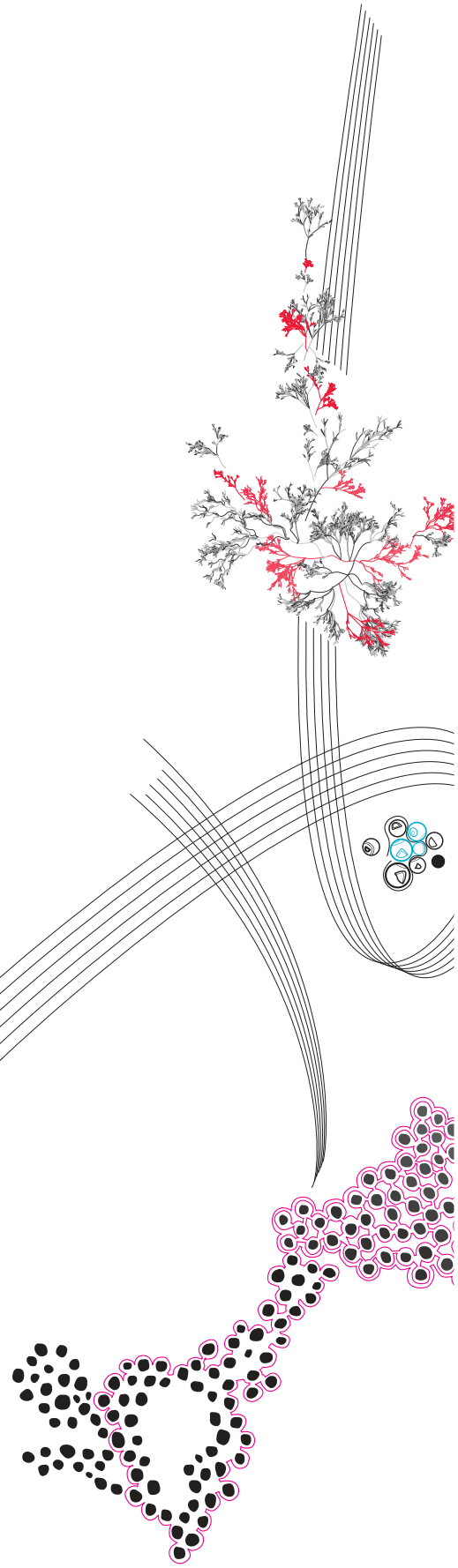

BSc Thesis
Applied Physics & Applied Mathematics

Excited State Forces in Semiconductor Crystals from the GW and Bethe-Salpeter Equation Methods

Alwin de Vries

Supervisors: L. Leppert, F.P. Schuller, P. Lechifflart

July, 2024

Department of Applied Physics
Faculty of Science and Technology

UNIVERSITY OF TWENTE.

Preface

This project was done as my Bsc graduation project for Applied Physics and Applied Mathematics. The work was performed in the CCP group of the TNW faculty of the University of Twente.

First of all, I would like to thank Prof. Dr. Linn Leppert, the chair of my committee. She was the first person I spoke to about this project and the one who originally conceived of it, so the project would not have been possible without her assistance. I would also like to express my gratitude for her guidance on choosing a supervisor from Applied Mathematics, as well as the great feedback she gave each time we met.

Now I want to express my gratitude towards Dr. Pierre Lechifflart, my daily supervisor throughout the project. I thoroughly appreciate all of his ideas about each of the directions we could take the project into; from the start, he constantly had plans for what should be the next step of the project. I am also very thankful for the way in which he explained the relevant theory to me and helped answer any questions I had throughout the project, whether he already knew the answer or not.

Next, I would like to thank Prof. Dr. Frederic P. Schuller. For this project, he was my supervisor from Applied Mathematics. Frederic gave very helpful advice on how to integrate a mathematical aspect into the project and gave very thorough feedback on the theory section of this report, which I really appreciated.

Finally, I am grateful towards all the members of the CCP group, for being ready to answer questions and help me out whenever I needed it. I especially appreciated the feedback and questions whenever I presented my progress within the group.

Excited State Forces in Semiconductor Crystals from the GW and Bethe-Salpeter Equation Methods

Alwin de Vries*

July, 2024

Abstract

Research into the forces caused by photo-induced excitation of semiconductors is currently very computationally costly. This report aims to set a benchmark of forces in silicon to which results of future analytical methods can be compared. Silicon is studied using first-principles numerical modelling techniques, namely density functional theory as implemented in QUANTUM ESPRESSO and the GW and Bethe-Salpeter Equation approach as implemented in the BERKELEYGW code. The lattice parameter of silicon was calculated to be 5.471 Å, which is in good agreement with experimental reference values. We then mapped out the excited-state energy landscape of silicon by calculating excited-state energies as a function of the interatomic distance of the two atoms in the primitive unit cell of fcc silicon. In these calculations, we observed exciton energies far above experimental values and no minimum energy was found. We therefore used a supercell approach, which allows us to calculate the energy of the indirect exciton. We found that even after relaxing the lattice parameter in the first excited state to 5.45 Å, there were still forces on the atoms, implying that some reconfiguration of the crystal also occurs during excitation.

Keywords: excited state forces; density functional theory; GW and Bethe-Salpeter equation methods

*Email: a.devries-1@student.utwente.nl

Contents

1 Introduction

For many electronics and solar energy applications, studying the fundamental electronic properties of semiconductor materials is vital, as it can lead to increased device efficiencies and allows for tailored implementation of specific functionalities. Some of the most important properties of these semiconductors are related to the changes that occur within them when exposed to light.

When interacting with a photon, an electron in the semiconductor crystal can be promoted from a valence state to a conduction state, this in turn leaves behind a (positively charged) hole in the valence state. Because of the attractive Coulomb force between them, the electron-hole pair forms a correlated quasiparticle called exciton.

These excitons are very interesting to study as they greatly influence the absorption spectrum of the materials they occur in, usually creating new states that are lower than the conduction band minimum of the material. When an exciton is created, it can also change the configuration of the semiconductor crystal. In some cases this can result in a self-trapped exciton, where the exciton causes a local distortion of the crystal, leading to it no longer being able to move around freely. Self-trapped excitons once again have a large impact on the photophysical properties of the materials they occur in.

The motivation for this project is that these excitations cause forces between the atoms in a crystal that can affect the structure and optical properties. Calculating these excited-state forces is currently very computationally costly, and state-of-the-art analytical methods rely on severe approximations. Computational methods and approximations for increasing the efficiency of these calculations have been created and tested for small molecules, but it is unknown to which extent the necessary approximations are accurate for semiconductors [11]. This project aims to set a benchmark for the excited-state forces in silicon, such that the accuracy of future implementations of these analytical methods can be evaluated.

2 Theory

This section is devoted to outline the theory and some of the history of Density Functional Theory and a short introduction to the concepts employed in GW+BSE calculations. In this project, all the ground-state calculations were done through the use of density functional theory, specifically using the Quantum ESPRESSO code [6, 5, 7].
All the calculations in the excited state were performed using GW+BSE methods, specifically with BerkeleyGW [15, 4, 10].

2.1 Density Functional Theory

Density Functional Theory (DFT) is a computational method used to determine the energy and other properties of the ground state of molecules and crystals. Unless stated otherwise, the theory in this section was adapted from the book *Electronic Structure: basic theory and practical methods* by Richard M. Martin [13].

The total energy of systems such as these, can be written as a sum of the Coulomb interactions between nuclei, nuclei and electrons and electron-electron interactions. Using the Born-Oppenheimer approximation, which allows for separating electronic and nuclear degrees of freedom, we assume the nuclei to be in fixed positions and classical objects, meaning that the Schrödinger equation of the nuclei is not solved and the interaction between nuclei contributes a constant energy to the total energy of the system. The nucleus-electron interactions can be modelled in terms of an external potential $V_{ext}(\mathbf{r})$. This leaves the electron-electron interactions as the largest challenge.

Instead of simulating individual electrons, DFT merely considers an overall electron density, since this scales to large systems far more efficiently. For a system of N particles, traditional approaches would have a many-body wavefunction with $3N$ spatial coordinates, DFT only requires the electron density with 3 spatial coordinates regardless of the amount of particles. The foundation that proved DFT could be a viable approach was created by Hohenberg and Kohn [9] in 1964, in the form of two theorems and their corollaries.

