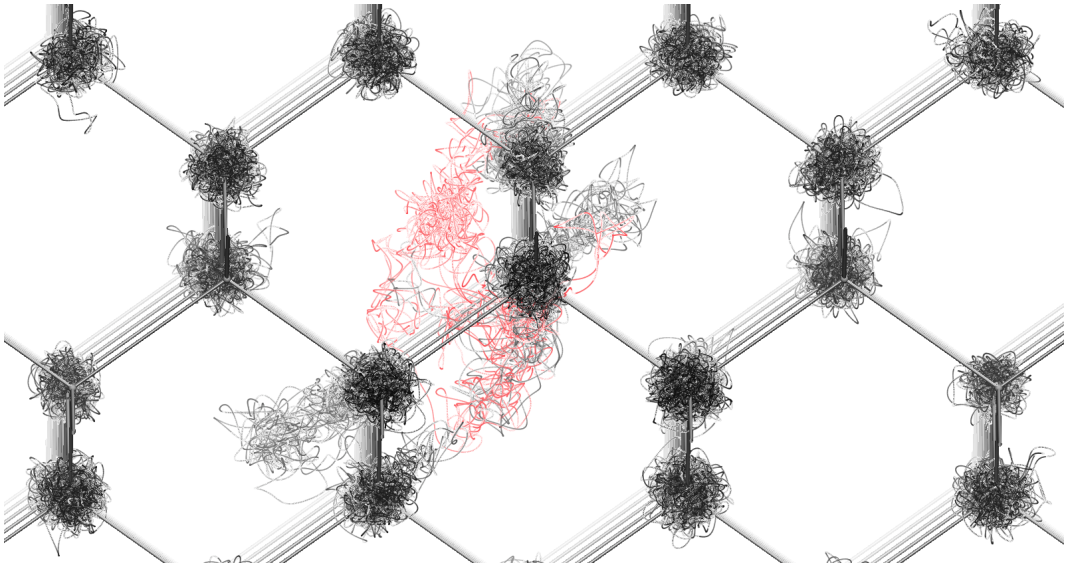
Ab initio molecular dynamics simulation of phosphorus diffusion in silicon

Dan Michael Olsen Heggø



THESIS

for the degree of

MASTER OF SCIENCE

Program for Materials Science and Nanotechnology

Faculty of Mathematics and Natural Sciences
University of Oslo

September 2012

Contents

1. Background

1.1. Introduction

Molecular dynamics with an *ab initio* potential, such as one obtained from DFT calculations, is very enticing from a theoretical point of view, but is often practically unfeasable due to the great computational cost involved. This work was started without knowing *a priori* if simulations could be carried out with simulation times long enough to obtain useful statistics, but some hope came from the successful simulations of the silicon self-interstitial by Sahli,[1] and from the recent development of 'very soft' PAW-potentials, that is potentials working accurately even with a very low energy cutoff (see section 2.6).

The alternative to using an *ab initio* potential is to use an empirical potential. Two well-tested empirical potentials for silicon are the Stillinger-Weber (SW) potential[2] and the Tersoff (T3) potential.[3] A drawback with empirical potentials is that they are defined for a limited range of conditions and problems, and it is generally not known how well they perform outside that range. A given potential may for instance capture elastic and cohesive properties of a material with the same accuracy as an *ab initio* potential, but fail for other properties.[4] For silicon self-diffusion, Posselt et al. found that the activation energies differed by about 1 eV between the SW potential and the T3 potential in a non-systematic way, and that the migration mechanism also depended on the choice of potential.[5]

With empirical potentials, transferability between different bonding environments can also not be expected, so rather than a general phosphorus potential, a specific parametrisation for phosphorus in silicon is needed. Over the years, many groups have developed parameters for more systems that can be used with the original form of the Tersoff potential, for instance for all III-V semiconductors.[4] Yet, at least to my knowledge, no such potential for phosphorus in silicon has been published.

A motivation for studying phosphorus diffusion in silicon has been the use of phosphorus-doped silicon in the emitter layer in classical silicon solar cells. This front layer of the solar cell is contacted with metal fingers, and the contact resistivity ρ_c depends strongly on the phosphorus doping concentration N_D. If the metal contacts take up 5% of the cell surface, a contact resistance $\rho_c < 2 \times 10^{-3}\,\Omega\,\mathrm{cm}^2$ is required to not degrade the output power by more than 0.5%. This in turn requires a doping concentration $N_D > 1 \times 10^{19}\,\mathrm{cm}^{-3}$,[6] and even higher doping concentrations is required if the area occupied by metal contacts is to be reduced further, or for contacting concentrator cells. Concentrations may reach $1 \times 10^{21}\,\mathrm{cm}^{-3}$ in the outermost part of a typical emitter layer,[7, p. 40] corresponding to about one phosphorus atom for every 50 silicon atoms, while the limit of electrically

active phosphorus is somewhat lower, about $2 \times 10^{20}\,\mathrm{cm}^{-3}$.[7, p. 40] In this work, a unit cell is simulated with one phosphorus atom to 64 silicon atoms, the concentration thus being close to the concentration in the outermost part of the emitter layer, but this unit cell approach does not allow for any phosphorus clustering, although phosphorus precipitation in reality can be expected at this concentration.

A privilege with studying phosphorus in silicon is that there exists a wealth of experimental and theoretical results from more than 60 years of studies on this system,[1] results that can be used to test new simulation results. Diffusion of dopants in silicon is discussed in e.g. the reviews by Fahey (1989)[8], Hu (1994)[9], and Pichler (2004)[10]. Some key properties of pure silicon are listed in table 1.1, while diffusion of phosphorus in silicon is discussed in section 1.3.

Crystal structure	Diamond
Lattice constant a / Å	5.43072 ± 0.00004 Å at 25 °C [11]
Si-Si distance / Å	$\sqrt{3}a/4 = 2.35$ Å
Atomic density	$8/a^3 = 4.9948 \times 10^{22}\,\mathrm{cm}^{-3}$
Melting point	1685 K / 1412 °C
Thermal expansion coefficient	$2.6 \times 10^{-6}\,\mathrm{K}^{-1}$[12]

Table 1.1.: Some key properties of silicon at room temperature.

1.2. Diffusion and diffusivity

Diffusion is macroscopic redistribution of atoms[2] due to their microscopic thermal motion. Such redistribution occurs in all materials, but at a negligible rate at temperatures considerably below the melting point of the material. For silicon, with a melting point of 1414 °C, diffusion is negligible at room temperature, but significant at elevated temperatures, and controlled diffusion is commonly carried out at temperatures of about 800-1000 °C.

How do we describe and model diffusion? As is not uncommon in physics, different pictures exist for different scales of interest. When it comes to describing diffusion, two main pictures exists: the continuum picture and the atomistic picture. Since materials consists of atoms, it is clear that the most fundamental picture of diffusion should be an atomistic one. But a description involving every single atom becomes highly impractical on a macroscopic scale, in which it becomes more practical to invoke a viscous fluid model in which the atoms are smoothed out into a continuum. The two pictures can be

[1] Silicon semiconductor technology took it first infant steps right before world war II, with for instance the discovery of the p-n junction (then called 'PN barrier') often attribtued to R.S Ohl (U.S. patent 2402662, 1939), and the understanding and development gained speed in the early 50s. In a 1953 patent by P. Robsin, the process of diffusion to form a junction is described (U.S. patent 2823149, 1953), and in a 1955 patent by Ross (2862160), phosphorus is mentioned as a donor, but I'm not sure just when phosphorus was first introduced.

[2] I will only be concerned with atoms here, but the concept of diffusion also applies to other small particles such as molecules, viruses and bacteria.

connected through parameters known as diffusion coefficients in order to produce integral diffusion models, but the diffusion coefficients are also commonly obtained empirically.

1.2.1. The continuum picture

Following Joseph Fourier's 1822 treatise on heat diffusion[13] and Georg Ohm's 1827 work on 'electricity diffusion' (Ohm's law),[14] Adolf Fick presented his phenomenological description of 'mass diffusion' in 1855.[15] In this picture, a species with concentration $c(\mathbf{r})$ will flow in the opposite direction of the concentration gradient of that species, $\nabla c(\mathbf{r})$. That is, it will show a flux,

$$\mathbf{J}(\mathbf{r}, t) = -D\nabla c(\mathbf{r}, t). \tag{1.1}$$

This relationship between the concentration gradient and the flux is known as Fick's first law, and the constant of proportionality D is known as the coefficient of diffusion or *diffusivity* for the species in question. Often reported in units of $\mathrm{m^2\,s^{-1}}$ or $\mathrm{cm^2\,s^{-1}}$, the diffusivity is in general a tensor, but it reduces to a scalar in cubic crystals. In general the diffusivity of species i will depend on both the concentration of the same species c_i and of other species c_j, on temperature, pressure, crystal defects and other factors that will be quickly discussed below.

If G and R are the generation and recombination rates for the species in question, a continuity equation could be set up as

$$\frac{\partial c}{\partial t} = -\nabla\mathbf{J} + G - R \tag{1.2}$$

In the absences of sinks and sources ($G = R = 0$), the continuity equation combined with Fick's first law forms what is known as Fick's second law,

$$\frac{\partial c}{\partial t} = \nabla(D\nabla c) = D\nabla^2 c \tag{1.3}$$

where the last equality is true only when D is independent of position (via the concentration).

