Predicting solid-state qubit candidates

by

Oliver Lerstøl Hebnes

Thesis

for the degree of

Master of Science

Faculty of Mathematics and Natural Sciences
University of Oslo

October 29, 2020

Abstract

This is an abstract. First coffee.

Acknowledgements

Acknowledgements. Coffe-time?

Contents

Part I

Introduction

Chapter 1

Introduction

This it The introduction. Another coffee.

1.1 Structure of thesis

1.2 Goals

Part II

Theory

Chapter 2

Introduction to quantum mechanics

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

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

2.1 The Schrödinger equation

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

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

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

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

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

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

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

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

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

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

The symmetry of an eigenfunction depends on the symmetry of the potential $V_{ext}(\mathbf{r})$ and the boundary conditions [1].

2.2 The many-particle Schrödinger equation

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

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

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

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

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

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

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

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

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

where Z_a is the atom number of nuclei number a.

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

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

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

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

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

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

2.3 The Born-Oppenheimer approximation

The many-particle eigenfunction describes the wavefunction of all the electrons and nuclei and we denote it as Ψ_κ^{en} for electrons (e) and nuclei (n), respectively. The Born-oppenheimer approximation assumes that nuclei, of

substantially larger mass than electrons, can be treated as fixed point charges. According to this assumption, we can separate the eigenfunction into an electronic part and a nuclear part,

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

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

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

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

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

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

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

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

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

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

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

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

and the nuclear eigenfunction as

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

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

2.4 The Hartree and Hartree-Fock approximation

The next question in line is to find a wavefunction $\Psi(\mathbf{r}, \mathbf{R})$ that depends on all of the electrons in the system. The Hartre [1] approximation to this is to assume that electrons can be described independently, suggesting the *ansatz* for a two-electron wavefunction

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

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

The Hartree-fock approach, however, overcame this challenge and presented an anti-symmetric wavefunction that made the electrons indistinguishable [3]:

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

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

2.5 Variational principle

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

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

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

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

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

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

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

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

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

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

Chapter 3

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 [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 [1]. A consequence of this is that the number of variables is reduced from $3N_e$ to 3, significantly reducing the computational efforts.

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

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

3.1 The Hohenberg-Kohn theorems

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

PROOF. Assume that two external potentials $V_{ext}^{(1)}$ and $V_{ext}^{(2)}$, that differ by more than a constant, have the same ground state density $n_0(r)$. The two different

[1] not including spin

potentials correspond to distinct Hamiltonians $\hat{H}^{(1)}_{ext}$ and $\hat{H}^{(2)}_{ext}$, which again give rise to distinct wavefunctions $\Psi^{(1)}_{ext}$ and $\Psi^{(2)}_{ext}$. Utilizing the variational principle, we find that no wavefunction can give an energy that is less than the energy of $\Psi^{(1)}_{ext}$ for $\hat{H}^{(1)}_{ext}$, that is

$$E^{(1)} = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle < \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle \tag{3.1}$$

and

$$E^{(2)} = \left\langle \Psi^{(2)} \middle| \hat{H}^{(2)} \middle| \Psi^{(2)} \right\rangle < \left\langle \Psi^{(1)} \middle| \hat{H}^{(2)} \middle| \Psi^{(1)} \right\rangle. \tag{3.2}$$

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 3.1 as

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

Thus,

$$E^{(1)} = E^{(2)} + \int \left(V^{(1)}_{ext} - V^{(2)}_{ext} \right) n(\mathbf{r}) \, d\mathbf{r} \tag{3.3}$$

A similar procedure can be performed for $E^{(2)}$ in equation 3.2, resulting in

$$E^{(2)} = E^{(1)} + \int \left(V^{(2)}_{ext} - V^{(1)}_{ext} \right) n(\mathbf{r}) \, d\mathbf{r}. \tag{3.4}$$

If we add these two equations together, we get

$$\begin{aligned} E^{(1)} + E^{(2)} < E^{(2)} + E^{(1)} &+ \int \left(V_{ext}^{(1)} - V_{ext}^{(2)} n(\mathbf{r}) \, d\mathbf{r} \right) \\ &+ \int \left(V_{ext}^{(2)} - V_{ext}^{(1)} n(\mathbf{r}) \, d\mathbf{r} \right) \\ E^{(1)} + E^{(2)} < E^{(2)} &+ E^{(1)}, \end{aligned} \tag{3.5}$$