Theorem 1 *For any system of interacting particles in an external potential, the external potential $V_{ext}(\boldsymbol{r})$ is determined uniquely, except for an additive constant, by the ground-state electron density $n_0(\boldsymbol{r})$.*

Corollary 1.1 *All observables of the system are completely determined by the ground-state density $n_0(\boldsymbol{r})$.*

This corollary is true, because the external potential determines the many-body wavefunction, which in turn determines the ground-state electron density.

Theorem 2 *A universal functional for the energy $E[n]$ in terms of the electron density $n(\boldsymbol{r})$ can be defined, valid for any external potential $V_{ext}(\boldsymbol{r})$.*
For any particular $V_{ext}(\boldsymbol{r})$, the global minimum of this functional is the ground-state energy, and it is achieved at the ground-state density $n_0(\boldsymbol{r})$.

Corollary 2.1 *The functional $E[n]$ alone is sufficient to determine the ground-state energy and density.*

The first corollary is a very strong statement, because it says that all the information of a many-body wavefunction of $3N$ dimensions is encoded in a 3-dimensional density.

The second theorem shows that the ground-state electron density can be determined by minimising the energy with respect to the density. The one shortcoming of these theorems is that they do not tell one how the energy functional should be constructed, which in practice makes it impossible to perform the minimisation.

Because of the non-constructiveness of these results, the thus emerging idea lay dormant until the creation of the Kohn-Sham Auxiliary System [13], which is based on the following statement.

Theorem 3 *The exact ground-state density can be represented by an auxiliary system of non-interacting particles.*

Within the Kohn-Sham approach, the energy functional is written as follows:

$$E[n] = E_{KS}[n] = T_s[n] + E_{Hartree}[n] + E_{II} + E_{xc}[n] + \int d\boldsymbol{r} V_{ext}(\boldsymbol{r}) n(\boldsymbol{r}). \tag{1}$$

Here the Kohn-Sham total energy is divided into five terms.
The first term is the non-interacting kinetic energy $T_s[n]$. This is followed by the Hartree potential $E_{Hartree}[n]$, which is the potential energy from electrostatic interaction resulting from the electron density. E_{II} is the energy from interactions between nuclei, and the value of the integral is equal to the energy from electrostatic reactions between the nuclei and electrons. Finally, $E_{xc}[n]$ is the exchange-correlation energy which is the combination of the remaining types of energy.

It is worth noting, that although $T_s[n]$ is written as a functional of the density, in practice it is calculated as a functional of single-particle orbitals. Writing it this way is allowed, because the Hohenberg Kohn theorems make it clear that these orbitals are also functionals of the density.

Next, a derivation of the Kohn-Sham equations will be presented [3].
Seeing that the ground-state density should minimise the energy, one can write

$$\frac{\delta E_{KS}[n]}{\delta n(\boldsymbol{r})} = \frac{\delta T[n]}{\delta n(\boldsymbol{r})} + \frac{\delta E_{Hartree}[n]}{\delta n(\boldsymbol{r})} + \frac{\delta E_{xc}[n]}{\delta n(\boldsymbol{r})} + \frac{\delta}{\delta n(\boldsymbol{r})} \left(\int d\boldsymbol{r} V_{ext}(\boldsymbol{r}) n(\boldsymbol{r}) \right) = 0. \tag{2}$$

This can then be written as

$$\frac{\delta E_{KS}[n]}{\delta n(\boldsymbol{r})} = \frac{\delta T[n]}{\delta n(\boldsymbol{r})} + V_{Hartree}(\boldsymbol{r}) + V_{xc}(\boldsymbol{r}) + V_{ext}(\boldsymbol{r}) = 0 \tag{3}$$

Now consider an auxiliary system of non-interacting particles moving in a potential v_s, then for the minimisation of its energy one has that

$$\frac{\delta E[n]}{\delta n(\boldsymbol{r})} = \frac{\delta T_s[n]}{\delta n(\boldsymbol{r})} + \frac{\delta V_s[n]}{\delta n(\boldsymbol{r})} = \frac{\delta T_s[n]}{\delta n(\boldsymbol{r})} + v_s = 0. \tag{4}$$

As interactions do not impact the kinetic energy term, the last two equations are exactly equal when

$$v_s = V_{ext}(\boldsymbol{r}) + V_{Hartree}(\boldsymbol{r}) + V_{xc}(\boldsymbol{r}). \tag{5}$$

This means the two equations are also solved by the same electron density. The density can then be found from the non-interacting system by solving

$$H_{KS}\psi_i(\boldsymbol{r}) = \epsilon_i\psi_i(\boldsymbol{r}), \tag{6}$$

where

$$H_{KS} = -\frac{1}{2}\nabla^2 + v_s(\boldsymbol{r}). \tag{7}$$

H_{KS} is the Hamiltonian operator for this Schrödinger equation. It consists of a kinetic energy term and a potential. ψ_i and ϵ_i are eigenfunctions of the Hamiltonian (also called Kohn-Sham orbitals) and their eigenvalues respectively. For these equations, the convention $\hbar = m = e = 1$ is used.

Once one has solved Equation 6 with Equations 7 and 5, the resulting electron density can be computed with

$$n(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2. \tag{8}$$

These last four equations together are known as the Kohn-Sham equations, and they need to be solved self-consistently, such that the resulting electron density is equal to the density that was used to calculate the potentials.

Methods for numerically solving these Schrödinger equations and computing most of these terms accurately are widespread, but the one term that causes problems is the final one: The exchange-correlation potential V_{xc}.

Many different approximations are used to calculate the exchange-correlation energy. The potentials are then obtained as the functional derivative with respect to density [13]. Two such approximations were used in this project, the first of these is the Local Density Approximation (LDA) which considers solids to be close to homogeneous electron gasses. The formula for the exchange-correlation energy is given by

$$E_{xc}[n] = \int d^3r\ n(\boldsymbol{r})\epsilon(n(\boldsymbol{r})), \tag{9}$$

where ϵ is the exchange-correlation energy of a single particle in a homogeneous electron gas, with density n. These energies are partially computable directly and partially need to be approximated through other means that will not be covered in this article.

The other class of approximation used in this project are the Generalised Gradient Approximations (GGAs). This approximation effectively works on the same basis as the LDA, considering solids to be close to homogeneous electron gasses. What differentiates GGAs is that they introduce gradients of the electron density, resulting in the functional for the exchange-correlation energy being written as

$$E_{xc}[n] = \int d^3r\ n(\boldsymbol{r})\epsilon_x^{hom}(n(\boldsymbol{r}))F_{xc}(n, |\boldsymbol{\nabla} n|). \tag{10}$$

The F_{xc} term here is a dimensionless correction including gradients of the density and ϵ_x^{hom} is the exchange energy of an unpolarised electron gas. The expression for the exchange enhancement factor F_x makes use of a dimensionless nth-order gradient which is given by

$$s_m = \frac{|\boldsymbol{\nabla}^m n|}{2^m(3\pi^2)^{m/3}(n)^{(1+m/3)}} \tag{11}$$

Finally, the first few orders of the Taylor expansion of F_x have been analytically found to be

$$F_x = 1 + \frac{10}{81}s_1^2 + \frac{146}{2025}s_2^2, \tag{12}$$

but many other forms are used, such as

$$F_x(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s_1^2/\kappa} \tag{13}$$

where μ and κ are both dimensionless constants [2].
The contribution of correlation, on the other hand, is quite complex and usually calculated using other numerical methods. This type of approximation always results in a higher exchange-correlation energy, which means that binding energies resulting from calculations using GGA are usually lower than those obtained when using LDA.

All of this gives rise to a general workflow for DFT calculations:

1. Determine some starting estimate for the electron density $n(\boldsymbol{r})$.
2. Using the estimate $n(\boldsymbol{r})$, calculate the Kohn-Sham potential.
3. Solve Equation 6 using Equations 7 and 5.
4. Compute the new electron density using the obtained eigenfunctions.
5. Optionally, mix the old and new electron densities.
6. If the resulting density satisfies the convergence conditions, move on to step 7. Otherwise, return to step 2 with the new density.
7. Determine desired quantities, such as total energy.

