

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

September 29, 2020

Abstract

This is an abstract. First coffee.

Acknowledgements

Acknowledgements. Coffe-time?

Contents

1	Introduction	1
I	Theory	3
2	Quantum mechanics	5
2.1	The single-electron Schrödinger equation	5
2.2	Eigenfunctions	7
2.3	Hartree-Fock approximations	8
3	Identifying materials	9
3.1	Crystallography	9
3.2	Semiconductors	10
II	Material Science Databases	11
4	Material Science Databases	13
4.1	Fundamentals of a database	15
4.1.1	modules	15
4.2	Novel Materials Discovery	15
4.2.1	Materials project	15
4.2.2	Open Quantum Materials Database	16
4.2.3	Joint Automated Repository for Various Integrated Sim- ulations	16
4.3	17

Contents

III	Implementation	19
IV	Results	21
V	Conclusion	23
VI	Appendices	25

Chapter 1

Introduction

This it The introduction. Another coffee.

Part I

Theory

Chapter 2

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.

2.1 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]

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V\Psi \quad (2.1)$$

for a convenient external potential $V_{\text{ext}}(r)$ that is independent of time. We will try to look for solutions for (2.1) by separating the wave function into a space-dependent and time-dependent function

$$\Psi(r, t) = \psi(r)\phi(t). \quad (2.2)$$

By inserting ordinary derivatives and dividing each side with equation (2.2), our Schrödinger equation (2.1) now reads

$$i\hbar \frac{1}{\phi(t)} \frac{d\phi(t)}{dt} = -\frac{\hbar^2}{2m} \frac{1}{\psi(r)} \nabla^2 \psi + V(r) \quad (2.3)$$

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;

$$i\hbar \frac{1}{\phi(t)} \frac{d\phi(t)}{dt} = E\phi(t) \quad (2.4)$$

and

$$\frac{\hbar^2}{2m} \frac{1}{\psi(r)} \nabla^2 \psi + V(r) = E\psi(r) \quad (2.5)$$

where the first equation (2.4) has a general solution $\phi(t) = C \exp(-iEt/\hbar)$ and $C = 1$ after normalization, and the second equation (2.5) is known as time-independent Schrödinger equation. These two equations are connected through the variable ϵ .

By utilizing variable separation to get equation (2.2), 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

$$\hat{H}(r, p) = \frac{p^2}{2m} + V(r) = -\frac{\hbar^2}{2m} \nabla^2 + V(r) \quad (2.6)$$

simplifying equation 2.5 to

$$\hat{H}\psi = E\psi \quad (2.7)$$

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 .

2.2 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

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

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

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

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)

$$\Psi(r, t) = \sum_{\kappa} c_{\kappa} \psi_{\kappa}(r, t), \quad (2.8)$$

where c_{κ} 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 2.7 into time-independent Schrödinger equation for eigenfunctions

$$\hat{H}\psi_{\kappa}(r) = \varepsilon_{\kappa}\psi_{\kappa}(r). \quad (2.9)$$

The shape of an eigenfunction has normally high spatial symmetry that depends on the symmetry of the potential $V_{\text{ext}}(r)$ 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.

2.3 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

$$\left(-\frac{\hbar^2 \nabla_1^2}{2m_e} - \frac{\hbar^2 \nabla_2^2}{2m_e} + \frac{q^2}{|r_1 - r_2|} + V_{\text{ext}}(r) \right) \Psi_{\kappa}(r_1, r_2) = E_{\kappa} \Psi_{\kappa}(r_1, r_2), \quad (2.10)$$

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

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

$$\Psi(r_1, r_2) = A \cdot \psi_1(r_1) \psi_2(r_2). \quad (2.11)$$

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;

$$\Psi(r_1, r_2) = \frac{1}{\sqrt{2}} \left(\psi_1(r_1) \psi_2(r_2) - \psi_1(r_2) \psi_2(r_1) \right) \quad (2.12)$$

Chapter 3

Identifying materials

3.1 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 ¹.

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 [3].

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

¹Need source

3.2 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

Chapter 4

Material Science Databases

There are multiple different databases for material science available for every day use, some of them completely open-source while others commercial. In this chapter we will investigate the different databases and what their scope of speciality is.