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

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

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

$$E[n] = \overbrace{T[n] + U_{ee}[n]}^{F[n]} + \overbrace{U_{en}[n]}^{\int V_{en} n(r) dr} \tag{3.6}$$

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 $n_0(r)$,

$$E_0 = E[n_0] = \langle \Psi_0 | H | \Psi_0 \rangle . \tag{3.7}$$

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

$$E_0 = E[n_0] = \langle \Psi_0 | H | \Psi_0 \rangle < \langle \Psi | H | \Psi \rangle = E[n] \tag{3.8}$$

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

3.2 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 [6] reformulated the Hohenberg-Kohn theorems by generating the exact ground-state density $n_0(r)$ using a Hartree-like total wavefunction

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, .., \mathbf{r}_{N_e}) = \psi_1^{KS}(\mathbf{r}_2)\psi_2^{KS}(\mathbf{r}_2)...\psi_{N_e}^{KS}(\mathbf{r}_{N_e}), \tag{3.9}$$

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

$$n(\mathbf{r}) = \sum_{j=1}^{N_e} |\psi_j^{KS}(\mathbf{r})|^2. \tag{3.10}$$

Recalling that equation 3.6 describes the total energy as a functional of the density,

$$E[n] = T[n] + U_{ee}[n] + U_{en}[n], \tag{3.11}$$

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

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

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

$$E_{xc}[n] = \Delta T + \Delta U = \big(T[n] - T_s[n]\big) + \big(U_{ee}[n] - U_s[n]\big), \tag{3.12}$$

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

The exact total energy functional can now be expressed as

$$\begin{aligned} E[n] = &\overbrace{\sum_j \int \psi_j^{KS*} \frac{-\hbar^2\nabla^2}{2m} \psi_j^{KS} d\mathbf{r}}^{T_s[n]} + \overbrace{\frac{1}{2}\int\int q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'}^{U_s[n]} \\ &+ \underbrace{\int V_{en}(\mathbf{r})n(\mathbf{r})d\mathbf{r}}_{U_{en}[n]} + \underbrace{\big(T[n] - T_s[n]\big) + \big(U_{ee}[n] - U_s[n]\big)}_{E_{xc}[n]}. \end{aligned} \tag{3.13}$$

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

$$\left\{-\frac{\hbar^2}{2m_e}\nabla_s^2 + V_H(\mathbf{r}) + V_{j\alpha}(\mathbf{r}) + V_{xc}(\mathbf{r})\right\}\psi_s^{KS}(\mathbf{r}) = \epsilon_s^{KS}\psi_s^{KS}(\mathbf{r}) \tag{3.14}$$

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

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

$$E[n] = \sum_j \epsilon_j^{KS} - \frac{1}{2}\int\int q^2 \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{xc}[n] - \int V_{xc}(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \tag{3.15}$$

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.

3.3 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 [6]:

$$V_{xc}(\mathbf{r}) = V_{xc}^{\text{electron gas}}\left[n(\mathbf{r})\right]. \tag{3.16}$$

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 [7]. 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 [8]. Considering the relatively low computational cost and high accuracy, the LDA overall makes a good model for estimation of the exchange-correlation functional for bulk-materials.

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

$$V_{xc}^{GGA}(\mathbf{r}) = V_{xc}\left[n(\mathbf{r}), \nabla n(\mathbf{r})\right]. \tag{3.17}$$

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) [9] and Perdew-Burke-Ernzerhof (PBE) [10].

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 [11]. The latter limitation is thought to be due to self-interaction in the Hartree potential in equation 3.14.

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 [12] introduced a 50% Hartree-Fock exact exchange and 50% LDA energy functional, while Perdew *et al.* [13] 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 [14], 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.* [15] 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 [16]. The functionals are expressed as

$$E_{xc}^{HSE} = aE_x^{HF,SR}(\omega) + (1-a)E_x^{PBE,SR}(\omega) + E_x^{PBE,LR}(\omega) + E_c^{PBE} \tag{3.18}$$

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.* [16] 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.

3.4 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 3.1.