There are a few more areas of importance to consider. The first one is the way in which the energy for a new cycle is determined. From previous results, it turns out that it is optimal to not simply take the resulting density, but to instead mix the new and old densities together linearly, with some mixing constant α. This massively increases the stability with which the energy converges to the correct value.

By virtue of the Hohenberg-Kohn theorem, if the equations converge, they will always converge to the ground-state. It is not guaranteed that the equations will converge in the first place, so it is valuable to understand what underlying principle leads to the convergence.
The most important part of this is in the approximation for the exchange-correlation potential. This potential strongly depends on the electron density, so as the electron density becomes more accurate, V_{xc} also converges. In most cases, the exact potential is not available. Instead, one employs approximations, such as LDA or GGAs, which provide the energy as a functional of the electron density. The functional $E_{xc}[n]$ does not change between iterations, but its value and functional derivative do converge with n.

2.2 Numerical Implementation of DFT

In practical calculations for solids, translational symmetry can be used in order to express the Kohn-Sham equations in reciprocal space using a plane wave basis. Usually, the orbitals are split into two segments: a localised part close to the nuclei of the atoms, as well as a smooth part for the connecting sections. This smooth part can be approximated through the use of plane waves very well.

The first important piece of information is that the electron density $n(\boldsymbol{r})$ should be periodic, as the crystal itself is periodic. The density is obtained through the use of Equation (8), which implies that a periodic electron density should also lead to a periodic magnitude of the wavefunctions. This gives some intuition for Bloch's theorem:

Theorem 4 ***Bloch's Theorem:***
In a crystal with lattice vectors $\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}$, any valid wavefunction satisfies

$$\psi(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u(\boldsymbol{r}). \tag{14}$$

Here $u(\boldsymbol{r})$ is a function that satisfies $u(\boldsymbol{r}) = u(\boldsymbol{r} + \boldsymbol{L})$, with $\boldsymbol{L} = h\boldsymbol{a_1} + k\boldsymbol{a_2} + l\boldsymbol{a_3}$ for any integers h, k, l. $\boldsymbol{k}$ is the crystal momentum.

Any value for $\boldsymbol{k}$ within the first Brillouin zone will give a different wavefunction, so in theory one would have to integrate the wavefunction over every value of $\boldsymbol{k}$ in order to obtain the electron density. Instead, a summation over finitely many k-points is often used. In DFT calculations, it is important to first make sure that sufficient k-points are used for the densities to converge.

Because $u_k(\boldsymbol{r})$ is a periodic function, it can be expressed in terms of a Fourier series. The only spatial frequencies that can contribute here are those that conserve the periodicity requirements set by the crystal, these are the reciprocal lattice vectors of the crystal.

Higher spacial frequencies are related to fast changes in the wavefunction, but as mentioned at the beginning of this section, the plane waves are mainly used for the smooth parts. Because of this, a truncation of the spatial frequencies is used; any frequency related to a kinetic energy larger than some cut-off energy E_{cut} is not used in the basis. Specifically,

$$\frac{1}{2}|\boldsymbol{k} + \boldsymbol{G}|^2 < E_{cut}, \tag{15}$$

where $\boldsymbol{G}$ is a reciprocal lattice vector.

With this plane wave basis, it is convenient to write the Kohn-Sham equations in their reciprocal form [13]

$$\frac{1}{2}\left|\boldsymbol{k} + \boldsymbol{G}\right|^2 C_{i,\boldsymbol{k}+\boldsymbol{G}} + \sum_{\boldsymbol{G'}} v_s(\boldsymbol{G} - \boldsymbol{G'}) C_{i,\boldsymbol{k}+\boldsymbol{G'}} = \epsilon_{i\boldsymbol{k}} C_{i,\boldsymbol{k}+\boldsymbol{G}}. \tag{16}$$

In this form, ϵ_i is the same as in equation 6. $\mathbf{G}$ and $\mathbf{G}'$ are the reciprocal lattice vectors included in the plane wave basis, one of these equations is obtained for every $\boldsymbol{G}$. The coefficients v_s are the coefficients of the Fourier series of the potential and $C_{i,\mathbf{k}+\mathbf{G}}$ are the coefficients of the plane waves that build up the orbital ψ_i. This turns the problem into a set of linear equations with equally many unknown coefficients, for which solution methods are widespread.

2.3 Hellmann-Feynman Theorem

In order to determine forces in the approach suggested by Ismail-Beigi and Louie [11], one theorem known as the Hellmann-Feynman theorem is essential. The theorem states the following:

Theorem 5 ***Hellmann-Feynman Theorem***
For a hermitian operator $\hat{H}_\lambda$ with eigenvalue E_λ and eigenfunction ψ_i, all of which are dependent on a parameter λ, the following holds:

$$\frac{dE_\lambda}{d\lambda} = \langle\psi_\lambda|\frac{d\hat{H}_\lambda}{d\lambda}|\psi_\lambda\rangle \tag{17}$$

Provided that these derivatives exist.

This theorem is very convenient for the calculation of forces, for example, if the parameter λ is x, denoting one of the coordinates of an atom, then:

$$F_x = -\frac{\partial E_x}{\partial x} = -\langle\psi_x|\frac{\partial\hat{H}_x}{\partial x}|\psi_x\rangle \tag{18}$$

The original proof by Feynman starts from the fact that E_λ is the expectation value of $\hat{H}_\lambda$, thus

$$E_\lambda = \int \psi_\lambda^* \hat{H}_\lambda \psi_\lambda dv. \tag{19}$$

From this expression, one takes the derivative with respect to λ, assuming that each part is differentiable, the integral is then split into three parts as

$$\frac{dE_\lambda}{d\lambda} = \int \frac{d\psi_\lambda^*}{d\lambda}\hat{H}_\lambda\psi_\lambda dv + \int \psi_\lambda^*\frac{d\hat{H}_\lambda}{d\lambda}\psi_\lambda dv + \int \psi_\lambda^*\hat{H}_\lambda\frac{d\psi_\lambda}{d\lambda}dv. \tag{20}$$

Now, because is hermitian, $\hat{H}_\lambda\psi_\lambda = E_\lambda\psi_\lambda$ and $\hat{H}_\lambda\psi_\lambda^* = E_\lambda\psi_\lambda^*$, the first and third terms can be combined as

$$\frac{dE_\lambda}{d\lambda} = E_\lambda\int \frac{d\psi_\lambda^*}{d\lambda}\psi_\lambda dv + \int \psi_\lambda^*\frac{d\hat{H}_\lambda}{d\lambda}\psi_\lambda dv + E_\lambda\int \psi_\lambda^*\frac{d\psi_\lambda}{d\lambda}dv. \tag{21}$$

$$= \int \psi_\lambda^*\frac{d\hat{H}_\lambda}{d\lambda}\psi_\lambda dv + E_\lambda\frac{d}{d\lambda}\int \psi_\lambda^*\psi_\lambda dv. \tag{22}$$

As the wavefunction is assumed to be normalised, the last term is equal to zero. This leads to the desired expression of

$$\frac{dE_\lambda}{d\lambda} = \int \psi_\lambda^*\frac{d\hat{H}_\lambda}{d\lambda}\psi_\lambda dv \tag{23}$$

which is equivalent to the form given in the theorem. It would be desirable to apply this to DFT, but that is not possible yet. In numerical implementations of DFT, the Kohn-Sham orbitals are approximated with a finite-dimensional basis set. It turns out that the Hellmann-Feynman theorem is a direct consequence of the Rayleigh-Ritz Variational principle. Because of this principle, the theorem also works with orbitals approximated using variational principles, including those obtained from Density Functional Theory.

2.4 Representation of Energy Surface

In this project, it is important to be able to visualise how the total energy relates to the displacement of the atoms. The total energy is a function of all $3N$ coordinates of the atoms within the system, so it is preferable to find a way in which any displacement can be reduced to one dimension.

For a displacement along one axis to be representative of a displacement along any axis, two requirements need to be met.