The picture is slightly complicated by the fact that the diffusivity D itself may vary with the concentration c. For the diffusivity of species at very low concentrations, such as dopants, this concentration dependence is generally negligible, but at high concentrations, such as the typical phosphorus concentrations in the solar cell emitter layer, the effect may be quite significant. As an example, A. Bentzen has obtained the concentration dependence of the phosphorus diffusion coefficient over a large range of concentrations using Boltzmann-Matano analysis of experimental data at 890 °C. From the intrinsic value of about $4 \times 10^{-14}\,\mathrm{cm^2\,s^{-1}}$, his diffusivity increases slowly up to a maximum of $1 \times 10^{-13}\,\mathrm{cm^2\,s^{-1}}$ at a phosphorus concentration of $10^{19}\,\mathrm{cm^{-3}}$, before it dips down to $1 \times 10^{-14}\,\mathrm{cm^2\,s^{-1}}$ at $10^{20}\,\mathrm{cm^{-3}}$, before it increases again.[7] Concentration gradients will nevertheless not be considered in this work.

The diffusivity depends strongly on the temperature, and the temperature dependence found from experiments often takes the simple Arrhenius form,

$$D(T) = D_0 \mathrm{e}^{-E_a/k_B T} \tag{1.4}$$

defined by a prefactor D_0 and an activation energy E_a. Taking the logarithm on both sides, we see that $\ln D$ as a function of $1/k_B T$ forms a straight line with slope $-E_a$ and intercept $\ln D_0$.

$$\ln D = \ln D_0 - E_a (k_B T)^{-1} \tag{1.5}$$

The parameters D_0 and E_a can thus easily be extracted if a good fit can be made to a straight line.

As mentioned, the diffusivity may also depend on other factors, such as crystal defects, pressure, et cetera, that will not be touched upon in this work, but these are discussed by e.g. Pichler.[10]

1.2.2. The atomistic picture

In the continuum picture, the diffusion coefficients are just empirical parameters, and we have to invoke a model involving atoms to explain them. Such a model was found independently by Einstein in 1905,[16] and Smoluchowski in 1906.[17]

Einstein related the Gaussian diffusion outwards from a point to the probability of finding a particle moving by a random walk process,

$$f(x,t) = \frac{n}{\sqrt{4\pi D}} \mathrm{e}^{-x^2/4Dt} / \sqrt{t}, \tag{1.6}$$

thus relating the mean square displacement of a particle (in one dimension), $\langle x^2 \rangle$, to the diffusivity[16],

$$\langle x^2 \rangle = 2Dt, \tag{1.7}$$

a relation that will be discussed further in section 3.1.5.

The diffusivity of a species can thus be found from a knowledge of the time-resolved displacement of all atoms of that species. Such a naïve approach to finding the diffusivity is often hindered, both in experimental and theoretical methods, by difficulties in obtaining this detailed microscopic information, but from a molecular dynamics simulation it's readily available.

The major problem with obtaining diffusivities from a molecular dynamics simulation, is that diffusion can be a quite slow process. For any given system, the simulation of a large number of atoms over a long enough amount of time, may not be computationally feasible.

In the solid phase, diffusion is several orders of magnitude slower than diffusion in the liquid or gas phase. At a finite temperature T, the atoms will have kinetic energies according to a Boltzmann distribution for that temperature (see section 3.1.4), but since the atoms are 'trapped' by the neighbouring atoms in the crystal, the kinetic energy goes

into vibrations most of the time, and averaged over the timescale of the vibrations, the atoms will appear to be more or less fixed in space.

Every now or then, however, an atom will obtain enough energy to 'jump' to a different position. A clear example of such a jump is illustrated in fig. 1.1, where a phosphorus atom is seen making a jump from its lattice site into a neighbouring vacant site at 1400 K, illustrating the vacancy mechanism of diffusion, at least one part of it. For this to really constitute diffusion, the vacant site left behind by the P atom must also diffuse away from the P atom before the P atom happens to jump back into it.

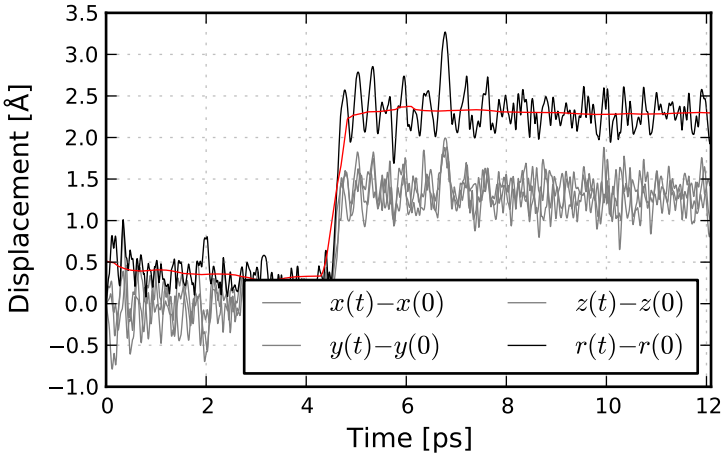


Figure 1.1.: A 'jump' from one lattice site to another. The red line is a symmetric running mean over 0.5 ps followed by a symmetric running median over 2.0 ps.

Due to the energy barrier that must be passed for the jump to take place, this is an example of what is called an *activated process*, and the threshold energy is called an activation energy E_a. In general there is not just one mechanism involved in diffusion, and the value for the activation energy obtained by experiment or simulation is a weighted average over all possible mechanisms. One mechanism may however be dominating to such an extent that other mechanisms can be ignored.

Back to the concept of 'jumps', we can find the fraction of atoms having kinetic energy greater than a given energy E_a as $\mathrm{e}^{-E_a/KT}$ from the Boltzmann distribution, and so the probability for a jump to happen is just e^{-E_a/k_BT}. In a simple model, we can then estimate the number of jumps per second by multiplying this probability by an 'attempt frequency' ν,

$$\Gamma = \nu \mathrm{e}^{-E_a/k_BT} \tag{1.8}$$

For a rough estimate, we may take the frequency to be 10^{13} Hz, a typical textbook value for the vibration of atoms in a crystal at room temperature.[18, p. 212] As seen in fig. 1.2, the atoms do vibrate at frequencies of roughly this order of magnitude, but the motion is neither simple nor harmonic. If it was, we would not need *simulation* anyway.

In fig. 1.3, the estimated number of jumps in 100 ps is plotted from (1.8), for three different activation energies. 100 ps was choosen, since it's about the maximum simulation

time we can conveniently sample.

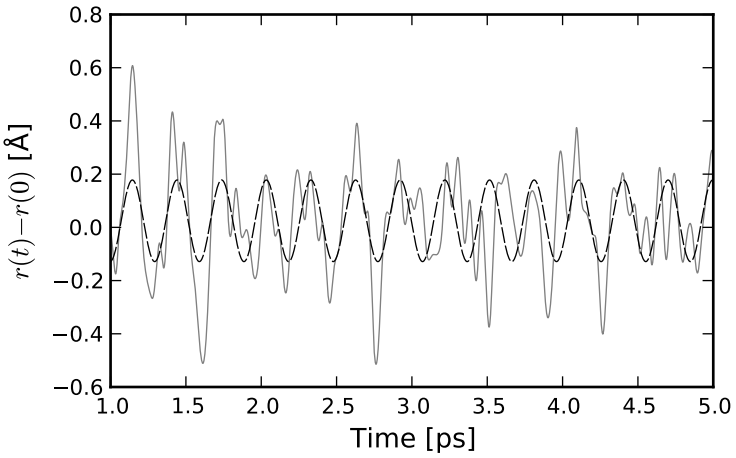


Figure 1.2.: Simulated motion of a silicon atom in a 64-atom supercell at 1400 K (solid line), with a least square fit to a function $f(t) = a\cos[(2\pi f)t + c] + d$ (dashed line). The fitted frequency is $f = 3 \times 10^{12}$ Hz, corresponding to a period of 0.3 ps.

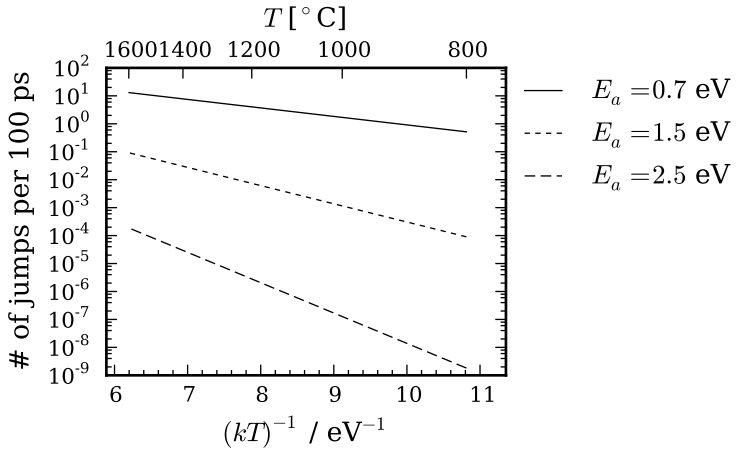


Figure 1.3.: Estimated number of jumps in 100 ps, from (1.8) with $\nu = 10^{13}$ Hz.

1.3. The case of phosphorus diffusion in silicon

1.3.1. Solubility limit and preferred site

According to the standard phosphorus-silicon binary phase diagram,[19] more than 2% P can be solved in silicon at temperatures between 1350 K and 1500 K without the

formation of new phases, and even at 1000 K more than half a percent can be solved.[3]

Sheet resistivity and Hall effect measurements have shown that below about 2–3 $\times 10^{20}\,\mathrm{cm}^{-3}$ (or 0.4-0.6%), the density of electrically active dopants in properly annealed samples follow the chemical concentration of phosphorus closely, but above such a concentration, the density of electrically active dopants n_e plateau.[21] A concentration of about half a percent therefore seems to be the maximum solubility of unclustered P,[22] but higher concentrations are common in the heavily doped emitter layer in solar cells.[23] The nature of the electrically inactive phosphorus has not been fully established yet, but a mix of SiP precipitates and mobile P defects is suggested.[24][25]

The atomic radii of silicon and phosphorus are 1.17 and 1.10 Å, respectively. Through the observation of a very slight lattice contraction following the addition of phosphorus, it was demonstrated in the 1940s that phosphorus enters mainly substitutional sites.[26]

1.3.2. Interstitials

While phosphorus prefers substitutional sites, phosphorus interstitials are of interest for their role in diffusion processes, as discussed in the next section. Direct experimental evidence for phosphorus interstitials is probably missing, so their binding energy have formerly been taken from fitting diffusion models to experimental data, and lately also from purely theoretical investigations,[10, p. 391] such as nudged elastic band studies.