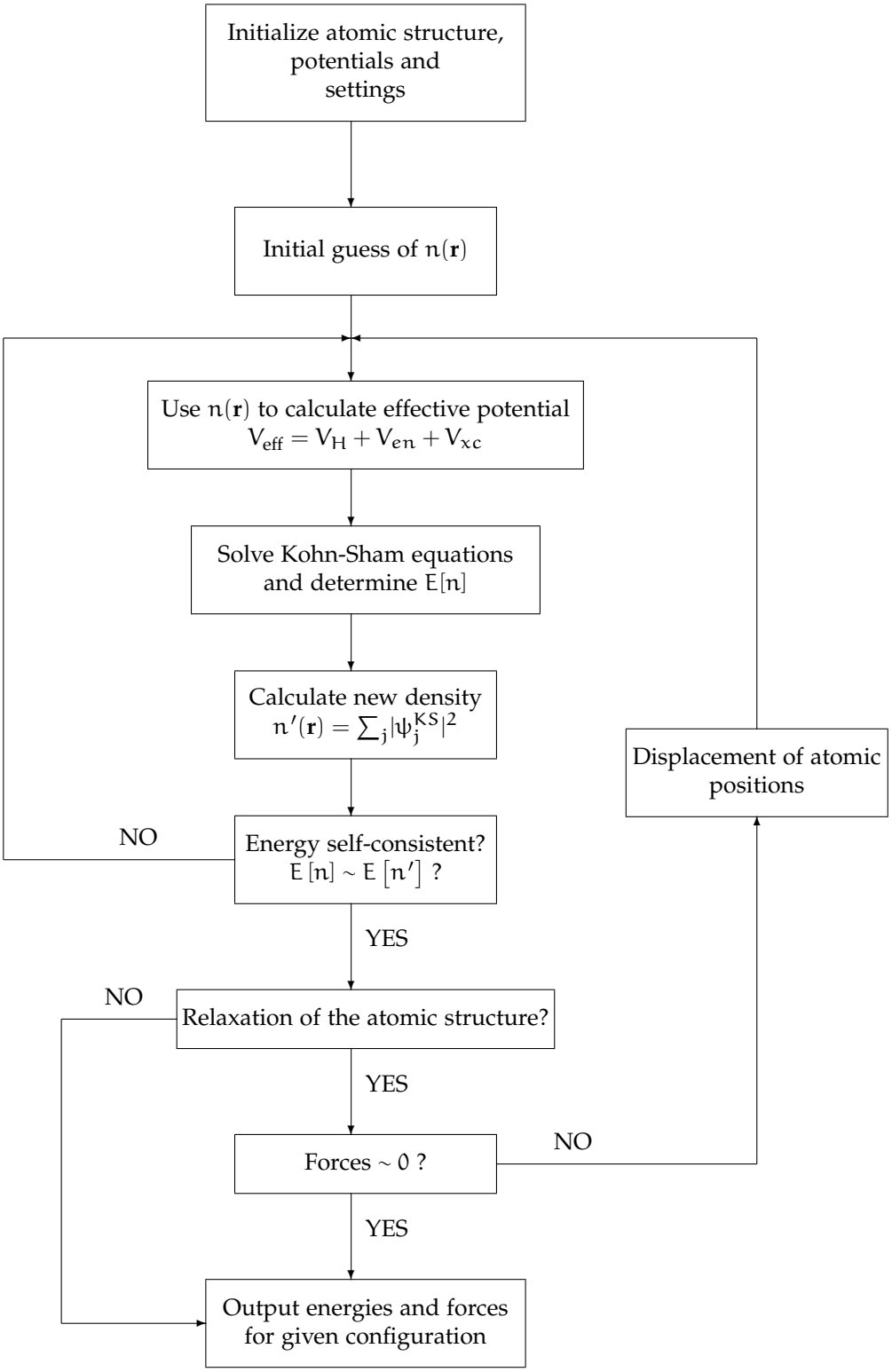

Figure 3.1: A flow chart of the self-consistent field method for DFT.

3.5 Limitations of the DFT

Chapter 4

Introduction to material science

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

This chapter will serve as an introduction to material science from a practical point of view, with a special emphasis on solid-state semiconductors.

4.1 Crystal structure

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 [17].

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 [18].

A third option is to have these atoms arranged with infinite periodicity, making the material a *crystalline solid* or more commonly named a *crystal*. [1]

The periodicity in a crystal is defined in terms of a symmetric array of points in space called the *lattice*, which can be simplified as either a one-dimensional array, a two-dimensional matrix or a three dimensional vector space, depending on the material. At each lattice point we can add an atom to make an arrangement called a *basis*. The basis can be one atom or a cluster of atoms having the same spatial arrangement. For every crystal, there exists periodically repeated building blocks called *cells* which represents the entire

[1] figur av de tre forskjellige strukturene? tikz plot?