The first of these requirements is that approximating the energy as a quadratic function of the displacement should be accurate. This is not generally true, unless very small displacements with respect to an equilibrium are used.
The second requirement is that the forces caused by a displacement should be directly opposed to the displacement. This requirement is physically intuitive, as it implies that the crystal will always try to return directly to its equilibrium configuration. If this requirement holds, the force can be written as

$$\mathbf{F} = -k(\mathbf{D})\mathbf{D}, \tag{24}$$

for some positive scalar function $k(\mathbf{D})$. Here $\mathbf{F}$ is the force and $\mathbf{D}$ is the displacement with respect to the equilibrium configuration. The force along one axis should also be equal to $\frac{\partial E}{\partial D}$, where D is the displacement along that one axis. Following the first assumption, E is a quadratic function of the displacement, so the derivative should be linear. This implies that while the first requirement holds, $k(\mathbf{D})$ is (roughly) constant. So the force can simply be written as

$$\mathbf{F} = -k\mathbf{D}. \tag{25}$$

2.5 Basics of GW+BSE Methods

DFT is very accurate for calculations of ground-state properties, but within the Kohn-Sham framework it is not suitable for excited-state calculations. We therefore use two approaches based on Green's function-based many-body perturbation theory, called the GW and Bethe-Salpeter Equation approaches.

The GW approach acts as a correction to DFT, which adds the neglected electron-electron interactions back in. In "GW", G is the one-particle Green's function, which is the main mathematical object used in the GW approach. G describes the evolution in time and space of a quasi-electron dressed with the Coulomb interaction. Being dressed with the Coulomb interaction means that it is surrounded by a slightly positively charged area, as it is less likely to find an electron close to another.

G is a complex function, and the poles of G give the energy levels of the quasi-electrons, which give the electronic states of the crystal. Applying this correction gives far more accurate band structures, because the electron-electron interactions are now treated far more accurately.

The GW approach gives a good description of charge excitation for systems that gain or lose one electron compared to the equilibrium state, but it does not describe neutral excitations where the electron stays in the crystal. For this, the Bethe-Salpeter Equation

(BSE) is needed.

For the BSE approach, instead of the one-particle Green's function, the two-particle Green's function L is used to describe the evolution of an electron-hole pair. It can be proven that L should satisfy the equation

$$L = L_0 + L_0 K L, \tag{26}$$

where L_0 is the non-interacting two particle Green's function and K is the Bethe-Salpeter kernel. This can then be written as

$$L = L_0 + L_0(v - W)L, \tag{27}$$

here v is the repulsive electron-hole interaction which is an interaction between dipoles. W is the direct (attractive) Coulomb interaction of the electron-hole pair. W is usually frequency-dependent, but this is usually approximated out ($\omega = 0$). Because of this approximation, the equation can be inverted, which gives a large excitonic Hamiltonian. This Hamiltonian has eigenvectors A^{vc} and eigenvalues E^{vc}, where v and c denote the valence and conduction states respectively. This means

$$H_{exc}^{vc,v'c'} A^{vc} = E^{vc} A^{vc}. \tag{28}$$

The lowest eigenvalue here represents the energy of the smallest direct exciton, which is the value that is usually computed in this project. The Hamiltonian can be written as

$$H_{exc}^{vc,v'c'} = (E_c - E_v)\delta_{vv'}\delta_{cc'} - I, \tag{29}$$

where δ is the Kronecker delta and I is an interaction term, this term is usually written in greater detail, but this is not relevant for this explanation. The first term here represents the electric band gap energy, and the second term is the (usually negative) binding energy of the electron-hole pair. This second term is the reason that the energy of an exciton is lower than that of the conduction state. Since these calculations are technically highly demanding, they were carried out all either by dr. Pierre Lechifflart or using the workflow he created for this purpose. All of the excited-state energies and forces referred to in the following sections rely on these calculations.

3 Method & Results

3.1 Convergence Parameters

The first step in this project was to find appropriate parameters such that the calculations converge to their numerically converged values. Convergence parameters include the energy cut-off of the plane-wave expansion and the number of k-points used to sample reciprocal space.
The choice was made to use 6 k-points in each direction, the convergence can be seen in Figures 1 and 2. This choice resulted in $6^3 = 216$ total k-points, as well as using an energy cut-off of 680 eV, this seemed to ensure that the resulting total energy values were precise up to 4 significant figures (when written in Ry).

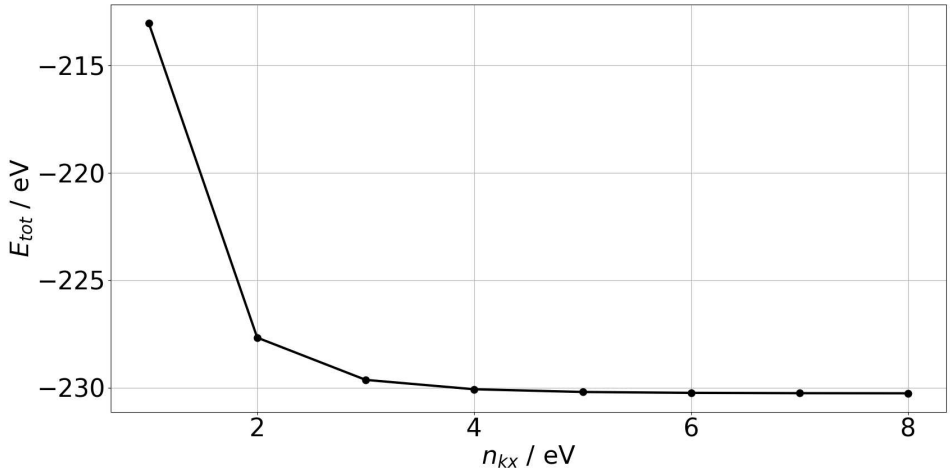

FIGURE 1: Plot of total energy against number of k-points on one axis. $n_{kx} = n_{ky} = n_{kz}$ and $E_{cut} = 680$ eV

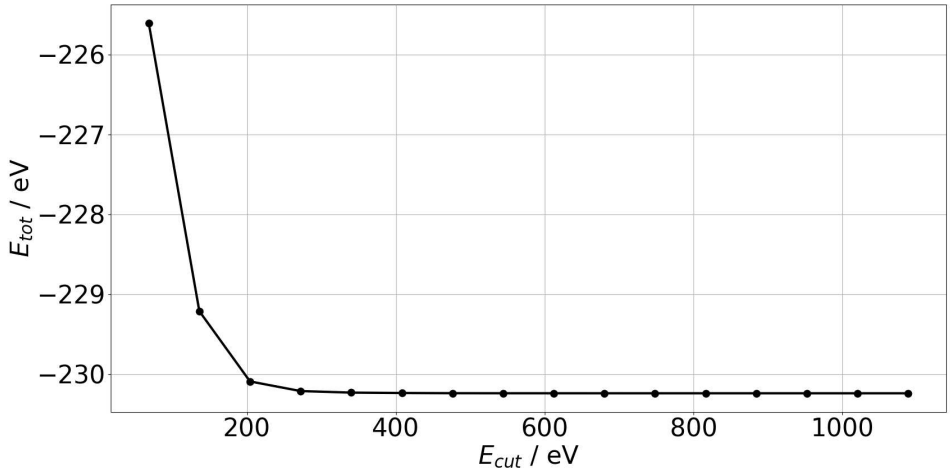

FIGURE 2: Plot of total energy against cut-off energy with $n_{kx} = n_{ky} = n_{kz} = 6$

3.2 Lattice Parameter of Silicon

Silicon crystallizes in a face-centered cubic (fcc) diamond structure, for which the relative positions of the atoms within the unit cell are at $(0, 0, 0)$ and $(0.25, 0.25, 0.25)$ in units of the lattice vectors. What should still be calculated, is the size of this unit cell, measured with the lattice parameter a. Figure 3 shows what the primitive unit cell of silicon looks like. Each of the lattice vectors (lines in the figure) has length $\frac{1}{2}\sqrt{2}a$. The lattice parameter was computed by varying a and looking for an energetic minimum.

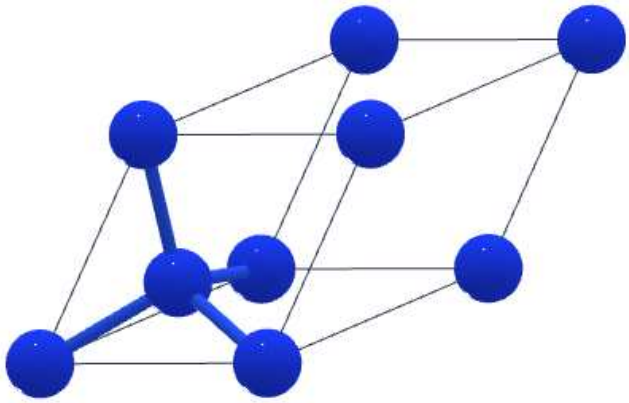

FIGURE 3: Image of silicon primitive unit cell layout, obtained from The Materials Project [14].

