor is an element or a compound that is

Semiconductors are elements or compounds that

# **Part II**

## **Material Science Databases**





# **Part III**

## **Implementation**



# **Part IV**

## **Results**



# **Part V**

## **Conclusion**



# **Part VI**

## **Appendices**





COFFEEEEEEEEEE appendix



# Bibliography

1. Griffiths, D. *Introduction to quantum mechanics* ISBN: 9781107179868 (Cambridge University Press, Cambridge, 2017).
2. Persson, C. *Brief Introduction to the density functional theory* 2020.
3. Top500. *SUPERCOMPUTER FUGAKU* June 2020. <https://www.top500.org/system/179807/> (2020).
4. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **136**, B864–B871 (Nov. 1964).
5. Toulouse, J. *Introduction to density-functional theory* Sept. 2019. [http://www.lct.jussieu.fr/pagesperso/toulouse/enseignement/introduction\\_dft.pdf](http://www.lct.jussieu.fr/pagesperso/toulouse/enseignement/introduction_dft.pdf) (2020).
6. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140**, A1133–A1138 (Nov. 1965).
7. 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).
8. 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](https://www.ebook.de/de/product/7207845/david_sholl_janice_a_steckel_density_functional_theory_a_practical_introduction.html) (WILEY, Apr. 1, 2009).
9. 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](https://doi.org/10.1016/b978-0-08-102205-4.00010-6).
10. Balachandran, P. V. *et al.* Predictions of new ABO<sub>3</sub> perovskite compounds by combining machine learning and density functional theory. *Physical Review Materials* **2**. doi:[10.1103/physrevmaterials.2.043802](https://doi.org/10.1103/physrevmaterials.2.043802) (Apr. 2018).
11. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* (2011). arXiv: [1201.0490v4](https://arxiv.org/abs/1201.0490v4) [cs.LG] (Jan. 2, 2012).
12. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2**, 18–22 (2002).

13. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29**, 1189–1232 (Oct. 2001).
14. Pilania, G., Balachandran, P. V., Gubernatis, J. E. & Lookman, T. Classification of  $\text{ABO}_3$  perovskite solids: a machine learning study. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* **71**, 507–513 (Sept. 2015).
15. Ward, L. *et al.* Matminer: An open source toolkit for materials data mining. *Computational Materials Science* **152**, 60–69 (Sept. 2018).
16. Kirklin, S. *et al.* The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **1**. doi:[10.1038/npjcompumats.2015.10](https://doi.org/10.1038/npjcompumats.2015.10) (Dec. 2015).
17. Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (June 2017).
18. 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).
19. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *Journal of Physics: Materials* **2**, 036001 (May 2019).
20. Curtarolo, S. *et al.* AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science* **58**, 227–235 (June 2012).
21. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002 (July 2013).
22. Ong, S. P. *et al.* The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Computational Materials Science* **97**, 209–215 (Feb. 2015).
23. Levin, I. NIST Inorganic Crystal Structure Database (ICSD) en. 2020. doi:[10.18434/M32147](https://doi.org/10.18434/M32147).
24. Choudhary, K. *et al.* JARVIS: An Integrated Infrastructure for Data-driven Materials Design. arXiv: [2007.01831v1](https://arxiv.org/abs/2007.01831) [cond-mat.mtrl-sci] (July 3, 2020).
25. Landis, D. D. *et al.* The Computational Materials Repository. *Computing in Science & Engineering* **14**, 51–57 (Nov. 2012).
26. Tanaka, I., Rajan, K. & Wolverton, C. Data-centric science for materials innovation. *MRS Bulletin* **43**, 659–663 (Sept. 2018).

27. Choudhary, K., Cheon, G., Reed, E. & Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Physical Review B* **98**. doi:[10.1103/physrevb.98.014107](https://doi.org/10.1103/physrevb.98.014107) (July 2018).
28. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **65**, 1501–1509 (Sept. 2013).