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. [2]

4.2 Basics of semiconductors

Isolated atoms have distinct energy levels, where the Pauli exlusion principle [19] states for fermions that each energy level can accomodate two electrons of opposite spin. In a solid, the discrete energy levels of the isolated atom spread into continuous energy bands for the solid since the wavefunctions of the electrons in the neighboring atoms overlap [3]. Hence, an electron is not neccessarily localized at a particular atom anymore as it would for a system with an isolated atom. Every material has a unique band structure, similar to every human have their unique fingerprint.

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

To be able to accelerate electrons in a solid using an electrical field, they must be able to move into new energy states. At 0 K, the entire valence band of a semiconductor is full with electrons and no available new states, making it impossible to flow current through the material. This can be solved by using either thermal or optical energy to excite energies from the valence band to the conduction band, in order to *conduct* electricity. In room temperature, many semiconductors will be able to have excited electrons in the conducting band solely from thermal energy matching the energy band gap [17].

In some scenarios, thermal or optical energy is not sufficient for an excitation since the en-

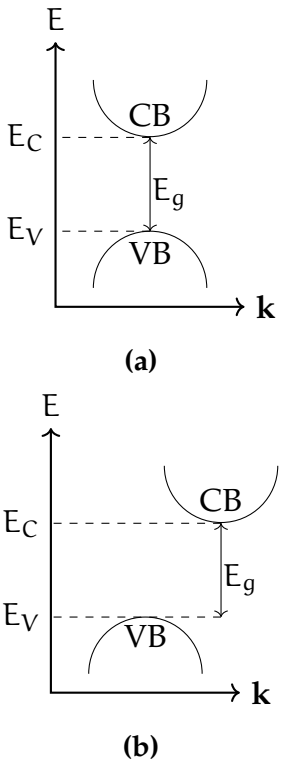

Figure 4.1: A schematic drawing of a (a) direct bandgap and an (b) indirect bandgap.

[2] Figur av enkleste gitter fcc, bcc og sc? tikz

[3] her er det mulighet for en kul figure fra diskret energi til continous band. Mena3000-boka

ergy bands are also dependent by a crystal momentum. A difference in the momentum of the minimal-energy state in the conduction band and the maximum-energy state in the valence band, is known to be an *indirect bandgap* as seen in figure 4.1a. If there is no difference at all, the material has a *direct bandgap*, which is visualized in figure 4.1a.

4.3 The perovskite structure

One known crystal structure is the perovskite structure. They have an ABO_3 stoichiometry whose symmetri belong to one of 15 space groups identified by Lufaso & Woodward [20], such as the cubic, orthorombic and tetragonal, to name a few. The A atom is nine- to 12-fold coordinated by oxygen, while the B atom is sixfold coordinated by oxygen, and the BO_6 octahedra are connected to the corners in all three directions.

The motivation behind the research of perovskites is reasoned by that there are a large amount of possible ABO_3 chemistries, whereas a significant portion of these that takes the perovskite structure. They have a broad specter of applications, ranging from high-temperature superconductors [21], ionic conductors [22], and multiferroic materials [23]. Additionally, adding a perovskite structured compound to solar cells has reportedly resulted in higher performance efficiency while being cheap to produce and simple to manufacture [24, 25], however, that includes the use of hybrid organic-inorganic compounds and excludes the use of oxygen.

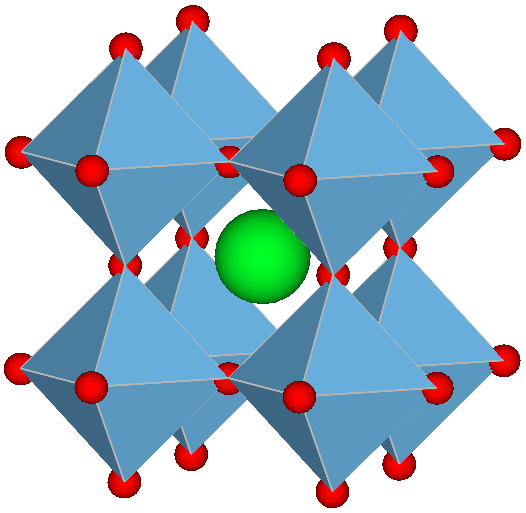

Figure 4.2: A crystal structure of $SrTiO_3$ which is a cubic perovskite. The red atoms are oxygen, whereas the green atom is strontium, and inside every corner-sharing BO_6 octahedral unit is a titanium atom.