This was done twice, once using LDA and once using the PBE functional, which is a type of GGA. All the pseudopotentials used were obtained from PseudoDojo [8]. The resulting plot, including a quadratic fit, can be seen in Figure 4 for GGA.

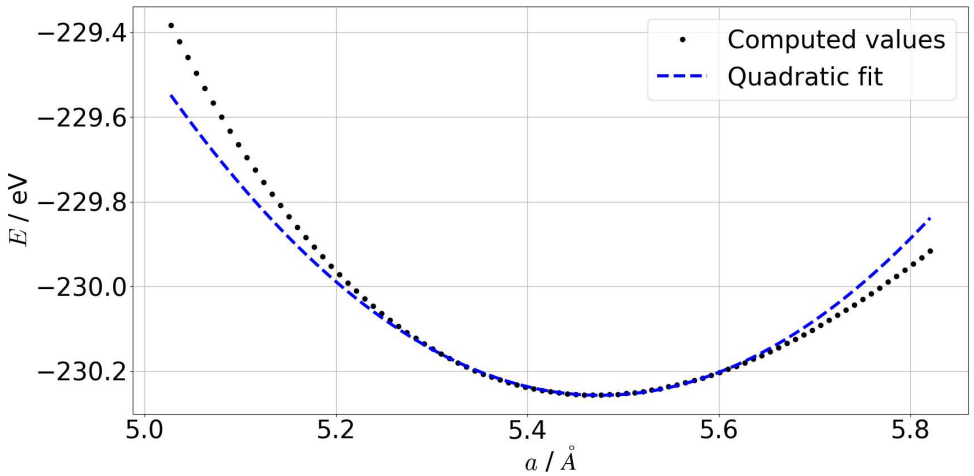

FIGURE 4: Plot of total energy against lattice parameter a for the primitive unit cell of Si using GGA.

From the quadratic fit, a minimum was obtained at $a = 5.471$ Å using GGA and $a = 5.397$

Å using LDA. The experimental value is $a = 5.431$ Å, so these values have errors of $+0.7\%$ and -0.6% respectively [1]. It is usual to see the LDA slightly underestimate the lattice parameter, while GGAs generally overestimates it. This is because, as was described before, atoms are not bound together as strongly within GGAs.

3.3 Primitive Unit Cell Displacements & Forces

After finding the lattice parameter, the next step was to start moving one of the two atoms in the primitive unit cell to find out what forces would act on the atom(s) when the cell is slightly perturbed. As the configuration of fcc silicon is well-known and the lattice parameter has been relaxed, the unit cell should be at an equilibrium. This means that for small displacements, moving along one axis should be representative (forces are assumed to be in the opposite direction of the displacement).

The displacements need to be small enough such that the energy can be reasonably approximated as a second-order polynomial. The first attempt at this was to move the atom such that the distance between the two atoms in the unit cell would increase and decrease by at most 10%. The energies obtained from the DFT calculations can be seen in Figure 5. The fit here is not quite good enough, so a second attempt was made by only making changes up to 3%. This can be seen in Figure 6, which is far closer to a second-order polynomial. This range of 3% was used for the rest of the displacements.

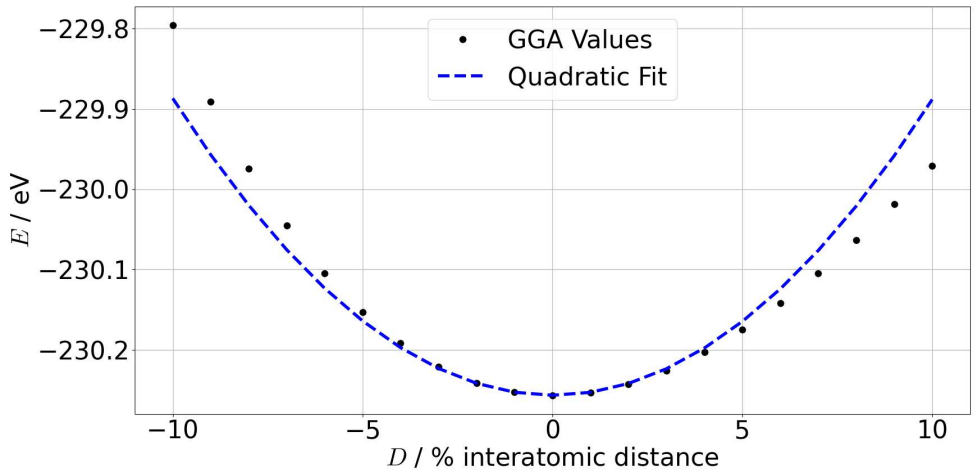

FIGURE 5: Plot of total energy against displacement of a single atom together with a second-order polynomial fit using GGA. Displacements of ±10% interatomic distance

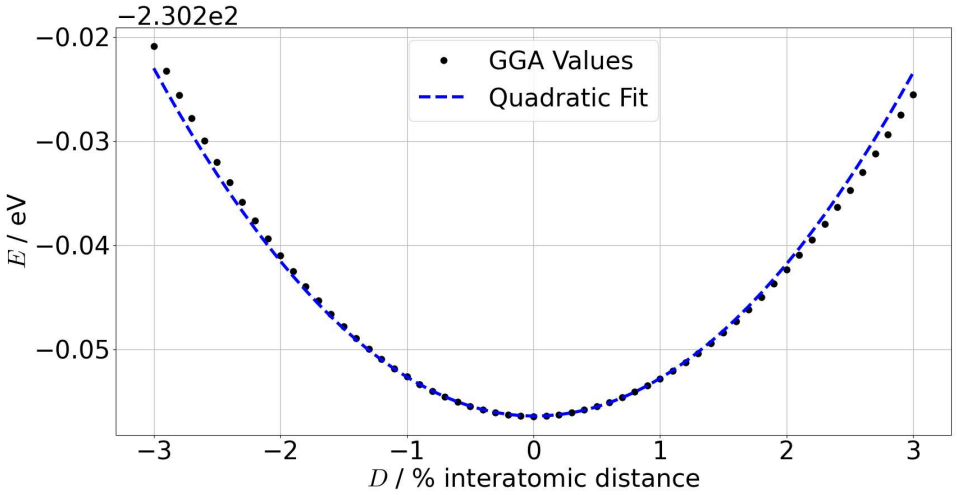

FIGURE 6: Same as Figure 5 with displacements of ±3% interatomic distance

The asymmetry in the energy here results from the configuration of the crystal; when the bond length of the bond between two atoms is shortened, the repulsive force between them grows asymptotically, for a bond length l, it should grow as at least $\frac{1}{l^2}$ because of electrostatic repulsion, a second-order pole is likely an underestimation, as well-known potentials such as the Lennard-Jones potential will often use $\frac{1}{l^{12}}$. When this same bond length is being increased, the energy for that bond does not increase as quickly and the contraction of the bonds with atoms in neighbouring unit cells does not have as much of a contribution because the atom is not being moved directly towards most of them, except one that is too far away to have a large impact, so the asymptote of the force is not approached.

As a final step for the ground state, the force along this axis was calculated as $F = -\frac{\partial E}{\partial D}$ by using finite differences. Specifically, most derivatives were computed using central differences, except the values at the edges which were calculated using forward and backwards differences. Generally central differencing is preferred as for a step size ΔD, the error scales as ΔD^2. Using forward or backward differences, the error scales as ΔD instead.

The resulting plot is visible in Figure 7. This plot also shows three red points resulting from doing relaxation calculations which already exist within Quantum-Espresso. This type of calculation uses quasi-Newton methods to iteratively find the relaxed state of a crystal structure. For this, it calculates the ground-state forces on every atom in each iteration by making use of the Hellmann-Feynman theorem. For three different starting displacements, the force on one of the atoms was added to Figure 7. The picked atom does not matter much, as the forces are identical, but opposite. The relaxation calculations are in line with the forces obtained by using finite differences. The obtained effective spring-constant is $k = 13.01$ eV/Å^2.