The tetrahedral structure of silicon is a relatively open one, with a spherical packing density of only $\pi\sqrt{3}/16 \approx 0.34$,[4]. The eight-atom cubic unit cell, shown in fig. 1.4a, contains eight equivalent tetrahedral voids (T_d symmetry),[5] all potentially large enough to hold an atom without distorting the surrounding structure considerably. Interstitials at two different tetrahedral sites are shown in fig. 1.4b and 1.4c.

A tetrahedral site can be distorted into a "hexagonal" site with D_{3d} symmetry, at the centre of a slightly distorted hexagon, with normal in the $\langle 111\rangle$ direction. Fig. 1.4d shows a hexagonal interstitial at $(\frac{3}{8}, \frac{5}{8}, \frac{3}{8})$, distorted from the tetrahedral position $(\frac{4}{8}, \frac{4}{8}, \frac{4}{8})$ in fig. 1.4b.

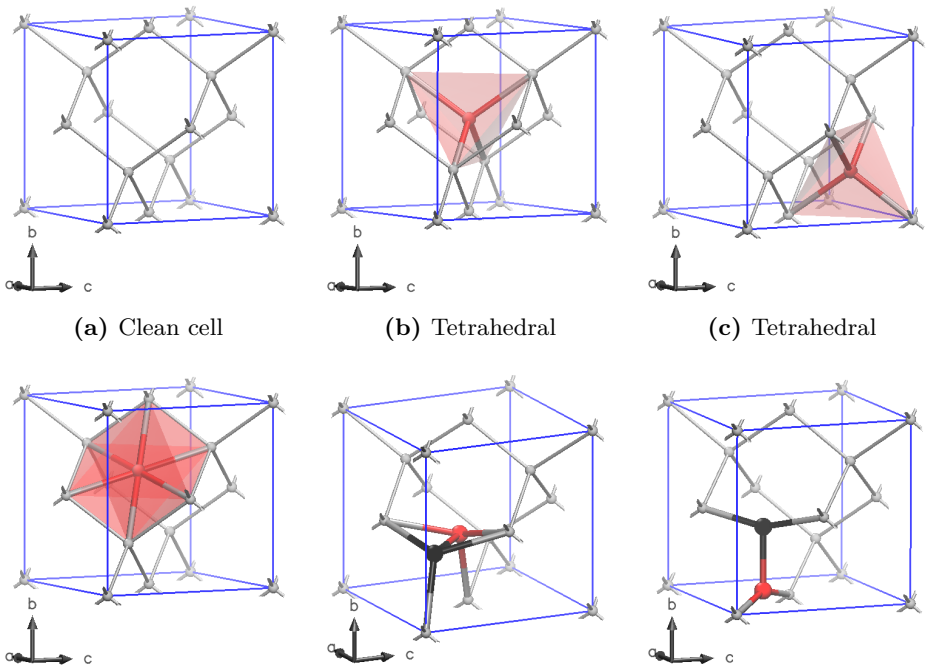
An alternative to the "pure" tetrahedral and hexagonal interstitials are the "interstitialcies", consisting of $N+1$ atoms sharing N lattice sites, of which the simplest, consisting of two atoms sharing one site, are called "split interstitials" or "dumb-bell interstitials".[10, p. 6] Two symmetrically inequivalent such configurations exists, the "split $\langle 110\rangle$" and the "split $\langle 001\rangle$" configurations depicted in fig. 1.4e and 1.4f.

For equivalent atoms, the split $\langle 001\rangle$ site has D_{2d} symmetry, reducing to C_{2v} for inequivalent atoms, while the split $\langle 110\rangle$ site has C_{2v} symmetry for equivalent atoms, reducing

[3] Solubility limits are not easy to measure to high precision. A recent compilation by Safarian and Tangstad shows the variation between different studies, and suggests lower solubilities at elevated temperatures.[20]

[4] Close-packed spheres have a spherical packing density of $\pi/3\sqrt{2} \approx 0.74$.

[5] Tetrahedral sites are found at $(\frac{1}{4}, \frac{1}{4}, \frac{3}{4})$, $(\frac{1}{4}, \frac{3}{4}, \frac{1}{4})$, $(\frac{3}{4}, \frac{1}{4}, \frac{1}{4})$, $(\frac{3}{4}, \frac{3}{4}, \frac{3}{4})$, at the edges $(0, 0, \frac{1}{2})$, $(0, \frac{1}{2}, 0)$, $(\frac{1}{2}, 0, 0)$, and at the centre $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$



(a) Clean cell **(b)** Tetrahedral **(c)** Tetrahedral

(d) Hexagonal **(e)** split ⟨110⟩ **(f)** split ⟨001⟩

Figure 1.4.: Interstitial and interstitialcy configurations in the cubic silicon unit cell, with edge atoms from neighbouring cells included.

to C_{1h} (just a mirror plane) for inequivalent atoms.[10, p. 12] An advantage of the lower symmetry of the split ⟨110⟩ site compared to the split ⟨001⟩ site is that there is freedom for relaxing the bond lengths. While the split ⟨001⟩ site is quite strained, the atoms in the split ⟨110⟩ position can relax freely in the ⟨001⟩ direction. In figure 1.4e normal bond lengths have been obtained without relaxing the neighbour atoms.

Early theoretical investigations by Car suggested the energetical stability of P at the hexagonal site over the tetrahedral site and P-Si interstitialcies.[27] In a 2003 nudged elastic band (NEB) study, Liu *et al.* identified four lowest energy configurations for insterstitial phosphorus in silicon;[28]

- P_i^X, a split ⟨110⟩ configuration (fig. 1.4e) distorted in the 110 direction so that symmetry is reduced from C_{1h} to C_1.
- $P_i^{X_2}$, a configuration with approximate C_2 symmetry, suggested to be viewed as a P_i^X configuration with local structural distortion.
- P_i^H, a hexagonal configuration with D_{3d} symmetry, as in fig. 1.4d
- P_i^S, a split ⟨001⟩-like configuration, with symmetry reduced from C_{2v} to C_{1h}.

Of these, the P_i^X and $P_i^{X_2}$ were reported to be the lowest energy structures for the neutral charge state, and P_i^H and P_i^S the lowest energy structures for the +1 charge state.[28]

A 2006 NEB study by Harrison et al. confirmed a hexagonal site as the most stable one for the +1 charge state, but suggested a "bond-centre" site to be preferred in the neutral case.[29] The suggested name is a bit confusing, since the configuration is not similar to the traditional bond-centre site described in Pichler[10, p. 12], which is literally at the center of a Si-Si bond and has C_{3v} symmetry. Instead, P is displaced 1.55 Å normal to the bond, forming a triangle, with the two Si atoms with the two Si-P bonds being of equal length. The Si-Si bond part of the triangle increased in length by 30 %, so its possible that the configuration could be viewed as a form of interstitialcy with three atoms sharing two lattice sites.

1.3.3. Diffusion

Substitutional phosphorus, like other substitutional defects, is considered immobile, but can enter a mobile state through a (random) encounter with mobile defects such as interstitials or vacancies, or by concerted exchange (direct interchange of a pair of atoms). Many different mechanisms come into play, but it is commonly accepted today that the dominant mechanism for phosphorus diffusion in silicon is mediated by self-interstitials (or -interstitialcies), not by vacancies or concerted exchange, except at very high doping concentrations, for which it has been suggested that vacancy-mediated diffusion becomes the dominant mechanism.[7, 28, 30]

In the classic interstitial mechanism, a self-interstitial, which migrates through the interstices of the lattice, approaches the substitutional defect and "kicks" it out into an

interstitial position, itself taking over the lattice site. The process is often called "kick out reaction";[8]

$$\mathrm{I} + \mathrm{P} \rightleftarrows \mathrm{P_i},$$

where I is the self-interstitial, P the substitutional phosphorus, and $\mathrm{P_i}$ the interstitial phosphorus,[6] which can migrate and eventually kick out a silicon atom to return to a lattice site.

< FIGUR ála fig. 17 i Fahey >

In the classic interstitial mechanism, the mobile defect is considered to be a single atom, either a host-interstitial or a dopant-interstitial. An alternative is to consider instead a "interstitialcy", a term coined by Frederick Seitz, better known for the Wigner–Seitz cell, to describe an entity that "has the same type of phase-like existence as a vacancy or a dislocation, since the pattern can move without having a fixed atom move with it",[31] and commonly associated with a pair of atoms.

Following Fahey,[8, p. 322] the migration of an interstitialcy with a pair of atoms centred around a lattice site A, takes place when the pair is shifted along its principal axis towards another (occupied) lattice site B to such an extent that one of the two atoms ends up forming a new pair with the atom at site B, while the other atom relaxes at site A. If the atom at site B happens to be a P atom, it becomes part of a P-Si pair (PI using defect nomenclature), that can then migrate in the same manner. The process is written (self-interstitials and -interstitialcies are both denoted I)

$$\mathrm{I} + \mathrm{P} \rightleftarrows \mathrm{PI}.$$

It's important to note that it is not the same pair of atoms that migrate over time. Each time the interstitialcy makes a jump, one of the atoms moves along, while the other is left behind, and each jump is considered independent of the previous one.[8, p. 323]

< FIGUR ála fig. 15 i Fahey >

In practice, distinguishing between interstitial and interstitialcy mechanisms is often not needed, and it's also very hard to distinguish the two experimentally.[8, p. 295]

< SETT INN FIGUR INSPIRERT AV Martin-Bragado 2005 > ?