Part III

Methodology and implementation

Chapter 5

Material Science Databases

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

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

To limit the quest of databases, we have restricted the search for databases and cloud services to include inorganic compounds obtained by first-principles calculations. Table 5.1 and 5.2 shows the databases and cloud services that meets the given criteries, respectively.

Database	API/REST	Free access	Number of compounds
AFLOW [26]	REST	True	3.27 M
MP [27]	MAPI [28]	True	0.66 M
OQMD [29, 30]	RESTful API (qmpy, matminer)	True	0.64 M
ICSD [31]	RESTful API	False	0.21 M
Jarvis-DFT [32]	API	True	0.04 M

Table 5.1: Databases of computational material science sorted after number of compounds. Abbreviations used are Novel Materials Discovery (NOMAD), Automatic-FLOW for Materials Discovery (AFLOW), Materials Project (MP), Inorganic Crystal Structure Database (ICSD) and Open Quantum Materials Database (OQMD).

Cloud service	API/REST	Free access
NoMaD [33]	API	True
CMR [34]	ASE	RESTful API
MatNavi	API	True
PRISMS	REST	True
Citrine	API	False
MPDS	API	False
MDF	API	False

Table 5.2: Cloud services that offers database-storage. Abbreviations used are Computational Materials Repository (CMR), NIMS Materials Database (MatNavi), PRedictive Integrated Structural Materials Science (PRISMS), Materials Platform for Data Science (MPDS) and the Materials Data Fascility (MDF).

5.1 Fundamentals of a database

5.1.1 modules

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

The Atomic Simulation Environment (ASE) is an environment in the Python programming language that includes several tools and modules for setting up, modifying and analyze atomistic simulations [35]. It is particularly used together with the cloud service Computational Materials Repository (CMR).

Another commonly used module is the Python Materials Genomics (pymatgen) [36]. This is a well-documented open module with both introductory and advanced use cases written in Jupyter Notebook for easy reproducibility, and is integrated with the Materials Project REST API.

The Materials Project are also behind a library named matminer [37], which is an open-source software platform written in Python. Matminer provides modules to extract data sets from many cloud-services and databases, with exampels in table 5.1 and 5.2, it can extract features from images (such as the band gap of a compound), and have modules for visualization of properties.

5.2 Databases and cloud services

5.2.1 Novel Materials Discovery

The Novel Materials Discovery (NOMAD) Repository is an open-access platform for sharing and utilizing computational materials science data. NOMAD also consists of several branches such as NOMAD Archieve, which is the representation of the NOMAD repository parsified into a code-independent format, NOMAD Encyclopedia, which is a graphical user interface (GUI) for characterizing materials, and lastly NOMAD Analytics Toolkit, which includes early-development examples of artificial-intelligence tools [33].

Databases that are a part of NOMAD data collection includes Materials Project, the Open Quantum Materials Database and AFLOW. They are all based on the underlying quantum engine Vienna ab initio simulation package (VASP) [38], which is a software based on DFT.

5.2.2 Materials project

Materials project is an open source project that offers a variety of properties of over one hundred thousand of inorganic crystalline materials. It is known

as the initiator of materials genomics and has as its mission to accelerate the discovery of new technological materials, with an emphasis on batteries and electrodes, through advanced scientific computic and innovative design [27].

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

Every compound has an initial relaxation of cell and lattice parameters performed using a 1000k-point mesh to ensure that all properties calculated are representative of the idealized unit cell for each respective crystal structure. The functional GGA is used to calculate band structures, while for transition metals it is applied $+U$ correction to correct for correlation effects in d- and f-orbital systems that are not addressed by GGA calculations [39]. The thermodynamic stability for each phase with respect to decomposition, is also calculated. This is denoted as E Above Hull, with a value of zero is defined as the most stable phase at a given composition, while larger positive values indicate increased instability.

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

5.2.3 Open Quantum Materials Database

The Open Quantum Materials Database (OQDM) is a completely free and available database of DFT-calculations [2]. It has included thermodynamic and structural properties of more than 600.000 materials, including all unique entries in the Inorganic Crystal Structure Database (ICSD) consisting of less than 34 atoms [29].