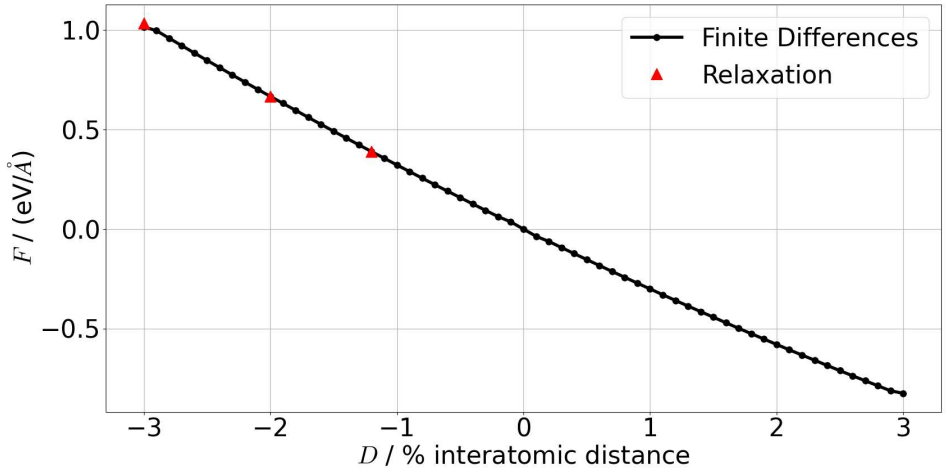

FIGURE 7: Plot of ground-state forces against displacement obtained using finite differences (black circles), as well as forces resulting from relaxation calculations (red triangles).

The next step is to calculate forces from excited-state energies. For this purpose, GW+BSE calculations were performed on each of the same unit cells that were used for the DFT calculations. From these calculations, the excitation energy of the first four excitons was obtained and added to the ground-state energy to obtain a plot of the excited-state energy against the same displacement, this can be seen in Figure 8.

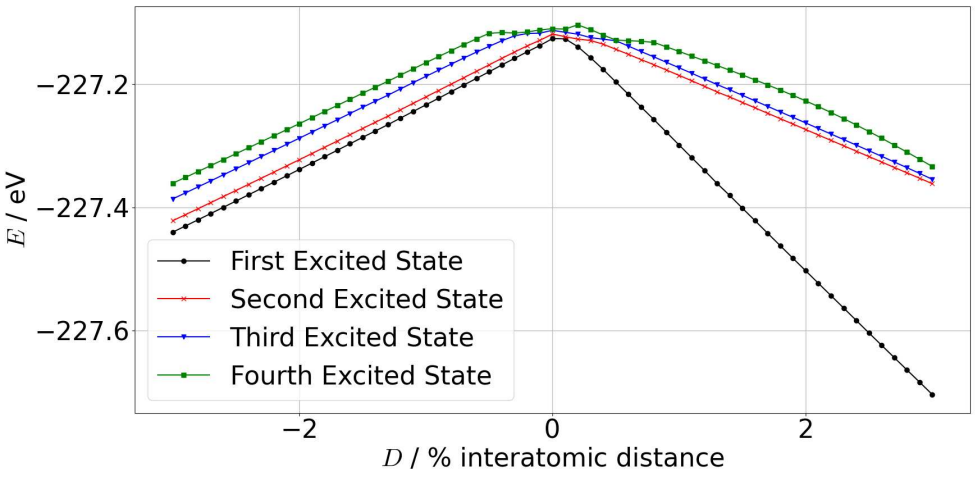

FIGURE 8: Plot of first four excited state energies against displacement of a single atom along one bond

What is surprising here is that the energies do not have a minimum anywhere close to the minimum of the ground-state energy, instead the data seems to be linear everywhere except at the original equilibrium, where there is a (local) maximum.
One possible cause of these results is the crossing of states. By looking close to the equilibrium, one can see that it might not be the same state on either side of the equilibrium, because the states cross through each other.

This behaviour also might be related to both the band structure of silicon and the way in which the calculations were carried out.

Figure 9 shows the band structure within the first Brillouin zone of silicon calculated using DFT with the GGA approximation. One can see that the valence band maximum is at Γ, while the conduction band minimum occurs right before X, this means that silicon has an indirect band gap. The calculated indirect band gap is about 0.6 eV, which is smaller than the experimental value of 1.1 eV [16]. This underestimation is known to happen for Kohn-Sham DFT, but this band structure is still useful for qualitatively understanding the excited-states of silicon.
In our GW+BSE calculations, we only take direct transitions from valence to conduction bands into account, which do not include a transfer of momentum, so the energy will not be accurate to the actual band gap. The lowest direct transition is at the Γ point, as can be seen in Figure 9. The DFT calculations using the GGA approximation give a band gap of ≈ 2.5 eV at this point. The GW band gap is ≈ 3 eV. What this all means, is that the excitation energy used for Figure 8 does not have the first excited state of silicon, but only the first direct excitation.

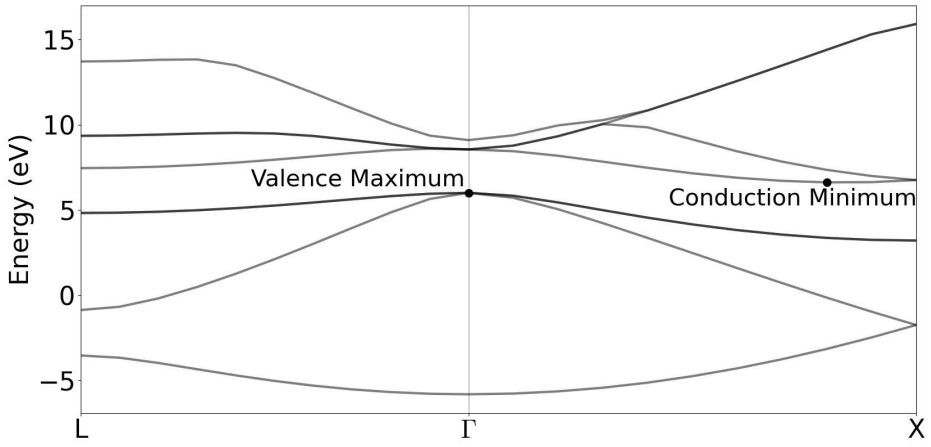

FIGURE 9: Band structure of silicon calculated with GGA. L, Γ and X are high-symmetry points in the first Brillouin zone of silicon. Their reciprocal crystal coordinates are given by $(0.5, 0.5, 0.5)$, $(0, 0, 0)$ and $(0.5, 0.5, 0)$ respectively.

One way to get around this restriction of only seeing direct excitons is by using a 2x2x2 supercell. When a unit cell is used which is twice as large in every direction, the first Brillouin zone of this unit cell shrinks to half its size in each dimension as it exists in the reciprocal lattice, which is a Fourier transform of the real lattice. This means that the conduction minimum and valence maximum are essentially folded close to, but not exactly onto one another. This makes the band gap close to direct. For this reason, all the calculations were repeated using this supercell.

3.4 Supercell Displacements & Forces

One essential difference for the supercell, compared to the primitive unit cell, is that still only one atom in the cell was displaced, while in the case of the primitive unit cell half of

the atoms would be displaced. This was chosen for two reasons, the first of these is that only displacing one atom in the cell leads to a simpler interpretation of which atom the resulting forces are acting on, as it almost has to be that specific atom. The other reason is that only displacing one atom in the supercell reduces the strength of the interactions between neighbouring distortions. This is more accurate to observations as distortions are generally very local, rather than being the same throughout the entire crystal.

The displacements for the ground state were repeated, however the results were virtually identical to those of the primitive unit cell, so those can be found in the appendix. The results for the first excited state are shown in Figure 10. The energy on the y-axis is the sum of the ground-state energy and the energy of the first exciton. Now that a supercell is being used, the energy does have a minimum close to the original equilibrium. This likely occurs because the exciton is now close to direct. The energetic minimum of the first excited state is found at $D = -0.045$ Å when using GGA. This displacement is defined as the change in distance between the atoms at the crystal coordinates $(0.5, 0.5, 0.5)$ and $(0.625, 0.625, 0.625)$. To make sure that the results were independent of the approximation used, they were repeated with LDA as well, which can also be seen in Figure 10. The minimum for LDA was found at a displacement of $D = -0.44$ Å. This is very close, the difference is possibly because the original unit cell for LDA was already slightly smaller.