On the overall activation energy for phosphorus diffusion, Fahey wrote in a 1989 review on point defects and diffusivities in silicon, that "determination of Q_P by different studies using a variety of techniques agree to within 0.2 eV and are considered accurate."[8] The situation is more or less the same today, with an accepted overall activation energy of about 3.5 eV.[10] Interestingly, two studies in the early 2000s suggested a significantly lower value of 2.7–2.8 Å at low temperatures,[32][33] indicating the possibility of a change in dominant diffusion mechanism at some temperature, but this is yet to be confirmed.

Table 1.2 lists a selection of reported diffusivities for phosphorus, including two review values, three 'non-standard' values, and two values from purely theoretical investigations. Most older results form part of the review results by Fahey[8] and Pichler,[10] and are not included separately.

[6] For an introduction to defect nomenclature, see e.g. Fahey[8].

E_a [eV]	Conditions	Ref.
3.51–3.67	selective listing of earlier work	Fahey(1989)[8, p. 320]
3.507	regression 900-1200 °C from 16 studies	Pichler(2004)[10, p. 395]
2.81	850, 900, 1000 °C	Haddara(2000)[32]
2.74 ± 0.07	810 – 1100 °C	Christensen(2003)[33]
2.88 ± 0.15	800 – 1050 °C	Naganawa(2008)[34]
3.1–3.5	NEB	Liu(2003)[28]
3.43	NEB	Harrison(2006)[29]

Table 1.2.: Activation energies for overall diffusion of phosphorus in silicon.

An activation energy of about 3.5 eV would require very long simulations to be able to see any jumps. While the estimate presented in fig. 1.3 is a very rough one, the indicationt that simulating diffusion with such an activation energy is several orders of magnitude out of reach, is not encouraging. On the other hand, the total activation energy of interstitial diffusion can be partitioned into a formation enthalpy of the PI complex, $H^{\mathrm{f}}_{\mathrm{PI}}$, and a migration energy of the complex, $E^{\mathrm{m}}_{\mathrm{PI}}$;[8, p. 323]

$$E_{\mathrm{PI}} = H^{\mathrm{f}}_{\mathrm{PI}} + E^{\mathrm{m}}_{\mathrm{PI}}, \tag{1.9}$$

with the formation enthalpy $H^{\mathrm{f}}_{\mathrm{PI}}$ being by far the larger term of the two. If we start the simulation with a phosphorus interstitial, we have effectively already supplied the enthalpy to create the PI complex, and we can expect diffusion measured from such a starting point to correspond more or less to the migration energy $E^{\mathrm{m}}_{\mathrm{PI}}$. More or less because in molecular dynamics, just like in experiment, different diffusion mechanisms are usually at the play at the same time. It may also happen during any given simulation that the complete activation energy E_{PI} is supplied, so that phosphorus, after having returned to a lattice site, forms a new complex and diffuses further. This may complicate the process of describing the diffusivity using a simple Arrhenius form (1.4).

Two quite recent studies have attempted to estimate the migration energy using *ab initio* calculations and the nudged elastic band method. Liu *et al.* (2003) found a migration energy of 0.6–0.7 eV for a mechanism where P migrates between split $\langle 110 \rangle$ sites through hexagonal sites.[28] Harrison *et al.* (2006) found a lower barrier of 0.18 eV for a different mechanism where P diffuses between what they call “bond-centred” sites, through another site that was not named.[29]

2. Density Functional Theory

This chapter consists of two parts. The short sections 2.1–2.9 give a general presentation of density functional theory (DFT), while the lengthy section 2.10 is concerned with the practical choices made in the application of the theory in this specific work.

The aim of the first part has not been to *introduce* DFT,[1] but rather to *describe* or summarise important concepts in the theory (sec. 2.1–2.4) as implemented in `VASP` (sec. 2.5–2.9) in a somewhat superficial way, and since notation varies between different authors, to provide a consistent notation to be used when referring to concepts later on. Readers experienced with DFT may safely jump directly to the second part (section 2.10), or even directly to the summary subsection 2.10.5 if in a rush.

2.1. Describing materials

As touched upon in section 1.2, matter – or materials, "matter from which a thing is or can be made"[2] – can be described in different pictures. With theories such as density functional theory, materials are pictured as *nuclei* and *electrons*. From a knowledge of the constituent nuclei and electrons, all properties of a material can in principle be calculated to the accuracy of interest, limited only by the available computational power (this is a most severe limitation though). And unless radioactivity is of interest, there is no need to worry about what happens within the nuclei (not to say what might happen within the electrons, if anything happens there).

While nuclei can be described well using classical mechanics in most situations, electrons require a quantum mechanical description, as their de Broglie wavelength is often larger than the average inter-particle separation.[3]

[1] A good practical introduction is given by Scholl(2009)[35], while a more theoretical and authoritative introduction is given by Parr(1994)[36]. Chapter 12 in Kantorovich(2004)[37] is somewhere in the middle, and is a personal favourite of mine. The comprehensive book by Martin(2004)[38] and the more compact book by Kohanoff(2006)[39] are good reference books that put DFT in context of other methods, but they may be a bit daunting on their own. Chemists may find chapter 8 in Cramer(2004)[40] a good read. This chapter borrows from all the aforementioned books.

[2] New Oxford American Dictionary 2nd ed. 2005. I will not delve into the definition of 'thing'.

[3] The de broglie wavelength is $\lambda = h/p$ after de Broglie's 1924 hypothesis of the wave properties of matter.[41] A (classical) particle with mass m has momentum $p = \sqrt{2mE}$ and thermal energy $E = k_B T$, giving a "thermal de Broglie wavelength"[42, p. 114] of $\lambda = h/\sqrt{2mk_BT} = 18\,\text{Å}/\sqrt{A_r T}$, where $A_r = m/m_u$ is the (dimensionless) relative atomic mass with $m_u \approx 1.66 \times 10^{-27}$ kg being the atomic mass constant, and T is the temperature in Kelvin. For an electron, $A_r = 1/1800$, and the thermal de Broglie wavelength at room temperature is about 77 Å, much larger than the expected inter-particle separation. For a Si-atom, on the other hand, $A_r = 28$ and λ 0.2 Å. It should be noted that for the lightest elements, quantum effects such as tunneling *do* have practical effects e.g. on the

In quantum mechanics, a system of particles is fully described by the concept of a (time-dependent) quantum state $|\Psi(t)\rangle$, or its representation in position-space; a many-body wavefunction $\Psi(\mathbf{X}, t) = \langle \mathbf{X}|\Psi(t)\rangle$, with $\mathbf{X} = \{\mathbf{x}_i, \mathbf{R}_k\}$ being the set of all electron spin coordinates $\mathbf{x}_i$ and all nuclei coordinates $\mathbf{R}_k$. The state (or states) with the lowest energy is called the *ground state* (GS), while the others are called excited states. Observables like energy, momentum, *et cætera* are extracted by letting the appropriate Hermitian linear operator act on the state. For instance, the energy is the expectation value of an energy operator, a *Hamiltonian* $\hat{H}$, consisting of operators for the kinetic energy[4] and various potential energy terms. The (non-relativistic) Hamiltonian for a system of N_e electrons and N_N nuclei, in Hartree atomic units[5], is

$$\hat{H} = \hat{K}_N + \hat{V}_{NN} + \hat{K} + \hat{V}_{ee} + \hat{V}_{Ne} \tag{2.1}$$

$$= \underbrace{-\frac{1}{2}\sum_{k=1}^{N_N}\frac{1}{m_k}\nabla_k^2 + \sum_{k<l}\frac{Z_k Z_l}{r_{kl}}}_{\text{nuclei}} \underbrace{-\frac{1}{2}\sum_{i=1}^{N_e}\frac{1}{m_e}\nabla_i^2 + \sum_{i<j}\frac{\mathrm{e}^2}{r_{ij}}}_{\text{electrons}} + \sum_{i=1}^{N_e}\sum_{k=1}^{N_N}\frac{Z_k}{r_{ik}} \tag{2.2}$$

with the five terms representing nucleus kinetic energy $\hat{K}_N$, nucleus-nucleus interactions $\hat{V}_{NN}$, electron kinetic energy $\hat{K}$, electron-electron interactions $\hat{V}_{ee}$ and electron-nucleus interactions $\hat{V}_{Ne}$, respectively. The indices i and j run over all electrons, while k and l run over all nuclei.

As seen, only electromagnetic (Coulombic) interaction is included. The particles also interact gravitationally, but this can safely be ignored since the gravitational force is so much weaker than electromagnetic one.[6] At the atomic scale, gravitation is completely negligible. But other terms could be included in the Hamiltonian depending on the problem at hand, such as interaction with external electric or magnetic fields, spin coupling terms, et cetera.

2.2. The many-particle problem and the Born Oppenheimer approximation

Our systems of interest consist of large number of nuclei and electrons, particles whose motion are correlated by their Coulombic interactions. While there is no known way to describe the mutual interaction of many particles in general, the reductionist principle of physics leads to a description in terms of a sum of pairwise interactions, for which

kinetics of reactions, but we are in the luck of working with no element lighter than Si, so we don't have to worry about this as a non-negligible source of error.

[4] Since quantum mechanical particles do not follow *trajectories* in the way classical particles do, the name 'kinetic energy' is a bit misleading, Reflecting the spatial variation of the wavefunction, Sauer and Kuhn suggested the name 'localisation energy' instead[43]. Yet, there is an analogy with classical kinetic energy, and with 'kinetic energy' being the dominating expression, it will be used here too.

[5] where the electron mass m_e, the elementary charge e, the reduced Planck constant $\hbar$ and the Coulomb constant $k_e = 4\pi/\varepsilon_0$ all are unity by definition.