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

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

Database	API/REST	Free access	Number of compounds
AFLOW [4]	REST	True	3.27 M
MP [5]	MAPI [6]	True	0.66 M
OQMD [7, 8]	RESTful API (qmpy, matminer)	True	0.64 M
Jarvis-DFT [9]	API	True	0.04 M

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

Cloud service	API/REST	Free access
NoMaD [10]	API	True
CMR [11]	ASE	RESTful API
MatNavi	API	True
PRISMS	REST	True
Citrine	API	False
MPDS	API	False
MDF	API	False

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

4.1 Fundamentals of a database

4.1.1 modules

Many of the databases share a common and convenient factor that happens to ease the life of their users. The modules used to adapt, visualize, calculate or predict have become established in their field of speciality.

The Atomic Simulation Environment (ASE) is an environment in the Python programming language that includes several tools and modules for setting up, modifying and analyzing atomistic simulations [12]. It is particularly used together with the cloud service Computational Materials Repository (CMR).

Another commonly used module is the Python Materials Genomics (pymatgen) [13]. This is a well-documented open module with both introductory and advanced use cases written in Jupyter Notebook for easy reproducibility and offline-scenarios. It is integrated with the Materials Project REST API (MAPI) by using a wrapper.

4.2 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 Archive, 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 [10].

Databases that are a part of NOMAD data collection include Materials Project, the Open Quantum Materials Database and AFLOW which are all based on the underlying quantum engine VASP ¹.

4.2.1 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 computing and innovative design².

It is built upon over 60 features³, some features being irrelevant for some

¹<https://www.vasp.at>

²add link MP here

³All features can be viewed in the documentation of the project: github.com/materialsproject/mapidoc/master/materials

materials while fundamental for others. Almost all of the data in the project is calculated using a theoretical technique called Density Functional Theory. 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. Some features, eg. density of state (dos), bandstructure, Pourbaix diagram etc., are characterized by using the module pymatgen.

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

4.2.2 Open Quantum Materials Database

The Open Quantum Materials Database (OQDM) is a completely free and available database of DFT-calculations ⁴. 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 [7].

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

4.2.3 Joint Automated Repository for Various Integrated Simulations

Joint Automated Repository for Various Integrated Simulations (JARVIS) - DFT is an open database based solely on DFT-calculations. It consists of roughly 40.000 3D and 1.000 2D materials using the vdW-DF-OptB88 van der Waals functional. This functional has shown accurate predictions for lattice-parameters and energetics for both vdW and non-vdW bonded materials [14]. There does also exist more data including 1D-nanowire and 0D-molecular materials, yet not publically distributed.

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.

⁴download the entire database: <http://www.oqmd.org/download/>

4.3

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. 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).
4. 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).
5. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002 (July 2013).
6. 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).
7. 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).
8. 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).
9. 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](https://arxiv.org/archive/cond-mat)] (July 3, 2020).
10. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *Journal of Physics: Materials* **2**, 036001 (May 2019).
11. Landis, D. D. *et al.* The Computational Materials Repository. *Computing in Science & Engineering* **14**, 51–57 (Nov. 2012).

12. 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).
13. 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).
14. 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).
15. 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](https://doi.org/10.1103/physrevmaterials.2.043802) (Apr. 2018).
16. 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).
17. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2**, 18–22 (2002).
18. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29**, 1189–1232 (Oct. 2001).
19. Pilania, G., Balachandran, P. V., Gubernatis, J. E. & Lookman, T. Classification of ABO₃ perovskite solids: a machine learning study. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* **71**, 507–513 (Sept. 2015).
20. Ward, L. *et al.* Matminer: An open source toolkit for materials data mining. *Computational Materials Science* **152**, 60–69 (Sept. 2018).
21. Levin, I. NIST Inorganic Crystal Structure Database (ICSD) en. 2020. doi:[10.18434/M32147](https://doi.org/10.18434/M32147).
22. Tanaka, I., Rajan, K. & Wolverton, C. Data-centric science for materials innovation. *MRS Bulletin* **43**, 659–663 (Sept. 2018).