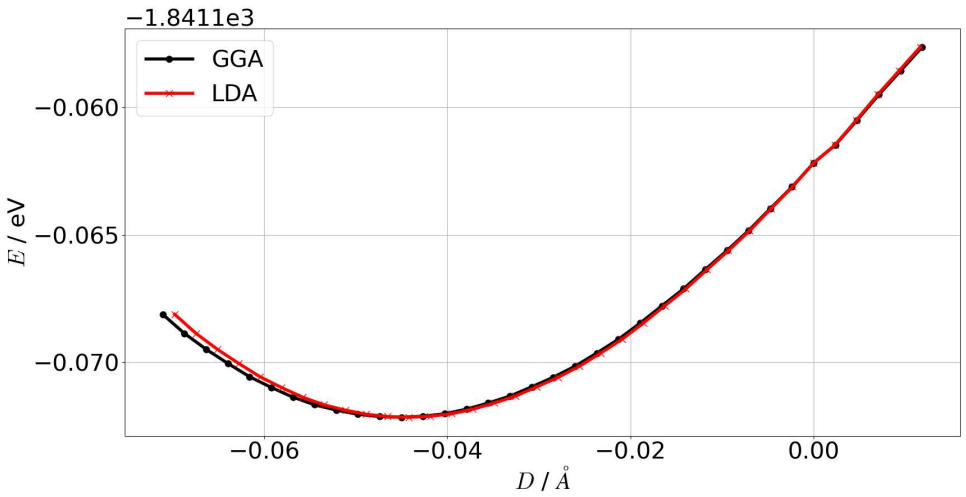

FIGURE 10: Plot of first excited state energy against displacement along one bond of a single atom for the 2x2x2 supercell. Both LDA and GGA results are shown.

Similarly to what was done for the primitive unit cell, the first four excited states were also plotted together to check if the states crossed. In this case, there were clearly no crossings.

The values of the exciton energies now also line up more closely with experimental values. The optical band gap of silicon is 1.11 eV [16, 12]. Figure 11 shows the energy of the first exciton with respect to displacement. Without displacement, the energy of the first exciton is 1.08 eV. This has an error of 2.7% when compared to the experimental value.

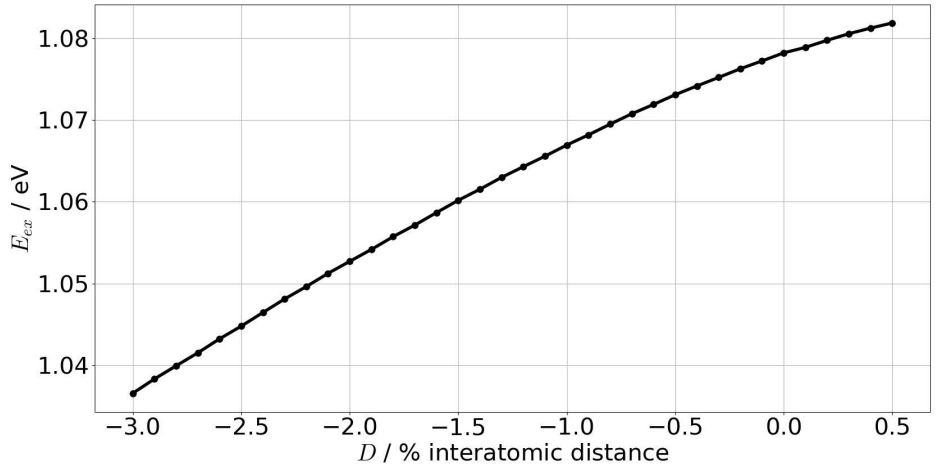

FIGURE 11: Plot of first exciton energy against displacement along one bond of a single atom for the 2x2x2 supercell. Results obtained with GGA.

From here the forces were computed once again using finite differences, these are visible in Figure 12. It is once again mostly linear, although slightly less consistently so than the ground state. One important artifact to mention is the sudden bump in the force around $D = 0$. This hints at the fact that the force may not quite be an entirely smooth function of the displacement, as was also seen in the unit cell. One possible cause for this could be that the band gap is still not quite direct, as in the band structure of silicon, the conduction minimum is not exactly at X, but right before it.

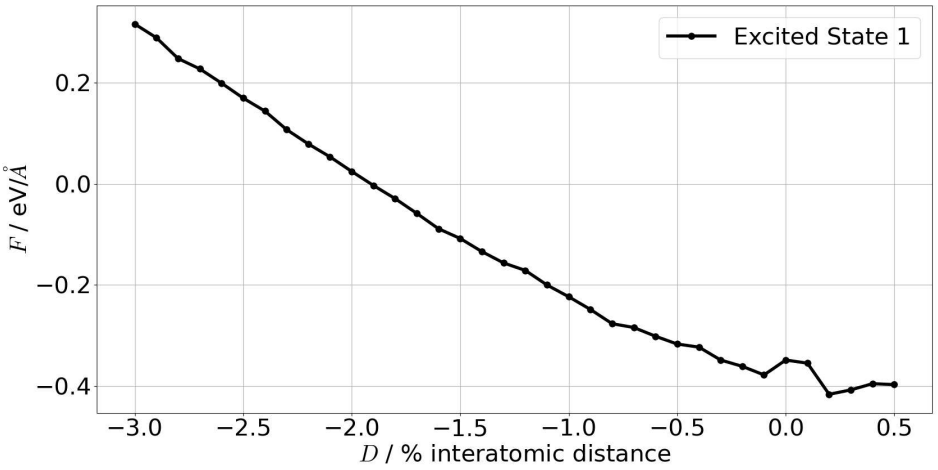

FIGURE 12: Plot of first excited state force against displacement along one bond of a single atom for the 2x2x2 supercell, using GGA.

Some physical interpretation of the way this changes the configuration of the crystal is also good to have. As this force supposedly acts along every bond in the same way, attempting to contract the bond, one likely interpretation is that the entire lattice would shrink by a small amount. To make sure that this is the case, instead of some type of local defor-

mations occurring, the lattice parameter was once again varied just like was done for the ground state of the primitive unit cell. These resulting energies can be seen in Figure 13, indeed a new optimal lattice parameter is found at $a = 5.45$Å.

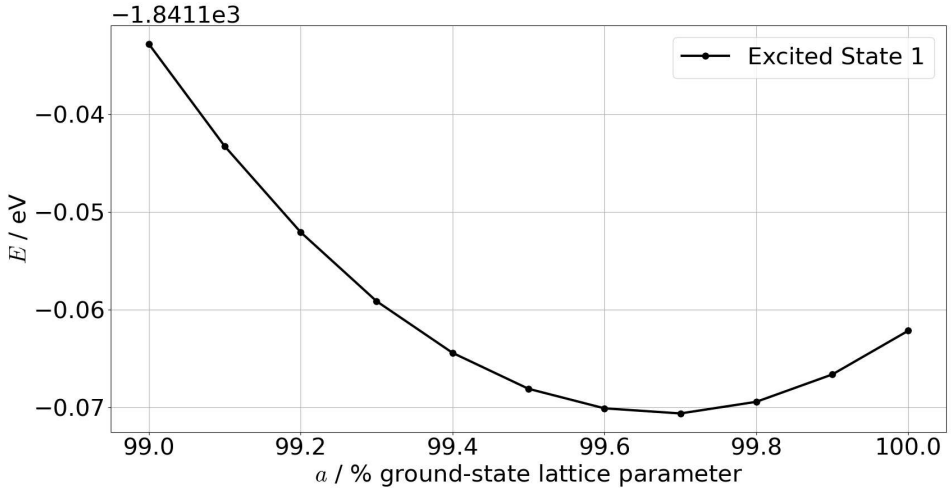

FIGURE 13: Plot of first excited state energy against lattice parameter for 2x2x2 supercell of silicon, using GGA.

With the newly obtained lattice parameter for the excited state, once again a single atom was slightly displaced. Figure 14 shows the forces resulting from applying finite difference to the total energies obtained from these calculations. The force is never zero in this range, so this configuration is not an equilibrium.
The most likely interpretation of this is that some amount of local deformation occurs.