[6] Remember that a small kitchen magnet counteracts the gravitational pull of the entire Earth.

simple expressions exist.[7] With such an approach, equations describing the motion of a general system of many particles can be set up, with 'many' now meaning 'more than two', but the correlated equations cannot be solved 'exactly' (analytically), no matter which specific equations are used for the pairwise interactions. This is the *many-particle problem* or *many-body problem*, perhaps first encountered by Newton as he studied the three-body problem of the motion of the Earth-Moon-Sun system interacting gravitationally.[8] Now, approximate solutions to the Earth-Moon-Sun system can be found quite easily by noting that the problem can be decoupled into the Earth-Sun and the Earth-Moon problems, due to the large mass difference between the three 'particles'.

Similarly, the quantum-mechanical two-body problem of a single electron interacting with a single nuclei is solvable, while the introduction of more particles makes approximate numerical solutions necessary. And with the electron being 1800 times lighter than the lightest nucleus, the hydrogen nucleus, approximate solutions may be found by decoupling the motion of the nuclei and the electrons. This is the central idea of the Born Oppenheimer approximation[45].

In the Born-Oppenheimer approximation, we assume that the full wavefunction $\Psi(\mathbf{R}, \mathbf{x}, t)$ can be decoupled into a nuclei wavefunction $\Theta_n(\mathbf{R}, t)$ and an electronic wavefunction $\Phi(\mathbf{x}; \mathbf{R})$ depending only parametrically on $\mathbf{R}$ (indicated following a semicolon here, but usually left out)

$$\Psi(\mathbf{R}, \mathbf{x}, t) = \Theta(\mathbf{R}, t)\Phi(\mathbf{x}; \mathbf{R}) \tag{2.3}$$

and the *electronic structure problem* is now to find solutions Φ to the Hamiltonian,

$$\hat{H}_\mathrm{e} = \hat{K} + \hat{V}_\mathrm{ee} + \hat{V}_\mathrm{Ne} \tag{2.4}$$

The expectation value of this Hamiltonian plus the (constant) nuclear-nuclear repulsion energy V_NN is the *electronic energy*,[40, p. 110] a parametric function of the nuclear coordinates called the potential energy surface. This parametric dependence on the nuclear coordinates amounts to the electrons adjusting their positions $\mathbf{x}$ *instantaneously* to a given *nuclei structure*. The Born-Oppenheimer approximation is fundamental in most electronic structure methods, including DFT, and further references to the Hamiltonian in this chapter will implicitly be to the *electronic* Hamiltonian unless otherwise stated. We will return to the nuclei structure in chapter 3.

While the Born-Oppenheimer approximation is justified in most cases,[40, p. 111] it's important to keep in mind the limitations it impose. A phenomena such as Joule heating, for instance, utilised in everyday electrical heating appliances, can not be described using the approximation, since it involves electrons dissipating energy into ionic vibrations, coupling the motion of the electrons and the nuclei.[46]

[7] Two-particle interactions are often sufficient for relatively weak forces such as gravitational forces or Coulomb forces, while for strong interactions the inclusion of three-particle interactions is hardly avoidable.[44]

[8] Principia, Book 1, section XI, Proposition 66 and its corollaries, http://en.wikisource.org/wiki/The_Mathematical_Principles_of_Natural_Philosophy_(1729)/Book_1/Section_11#Prop66 http://books.google.com/books?id=ySYULc7VEwsC&pg=PA173 (Newton, 1687, english transl. by Motte 1729) The problem received great attention, as exemplified by the two-volume work La Théorie du mouvement de la lune, 1800 pages in length, published in1860 and 1867 by Charles-Eugène Delaunay on the system.

2.3. Electron density as the fundamental variable

In density functional theory (DFT), the (observable) electron density $\rho(\mathbf{r})$ rather than the (non-observable) quantum state $|\Phi\rangle$ is considered the fundamental entity for the electronic structure problem, and other observables are extracted from functionals of the electron density $\rho(\mathbf{r})$ rather than from operators acting on the state. This approach was first tried by Thomas[47] and Fermi[48] in the late 1920s, and formally justified in 1964 by Hohenberg and Kohn (HK), who proved that a degenerate ground state density determines the external potential and thus the Hamiltonian (first HK theorem), and also that the energy functional $E[\rho(\mathbf{r})]$ is variational with respect to the density (second HK theorem).[49] Later, Levy generalised their proofs to also include degenerate ground states, and to allow a larger set of trial densities.[9][50, 51]

The energy functional consists of the same three contributions as the Hamiltonian (2.4); with $\hat{V}_{\text{ee}}(\{\mathbf{r}\}) \leftrightarrow E_{\text{int}}[\rho(\mathbf{r})]$, the potential energy arising from the 'internal' electronic field, and $V_{\text{Ne}}(\{\mathbf{r}\}) \leftrightarrow E_{\text{ext}}[\rho(\mathbf{r})]$, the potential energy arising from the 'external field' set up by the nuclei (and in general other external fields).

The energy functional can then be written

$$E[\rho(\mathbf{r})] = K[\rho(\mathbf{r})] + E_{\text{int}}[\rho(\mathbf{r})] + E_{\text{ext}}[\rho(\mathbf{r})]. \tag{2.5}$$

With the nuclei at fixed positions $\{\mathbf{R}_k\}$, the potential energy functional for the electron density $\rho(\mathbf{r})$ is just

$$V_{\text{ext}}(\mathbf{r}) = \sum_{k=1}^{M} \frac{Z_k}{|\mathbf{r} - \mathbf{R}_k|}, \qquad \text{and then} \qquad E_{\text{ext}}[\rho(\mathbf{r})] = \int \rho(\mathbf{r}) V_{\text{ext}}(\mathbf{r})\, d\mathbf{r}. \tag{2.6}$$

where Z_k is the atomic number

The repulsive potential between the electrons V_{int} is more difficult to describe in a simple manner. A simple first approximation is the classical electrostatic interaction described by the Hartree potential,

$$V_{\text{H}}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \qquad \text{and then} \qquad E_{\text{H}}[\rho(\mathbf{r})] = \int \rho(\mathbf{r}) V_{\text{H}}(\mathbf{r}) d\mathbf{r}, \tag{2.7}$$

Since the potential depends on the density, this is a self-consistent field problem, but the main problem is that the potential does not take account of quantum mechanical exchange and correlation, and that it includes self-interaction. For these reasons, approximating $E_{\text{int}}[\rho] \approx E_{\text{H}}[\rho]$ is considered a very poor approximation,[52, p. 208] and better approximations are discussed in the next two sections below.

But first, how is the kinetic energy to be found from an electron density ($K[\rho]$)? That's not a trivial problem, and the most successful solution involves the re-introduction of a one-particle formalism, in what is known as the Kohn-Sham method.

[9] by only requiring densities to be N_e-representable, not necessarily V-representable

2.4. The Kohn-Sham self-consistent field method

The *ansatz* in the Kohn-Sham (KS) method[53] is that a system of correlated electrons can be described by an auxillary system of independent electrons (or formally *quasielectrons*) moving in a potential altered to make the density *the same* for both systems. Since the density determines all properties of the full many-body system according to the Hohenberg-Kohn (HK) theorems, an exact solution to the independent-particle problem should yield an exact solution to the correlated-particle problem. This is represented schematically in figure 2.1.

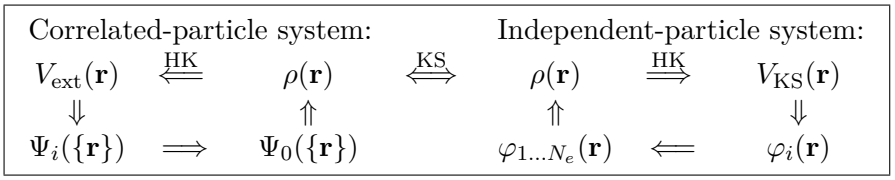


Figure 2.1.: Schematic representation of the Kohn-Sham *ansatz*. Adopted from [38, p. 137].

The unlabelled arrows in the left part indicate the usual solution of the Schrödinger equation of a correlated many-electron system, in which the potential $V_{\text{ext}}(\mathbf{r})$ determines all the states of the system $\Psi_i(\{\mathbf{r}\})$, including the ground state $\Psi_0(\mathbf{r})$ with the corresponding density $\rho(\mathbf{r})$.[10] The Hohenberg-Kohn theorem completes the circle by proving that the ground state density $\rho(\mathbf{r})$ determines $V_{\text{ext}}(\mathbf{r})$. Note that in theory the ground state density determines excited states as well as the ground state, but practical methods for calculating properties of excited states still represents a big challenge.[38, p. 136] We will, however, only be concerned with the ground state here.

While the left part of the figure represents the "real" correlated many-body problem, the right part is the fictitious independent-particle system introduced by Kohn and Sham. If $\{\varphi_i\}$ is the set of the lowest N_e quasielectron orbitals, the ground state density is[11]

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_e} |\varphi_i(\mathbf{r})|^2, \tag{2.8}$$

Now, the kinetic energy of these quasielectrons is just

$$K_0[\rho] = \sum_{i=1}^{N_e} \langle \varphi_i | \hat{t} | \varphi_i \rangle, \qquad \text{where} \qquad \hat{t} = -\frac{1}{2}\nabla^2 \tag{2.9}$$

is the single-particle kinetic energy operator.