For general DFT-settings, see http://oqmd.org/documentation/vasp. (read a bit more about VASP before entering this labyrinth)

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

[2] download the entire database: http://www.oqmd.org/download/

5.2.4 AFLOW

The AFLOW repository uses the GGA-PBE functional within VASP with projector-augmented wavefunction (PAW) potentials to relax twice and optimize the ICSD-sourced structur. They are using a 3000-6000 k-point mesh, indicating a more computationally expensive calculation compared to the Materials Project.

5.2.5 JARVIS

Joint Automated Repository for Various Integrated Simulations (JARVIS) - DFT is an open database based based on the VASP software to perform a variety of material property calculations. It consists of roughly 40.000 3D and 1.000 2D materials using the vdW-DF-OptB88 van der Waals functional, which was originally designed to improve the approximation of properties of two-dimensional van der Waals materials, but has also shown to be effective for bulk materials [40, 41]. The functional has shown accurate predictions for lattice-parameters and energetics for both vDW and non-vdW bonded materials [42].

Structures included in the data set are originally taken from the materials project, and then re-optimized using the OPT-functional. Finally, the combination of the OPT and modified Becke-Johnson (mBJ) functionals are used to obtain a representative band gap of each structure, since both have shown unprecedented accuracy in the calculation of band gap compared to any other DFT-based calculation methods [43].

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

5.3

Bibliography

1. Persson, C. *Brief Introduction to the density functional theory* 2020.
2. Top500. *SUPERCOMPUTER FUGAKU* June 2020. https://www.top500.org/system/179807/ (2020).
3. Griffiths, D. *Introduction to quantum mechanics* ISBN: 9781107179868 (Cambridge University Press, Cambridge, 2017).
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 (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 (WILEY, Apr. 1, 2009).
9. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **45,** 13244–13249 (June 1992).
10. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **77,** 3865–3868 (Oct. 1996).
11. Freysoldt, C. *et al.* First-principles calculations for point defects in solids. *Reviews of Modern Physics* **86,** 253–305 (Mar. 2014).
12. Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics* **98,** 1372–1377 (Jan. 1993).

13. 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).

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

15. 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).

16. 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).

17. 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, Mar. 18, 2015).

18. 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.

19. Pauli, W. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik* **31,** 765–783 (Feb. 1925).

20. 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).

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

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

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

24. 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).

25. Chen, P.-Y. *et al.* Environmentally responsible fabrication of efficient perovskite solar cells from recycled car batteries. *Energy Environ. Sci.* **7,** 3659–3665 (2014).

26. 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).
27. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1,** 011002 (July 2013).
28. 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).
29. Kirklin, S. *et al.* The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **1.** doi:10.1038/npjcompumats.2015.10 (Dec. 2015).
30. 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).
31. Levin, I. *NIST Inorganic Crystal Structure Database (ICSD)* en. 2020. doi:10.18434/M32147.
32. Choudhary, K. *et al.* JARVIS: An Integrated Infrastructure for Data-driven Materials Design. arXiv: 2007.01831v1 [cond-mat.mtrl-sci] (July 3, 2020).
33. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *Journal of Physics: Materials* **2,** 036001 (May 2019).
34. Landis, D. D. *et al.* The Computational Materials Repository. *Computing in Science & Engineering* **14,** 51–57 (Nov. 2012).
35. 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).
36. 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).
37. Ward, L. *et al.* Matminer: An open source toolkit for materials data mining. *Computational Materials Science* **152,** 60–69 (Sept. 2018).
38. Kresse, G. & Furthmüller, J. Efficient iterative schemes forab initiototal-energy calculations using a plane-wave basis set. *Physical Review B* **54,** 11169–11186 (Oct. 1996).

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

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

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

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

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

44. Balachandran, P. V. *et al.* Predictions of new ABO3 perovskite compounds by combining machine learning and density functional theory. *Physical Review Materials* **2.** doi:10.1103/physrevmaterials.2.043802 (Apr. 2018).

45. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research (2011).* arXiv: 1201.0490v4 [cs.LG] (Jan. 2, 2012).

46. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2,** 18–22 (2002).

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

48. Pilania, G., Balachandran, P. V., Gubernatis, J. E. & Lookman, T. Classification ofABO3perovskite solids: a machine learning study. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* **71,** 507–513 (Sept. 2015).

49. Tanaka, I., Rajan, K. & Wolverton, C. Data-centric science for materials innovation. *MRS Bulletin* **43,** 659–663 (Sept. 2018).