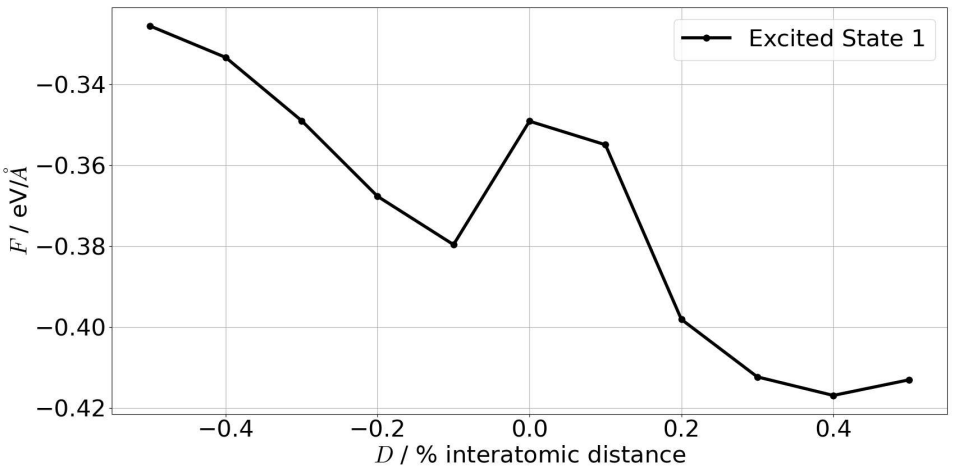

FIGURE 14: Plot of first excited state force against displacement along one bond of a single atom for the 2x2x2 supercell, using GGA.

4 Conclusions

First, the lattice parameter of silicon was calculated to be 5.471 Å, which in good agreement with the experimental values of 5.431 Å.
We then mapped out the excited-state energy landscape of silicon by calculating excited-state energies as a function of the interatomic distance of the two atoms in the primitive unit cell of fcc silicon. In these calculations, we observed exciton energies far above experimental values and no minimum energy was found. We therefore used a supercell approach, which allows us to calculate the energy of the indirect exciton as if it was (almost) direct. Some behaviour similar to that of the primitive unit cell was still visible, but the effects of it seemed to be reduced a lot. The exciton energy in the supercell was 1.08 eV, rather than the ≈ 3 eV of the primitive unit cell. This is in far better agreement with the experimental value of 1.11 eV.
We found that even after relaxing the lattice parameter in the first excited state to 5.45 Å, there were still forces on the atoms, implying that some reconfiguration of the crystal also occurs during excitation.

The lack of smoothness around the equilibrium state for most of the plots does suggest that the forces there are not quite accurate, likely because the supercell is still too small. It would be interesting to see if the force smooths out as the size of the cell increases, but this would not be feasible with the current methods as the cell would need to be much larger for the conduction minimum to approach the valence maximum.

There are some more aspects of this report that can be expanded on. The first of these would be to relax the structure of the crystal to see what the new optimal configuration is.
Semiconductors other than silicon should also definitely be used. The results using silicon usually have some unexpected behaviour around the equilibrium, which likely results from an indirect band gap. This would not be present in a material that has either a direct band gap, or at least has both its valence maximum and conduction minimum at a high-symmetry point.
Some more evaluation of the accuracy of the results could also be done, such as by investigating the finite difference method more closely or comparing with other types of calculations like Quantum Monte Carlo.
Finally, one of the next steps of research should be to implement the method from Ismail-Beigi and Louie[11] for crystals, in order to test the accuracy of the approximations made for that approach.

5 Bibliography

References

[1] CODATA Value: lattice parameter of silicon.

[2] Phys. Rev. Lett. 77, 3865 (1996) - Generalized Gradient Approximation Made Simple.

[3] Klaus Capelle. A bird's-eye view of density-functional theory. *Brazilian Journal of Physics*, 36(4a):1318–1343, December 2006.

[4] Jack Deslippe, Georgy Samsonidze, David A. Strubbe, Manish Jain, Marvin L. Cohen, and Steven G. Louie. BerkeleyGW: A Massively Parallel Computer Package for the Calculation of the Quasiparticle and Optical Properties of Materials and Nanostructures. *Computer Physics Communications*, 183(6):1269–1289, June 2012. arXiv:1111.4429 [cond-mat].

[5] P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo, A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio Jr, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, E Küçükbenli, M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V Nguyen, A Otero-de-la Roza, L Paulatto, S Poncé, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with QUANTUM ESPRESSO. *Journal of Physics: Condensed Matter*, 29(46):465901, 2017.

[6] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502 (19pp), 2009.

[7] Paolo Giannozzi, Oscar Baseggio, Pietro Bonfà, Davide Brunato, Roberto Car, Ivan Carnimeo, Carlo Cavazzoni, Stefano de Gironcoli, Pietro Delugas, Fabrizio Ferrari Ruffino, Andrea Ferretti, Nicola Marzari, Iurii Timrov, Andrea Urru, and Stefano Baroni. Quantum ESPRESSO toward the exascale. *The Journal of Chemical Physics*, 152(15):154105, 2020. tex.eprint: https://doi.org/10.1063/5.0005082.

[8] D. R. Hamann. Optimized norm-conserving vanderbilt pseudopotentials. *Phys. Rev. B*, 88:085117, Aug 2013.

[9] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864–B871, 1964.

[10] Mark S. Hybertsen and Steven G. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Physical Review B*, 34(8):5390–5413, October 1986. Publisher: American Physical Society.

[11] Sohrab Ismail-Beigi, Steven G Louie, and Berkeley Lawrence Berkeley National Lab (LBNL), CA (United States). Excited-state forces within a first-principles green's function formalism. *Physical Review Letters*, 90(7), 2002.

[12] G. G. Macfarlane, T. P. McLean, J. E. Quarrington, and V. Roberts. Exciton and phonon effects in the absorption spectra of germanium and silicon. *Journal of Physics and Chemistry of Solids*, 8:388–392, January 1959.

[13] Richard M. Martin. *Electronic structure: basic theory and practical methods*. Cambridge University Press, Cambridge, UK, 2004. Section: xxiii, 624 pages : illustrations ; 26 cm.

[14] The Materials Project. mp-149: Si (Cubic, Fd-3m, 227).

[15] Michael Rohlfing and Steven G. Louie. Electron-hole excitations and optical spectra from first principles. *Physical Review B*, 62(8):4927–4944, August 2000. Publisher: American Physical Society.

[16] W. H. Strehlow and E. L. Cook. Compilation of Energy Band Gaps in Elemental and Binary Compound Semiconductors and Insulators. *Journal of Physical and Chemical Reference Data*, 2(1):163–200, January 1973.

A Ground-state Forces of Supercell

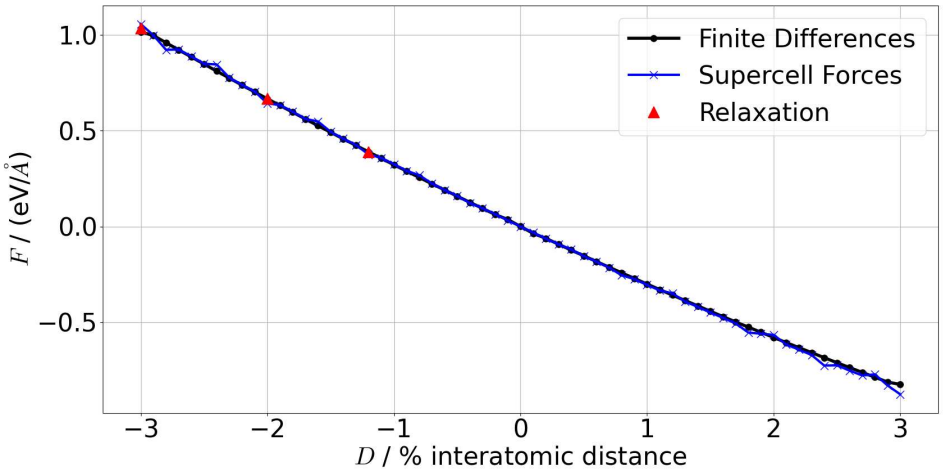

FIGURE 15: Plot of 2x2x2 supercell ground-state forces against displacement obtained using finite differences (black circles), primitive unit cell ground-state forces against displacement obtained using finite differences (black circles), as well as forces resulting from relaxation calculations (red triangles).