Algebraically we can rewrite the energy functional (2.5),

$$\begin{aligned} E[\rho] &= K[\rho] + E_{\text{int}}[\rho] + E_{\text{ext}}[\rho] && (2.10)\\ &= K_0[\rho] + E_{\text{H}}[\rho] + E_{\text{ext}}[\rho] + (K[\rho] - K_0[\rho]) + (E_{\text{int}}[\rho] - E_{\text{H}}[\rho]) && (2.11)\\ &= K_0[\rho] + E_{\text{H}}[\rho] + E_{\text{ext}}[\rho] + E_{\text{xc}}[\rho], && (2.12) \end{aligned}$$

[10] An index 0 is omitted since ρ will always refer to the ground state density

[11] Single-electron orbitals are assumed here. If closed-shell orbitals are used, the sum is multiplied by two and upper-limited by $N_e/2$. The same goes for the sums below.

where all the 'difficulties' are now collected in the exchange-correlation (xc) functional,

$$E_{\mathrm{xc}}[\rho] = (E_{\mathrm{int}}[\rho] - E_{\mathrm{H}}[\rho]) + (K[\rho] - K_0[\rho]), \tag{2.13}$$

to which we will return in the next section. Note that the exchange-correlation functional includes not only quantum mechanical exchange and correlation, but also corrections for the classical self-interaction in the Hartree potential, and for the difference in kinetic energy between the non-interacting system and the real one; $K - K_0$. The name is therefore really a misnomer.[40, p. 256]

The quasielectrons[12] move in an effective potential

$$V_{\mathrm{KS}}(\mathbf{r}, [\rho]) = V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{ext}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r}), \qquad \text{where} \qquad V_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[\rho]}{\delta \rho} \tag{2.14}$$

is not known, and we can find the quasielectron orbitals from the set of equations

$$\left[\hat{t} + V_{\mathrm{KS}}(\mathbf{r}, [\rho])\right] \varphi_n(\mathbf{r}) = \varepsilon_n \varphi_n(\mathbf{r}) \tag{2.15}$$

and the corresponding energy from (2.8) and (2.12). Note that the total energy is not just the sum of the quasielectron eigenergies ε_n. These are self-consistent field problems requiring iterative solutions, since $V_{\mathrm{KS}}(\mathbf{r}, [\rho])$ depends on the density.

2.5. The exchange-correlation potential

The Kohn-Sham scheme provides a neat solution to the problem of defining a kinetic energy functional, and the mapping of the of interacting electrons onto non-interacting quasielectrons is exact in the limit of an exact effective potential $V_{\mathrm{KS}}(\mathbf{r})$. The remaining problem is to provide a good approximation to $V_{\mathrm{KS}}(\mathbf{r})$, or rather to $V_{\mathrm{xc}}(\mathbf{r})$, in which all the difficulties are collected, and this problem remains the top challenge for the DFT method.[37, p. 527][36, p. 153]

While the exchange-correlation potential is inherently non-local, a popular starting point is a local approximation known as the local density approximation (LDA), as taken by Kohn and Sham themselves in their 1965 paper.[53] LDA utilizes the well-understood behaviour of the *uniform* gas.[13] If $\varepsilon_{\mathrm{xc}}^{\mathrm{uni}}[\rho]$ is the exchange-correlation functional for a uniform electron gas of density ρ, LDA corresponds to evaluating this functional for each volume element $d\mathbf{r}$;

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho] = \int \varepsilon_{\mathrm{xc}}^{\mathrm{uni}}[\rho]\rho(\mathbf{r})\, d\mathbf{r} \tag{2.16}$$

[12] At some point I may forget to include 'quasi-' and just write 'electrons'. To avoid any confusion, Kohn-Sham DFT is *only* concerned with quasielectrons, so even if omitted, 'quasi-' should still really be there.

[13] The exact local exchange is given by the Dirac exchange functional $\varepsilon_{\mathrm{x}}[\rho] = C_x \rho^{1/3}$, with $C_x = -3/4(3\rho/\pi)^{1/3}$,[54][36, p. 108] while correlation has been tabulated with increasing levels of accuracy, with the currently most accurate results found from quantum monte carlo (QMC) calculations.[38, p. 109 and references 297-299 within]

Such an approach is expected to work well only for sufficiently slowly varying densities ("uniform-like"), but in addition it has turned out that in many cases there are fortunate cancellation of errors resulting in better-than-expected results.[55][37, p. 529][36, p. 180] It's hard to come up with a better local approximation than LDA, at least one that is somewhat intuitive since the effects that are to be described are inherently non-local, so to improve on LDA the logical next step is to move to a gradient approximation, including some dependency on the density gradient $\nabla\rho$, but exactly how to do this is far from straightforward and different and many different functionals have evolved over the years.

In this work we will make use of the PBE-GGA functional introduced by Perdew, Beckge and Eizenhover in 1996,[56] where GGA is short for generalised gradient approximation, Perdew and Yue's contrasting term to the earlier gradient expansion approximation (GEA).[57] Compared to other functionals PBE-GGA has a relatively simple form, satisfying what the authors consider the most important conditions, and introduced as an alternative to what the authors called "the Byzantine PW91" functional in a paper titled "GGA made simple". The other defining feature of PBE-GGA is that is non-empirical, making it a *general-purpose* functional in contrast to 'semi-empirical' functionals like B3LYP or revPBE that provide better results for *specific* problems. PBE-GGA can be sketched as

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = \int \varepsilon_{\mathrm{xc}}^{\mathrm{uni}}[\rho]\rho(\mathbf{r})F_{\mathrm{xc}}(s)\, d\mathbf{r}, \qquad \text{where} \qquad s = \frac{|\nabla\rho|}{2k_F\rho} \tag{2.17}$$

is the first-order reduced density gradient, $k_F = (3\pi^2\rho)^{1/3}$ is the Fermi vector, and $F_{\mathrm{xc}}(s)$ is some dimensionless function called the enhancement factor chosen to fulfil a set of conditions. $F(0) = 1$, so the functional reduces to LDA for $s = 0$. See [57], [56] and [38, p. 154] for details.

2.6. Pseudopotentials

So far, two important approximations have been introduced; the Born Oppenheimer approximation and the approximate handling of exchange and correlation using the PBE-GGA functional. Still, calculations using a plane wave basis, except for very small systems, would not be feasible without one more important approximation.

When expanding the one-particle orbitals in a plane wave basis, (2.20), it is evident that a much larger amount of plane waves must be included to accurately describe the more localised core orbitals than the less localised valence orbitals. Further, the core orbitals for a givent atom resemble each other when the atom is in different environments, since it's the valence orbitals that take part in chemical binding and most interesting material properties.

The basic idea of the pseudopotential approach is to replace the inner orbitals with a fixed potential, and only threat the valence orbitals variationally. For the element of silicon for instance, the common pseudopotential approach is to threat only the four valence electrons variationally, leaving the remaining ten core electrons to some fixed

pseudopotential. Of course all kinds of practical difficulties may arise, such as the question of which electrons to include in the core, and how to actually do the partitioning without introducing unphysical effects. Several different methods have therefore evolved over the years. Here, the projector augmented-wave (PAW) method[58] is used, which has the elegant feature that the wavefunction in the core region can be fully reconstructed. For more information, see e.g. the chapter on pseudopotentials in Martin(2004)[38, pp. 204-299].

The specific pseudopotentials usd in this work were provided by Georg Kresse and are referred to as "very soft"[59] (not to be confused with "ultrasoft" potentials which are not based on the PAW method). They are 'softer' than the potentials included in the standard `VASP` distribution,[60] since they have 26% larger core radius.[14]

2.7. Bloch states and the plane wave basis

A few words must be said about the orbitals φ_n used in the Kohn-Sham equations (2.15). In a periodic system, they should be Bloch states. In general for some reciprocal vector $\mathbf{k}$,

$$\varphi_{n\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \tag{2.18}$$

where $u_{n\mathbf{k}}(\mathbf{r})$ is a function that shares the periodicity of the lattice, $u_k(\mathbf{r}+\mathbf{G}) = u_k(\mathbf{r})$, where $\mathbf{G} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ and $\mathbf{k}$ is any reciprocal space vector.

In the `VASP` code used in this work, a plane wave basis is used. Such a basis make use of the fact that a periodic function can always be represented by a Fourier series,

$$u_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n\mathbf{k}\mathbf{G}} \mathrm{e}^{\mathrm{i}\,\mathbf{G}\cdot\mathbf{r}} \tag{2.19}$$

making (2.18)

$$\varphi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n\mathbf{k}\mathbf{G}} \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{2.20}$$

This makes for a well-defined basis, exact in the limit of all $\mathbf{G}$, and just as important, one that can be systematically truncated by including only components with

$$\frac{|\mathbf{G}+\mathbf{k}|^2}{2} \leq E_{\mathrm{cut}} \tag{2.21}$$

for some scalar cutoff energy E_{cut}. Note that leaving out long reciprocal space vectors $\mathbf{G}$ corresponds to leaving out short real space vectors $\mathbf{r}$. Introducing the cutoff is therefore associated with discretising real space.

Why plane waves? One advantage is how effectively they can be transformed between real and reciprocal space using Fast Fourier Transform (FFT), allowing each part of the Hamiltonian to be evaluated in the most appropriate space. In particular, the kinetic

[14] The verysoft potentials have RCORE=2.4. Since no versioning is used, for identification, the SHA-1 of the Si and P potentials are cf786bc703f03a562aae3661fa6cc65f86bc3097 and 8058b1b1a564e2d79696e3c30bdac2d44ff524a0, respectively.

term of the Hamiltonian is diagonal in reciprocal space when using a PW basis. Another advantage of a PW basis is that the Hellman-Feynman force theorem can easily be applied since the basis functions don't depend on the nuclear coordinates (see section 2.9). Third, the PWs (and their derivatives) are quite simple to evaluate analytically.[39, p. 187]

2.8. Brillouin zone integration and the k point density

For periodic systems, calculations of observables generally involves integrals over the Brillouin zone[38, p. 89][15]. For example, we may compute the total energy per unit cell as[61, p. 24]

$$E_{\text{tot}} = \frac{1}{\Omega}\int_{\text{BZ}} E(\mathbf{k})\, d\mathbf{k} \tag{2.22}$$

where Ω is the Brillouin zone volume and $E(\mathbf{k})$ is the total energy found from solving the Kohn Sham equations for a given $\mathbf{k}$. For computational evaluation such integrals have to be turned into sums over a discrete set of k-points. In 1973 Baldereschi suggested a first approximation could be made by including only a single k-point,

$$E_{\text{tot}} = \frac{(2\pi)^3}{\Omega}\bar{E}, \qquad \bar{E} = E_{\mathbf{k}^*} \tag{2.23}$$

for a 'mean value point' $\mathbf{k}^*$ determined by crystal symmetry[62]. This idea has later been generalised into sets of optimal (so-called 'special') k points by Chadi and Cohen[63], Monkhorst and Pack[64] and others. Including only points in the irreducible Brillouin zone (IBZ) results in a sum over points with weights $w_{\mathbf{k}^*}$;

$$E_{\text{tot}} = \frac{(2\pi)^3}{\Omega}\sum_{\mathbf{k}^*} w_{\mathbf{k}^*} E(\mathbf{k}^*) \tag{2.24}$$

When supercells are used, as will be done here, the k point density can be reduced, since a larger supercell (real space) corresponds to a smaller Brillouin zone (reciprocal space), but the benefit of fewer k points is cancelled by the increased number of bands. This is illustrated, for the sake of simplicity, for a one-dimensional lithium chain with bond length 2.38 Å in figure 2.2.[16] Since the atoms are evenly spaced, the system is periodic, and we can define a unit cell with lattice parameter $a = 2.38$ Å containing a single atom. The brillouin zone is then a line segment of length $2\pi/a$. The left part of figure 2.2 shows the energy bands sampled at 17 k points between $-\pi/a$ and π/a.[17]

Now, we could define a cell with lattice parameter $a' = 2a = 4.76$ Å containing two atoms. The brillouin zone is then reduced to a line segment of length $2\pi/a' = \pi/a$,

[15] The term 'Brillouin zone' will be used throughout this work as a shorthand for 'first Brillouin zone' as no other Brillouin zones will be encountered.

[16] I really wanted to see such an illustration myself, but for some reason I didn't find one in any textbook I've seen, which is one reason why I include it here even though it's slightly off-topic.

[17] Of course there is some symmetry here. Really, we would only have to sample the irreducible brillouin zone of length π/a, but the whole Brillouin zone was sampled here for the purpose of illustration.

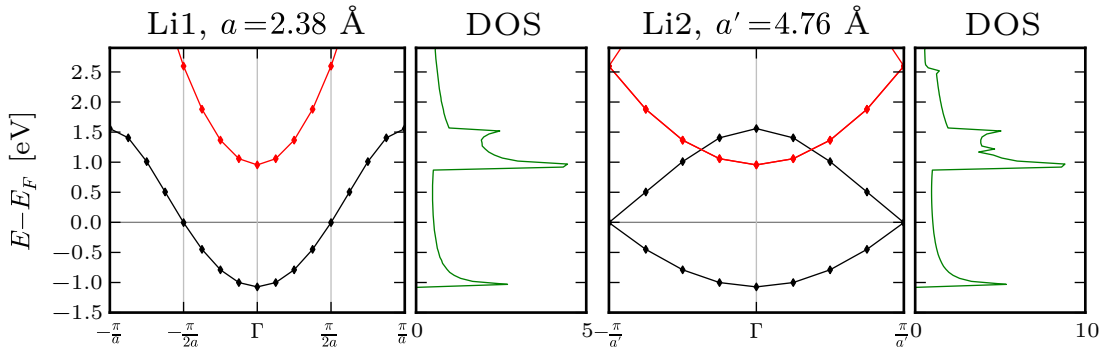


Figure 2.2.: Illustration of the effect of using supercells in one dimension. The left figure shows the band structure calculated from a 'unit cell' with lattice parameter 2.38 Å containing a single Li atom. The right figure shows the band structure calculated from a 'supercell' with lattice parameter 4.76 Å containing two Li atoms.

but should still contain the whole band structure, since the choice of unit cell size is arbitrary (as long as translational symmetry is retained). In the right part of the figure it is shown how the lowermost band "wraps back" into the first brillouin zone. Even though the number of k points used for the 'double cell' (right) is half the number used in the 'single cell' (left), the distance between k points are the same; $\pi/(8a) \approx 0.165\,\text{Å}^{-1}$ in both cases.

2.9. Forces from the Hellman-Feynman force theorem

For the dynamics discussed in the next chapter, a method to calculate the *force* acting on each atom is also needed. If we take the nuclei to be classical point particles, the force $\mathbf{F}_k$ acting on a nucleus k at position $\mathbf{R}_k$ is determined from the potential energy

$$\mathbf{F}_k = -\frac{\partial E}{\partial \mathbf{R}_k} \tag{2.25}$$

The energy E here is the sum of the expectation value of the electronic Hamiltonian $\langle\Phi|\hat{H}_e|\Phi\rangle$ and the nuclear-nuclear interactions E_{NN}. In both cases we are interested in finding the derivative of E with respect to some perturbation λ, for instance a component of $\mathbf{R}_k$,

$$\frac{\partial E}{\partial \lambda} = \left\langle \frac{\partial \Phi}{\partial \lambda} \middle| \hat{H} \middle| \Phi \right\rangle + \left\langle \Phi \middle| \hat{H} \middle| \frac{\partial \Phi}{\partial \lambda} \right\rangle + \left\langle \Phi \middle| \frac{\partial \hat{H}}{\partial \lambda} \middle| \Phi \right\rangle + \frac{\partial E_{\text{NN}}}{\partial \lambda} \tag{2.26}$$

If $|\Phi\rangle$ is an eigenstate of $\hat{H}$ with energy E_e, the first two terms become

$$E_e \left(\left\langle \frac{\partial \Phi}{\partial \lambda} \middle| \Phi \right\rangle + \left\langle \Phi \middle| \frac{\partial \Phi}{\partial \lambda} \right\rangle \right) = E_e \frac{\partial}{\partial \lambda} \langle \Phi | \Phi \rangle = E_e \frac{\partial}{\partial \lambda}(1) = 0 \tag{2.27}$$

using the normalisation $\langle\Phi|\Phi\rangle = 1$. Thus the force is

$$\mathbf{F}_k = -\left\langle \Phi \middle| \frac{\partial \hat{H}_e}{\partial \mathbf{R}_k} \middle| \Phi \right\rangle - \frac{\partial E_{\mathrm{NN}}}{\partial \mathbf{R}_k} \tag{2.28}$$

This is often referred to as the Hellmann-Feynman theorem after Hellmann[65] and Feynman[66], but also as the 'force theorem'.[38, p. 56] The theorem is also used when calculating the stress tensor, where $\lambda = \varepsilon_{\alpha\beta}$, a component of the strain tensor.

We assumed above that $|\Phi\rangle$ is an exact eigenstate of $\hat{H}$, but when $|\Phi\rangle$ is expanded in a finite basis set, it is generally not so. However, it turns out that as far as the basis functions do not depend on the nuclei coordinates (plane waves do not), (2.28) is still valid.[see e.g. 37, p. 557]

A further simplification arises from the fact that the only term in the electronic Hamiltonian that depends on the nuclei coordinates is the term describing electron-ion interaction:

$$\mathbf{F}_k = -\int \frac{\partial V_{\mathrm{ext}}(\mathbf{r})}{\partial \mathbf{R}_k} \rho(\mathbf{r})\, d\mathbf{r} - \frac{\partial E_{\mathrm{NN}}}{\partial \mathbf{R}_k} \tag{2.29}$$

where the first term is the attraction to the (unperturbed) electronic density $\rho(\mathbf{r})$, and the second term the repulsion to the other nuclei (which are kept fixed).

Note that the final simplification assumes $V_{\mathrm{ext}}(\mathbf{r})$ to be a local potential. In the case of non-local pseudopotenstials, (2.29) can not be used, but more complicated expressions can be derived from (2.28), which is still valid.[38, p. 57]

2.10. Choice of parameters

As described in chapter 3 on Born Oppenheimer molecular dynamics below, the purpose of using density functional theory is to accurately calculate the forces between atoms. From a knowledge of the forces, the dynamics of the atoms can be simulated using simple newtonian dynamics. Since atoms vibrate at frequencies in the order of 10^{12} Hz (see fig. 1.2), forces will have to be re-calculated in the order of once every femtosecond to produce trajectories that are smooth and hopefully, realistic simulations of Nature.

With the time step more or less fixed, the simulation length is then limited mainly by the time it takes to carry out each force-calculation. If each force-calculation takes about 1 CPU second to carry out, and we re-calculate the forces every 1 fs, simulating 1 ns takes 10^6 CPU seconds or 278 CPU hours, or if parallelised over 8 CPU cores, 35 clock hours.[18] It's therefore fundamentally important to optimize the force calculation as much as possible, which is the topic of most of this section on "choice of parameters". Our system must be as simple as possible, yet behave as similarly as possible to the real physical system we are trying to describe. The art of computer simulation is to simplify as much as possible, yet not too much. However, we will start discussing an

[18] If we could just simulate 1 ns in "less" than 1 ns we could actually predict the future, but such simulation speeds are apparently far into the future themselves.

important parameter that does not affect the calculation time; the lattice parameter of the simulation cell.

2.10.1. Lattice constant

Choosing a too small lattice parameter a will add an artificial positive external pressure on the simulation cell, while choosing it too large will add a negative external pressure (corresponding to an isotropic pull). In figure 2.3, the pressure and total energy of the system is plotted as a function of the lattice constant, varied in steps of 0.01 Å. The pressure is calculated as $P = \frac{1}{3}\sum_\alpha \sigma_{\alpha\alpha}$ where σ is the stress tensor. A realistic lattice parameter is one that zero out the external pressure,[19] and following the variational principle, that minimizes the total energy. The lattice parameter fulfilling such conditions is indicated by a vertical line for each set. The first two sets make use of fully converged parameters ('standard' potentials, 125 k points in the Brillouin zone[20] and plain wave energy cutoff of 800 eV), while the last set is discussed below.

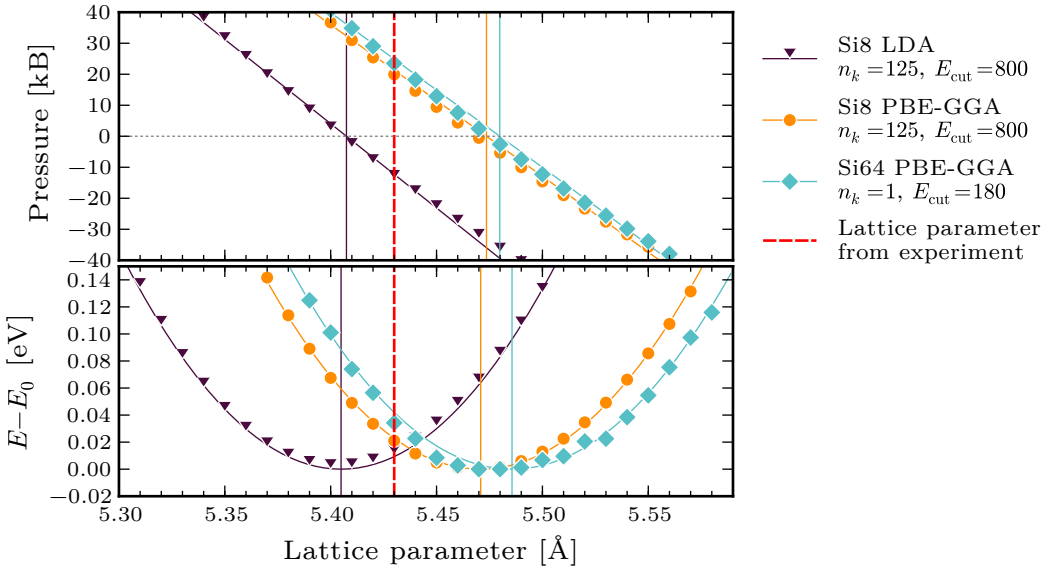


Figure 2.3.: Total energy and pressure as a function of varying the lattice parameter a of a 8-atom cubic unit cell. For the 64-atom cell, the lattice parameter is $2a$, and the energy plotted is the cube root of the total energy, for comparison with the 8-atom cell. All energies are given relative to the minimum energy E_0 for each configuration. The absolute minimum energies of the different configurations are not the same.

The first to notice about the figure is that even for fully converged parameters, using LDA results in a lattice parameter 0.46 % (2.5 pm) shorter than the experimental value of

[19] Ideally one that makes the pressure equal the standard atmosphere of one bar, but such a small pressure is not really distinguishable from zero bar within the precision we work with. The scale on fig 2.3 should make this clear.

[20] although only 10 are actually used for calculation due to symmetry

5.43 Å[21] (indicated with a thick, dashed line), while using PBE-GGA results in it being 0.75 % (4.1 pm) too long. This is the well-known tendency of LDA to overbind, and the rather common tendency of PBE-GGA to underbind. Now, in this particular case of pure silicon, it turns out that the LDA-value comes closer to the experimental value than the PBE-GGA value, in line with the results of Haas et al.[69][22] This is most likely due to fortunate cancellation of errors,[see e.g. 38, p. 167] and since the purpose here is not to calculate accurate lattice constants, we will stay with the theoretically more robust PBE-GGA as it would be rather fortenous to expect the same level of error cancellation for every different configuration visited during a simulation at elevated temperatures. Rather, we could expect the cancellation to work better for some configurations than others, and thus produce non-systematic errors.

Now, foreseeing the results of the next sections, we will look quickly into the effects of the approximations decided upon in those sections. First, replacing the 'standard' potential with the 'very soft' potential (see section 2.6) lead to a 0.15 % (0.8 pm) reduction in a. Then lowering the plane wave energy cutoff from 800 eV to 180 eV lead to a 0.09 % (0.5 pm) increase, largely cancelling the previous effect. Finally, reducing the k point density from 20.5 to $1.3 \times 10^3/\text{Å}^{-3}$ lead to a 0.29 % increase (1.6 pm). The plot for this final configuration is included in 2.3 as the third curve.

To sum up, a lattice constant of 5.47 Å will be used in the simulations. This is 0.74% off the experimental lattice constant of 5.43 Å, and since the Si-Si bond length is $(\sqrt{3}/4)a$ and thus proportional to a, the silicon bond length is also 0.74 % off. So while the experimental equilibrium Si-Si bond length is 2.352 Å, the equlibrium Si-Si bond length in our simulation cell is 2.369 Å. This should be kept in mind when bond lengths are discussed in the results section.

Now, we may ask what happens when we add temperature in the form of ionic motion. Should the lattice parameter then be increased? This question is not as easy to answer as it may seem. First, a solid heated to diffusion temperatures will usually require very long times to actually expand to its equilibrium volume, much longer than the timescale of the experiment itself, so it is not very realistic to expect the solid to be at its equilibrium volume. Second, we have no simple way of actually finding the volume expansion. Or do we?

[21] The lattice parameter of silicon is lattice parameter of silicon is 5.431 Å at room temperature (293.15 K),[67], but we should really compare with the lattice parameter at zero kelvin since the ions are at rest. Using the linear thermal expansion coefficient of $2.6 \times 10^{-6}\,\text{K}^{-1}$ at room temperature,[CITATION NEEDED] we can extrapolate to find a lower bound of 5.427 Å for the lattice parameter at zero kelvin. This is a lower bound since the thermal expansion coefficient itself is a function of temperature, and has been shown to decrease to zero near 120 K and even becoming slightly negative for even lower temperatures.[68] Experimental data are scarce for temperatures below 100 K, but if we are content with two decimals, we can take the experimental lattice parameter of silicon at zero kelvin to be 5.43 Å.

[22] Following Haas, some minor adjustments to the PBE functional actually make the functional calculate lattice parameters much closer to experimental values. On the other hand, these adjustments tended to produce worse atomisation energies. Since neither property is the interest of this work, we will stick to the very well-tested and theoretically robust standard PBE GGA functional.

2.10.2. k point density

As illustrated in section 2.8, the number of k points needed is inversely proportional to the cell size. For large supercells only a single k point, the Γ point, is needed, but we may well ask if the $a = 10.94\,\text{Å}$ cell ($2 \times 2 \times 2$) used in the simulations here is 'large enough' to justify including only a single k point, making the distance between k points $2\pi/a = 0.574\,\text{Å}^{-1}$.

Since the purpose of the density functional calculations here is to provide forces for *dynamics*, the effect of reducing the k point density was studied on a simple dynamics model. The model is a cell with 62 Si atoms, one P atom and one vacancy, and we calculate the energy as the P atom moves from its lattice position into the adjacent vacancy while the remaining Si atoms are kept still. Figure 2.4 shows the resulting force and total energy change as a function of the reaction coordinate of phosphorus, r. In order to move from its original lattice site into the vacancy it has to pass a barrier, the height of which is the activation energy for this model process. The activation energy for the converged configuration (5x5x5 k points) is 1.55 eV, while the activation energy increases to 1.61 eV for the Γ-only configuration. If the error of 3.8% in the activation energy of this model diffusion process is representable for other diffusion processes as well, the low k point density is expected to be a noticeable, but acceptably small source of error.

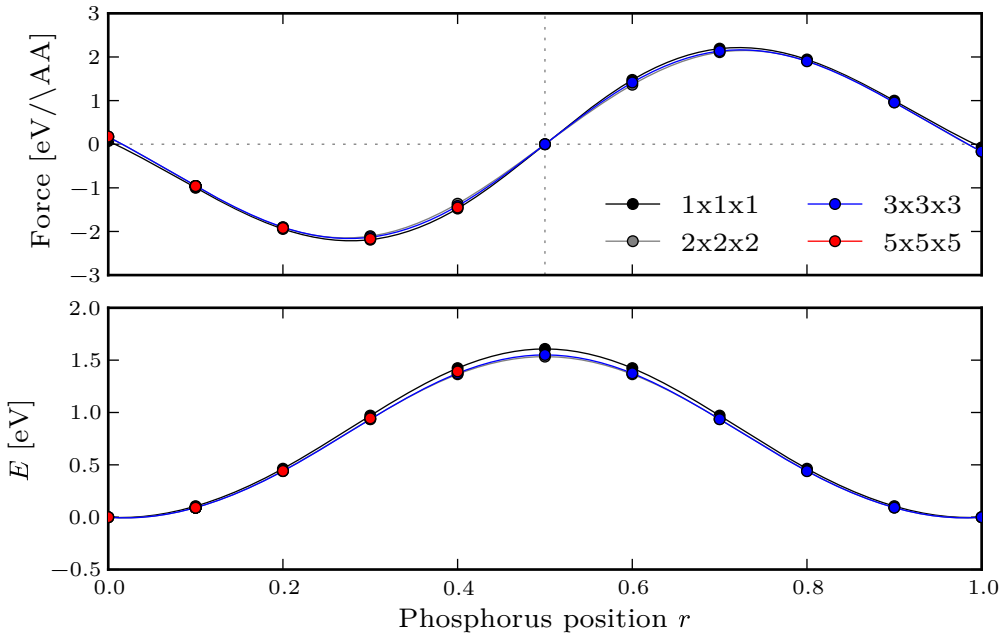


Figure 2.4.: Phosphorus atom moving in the [111] direction from its initial lattice site ($r = 0$) into a neighbour vacancy ($r = 1$), with all other atoms fixed. *Top*: Force acting on the phosphorus atom in the [111] direction. *Bottom*: Difference in the total energy between the system with phosphorus at r and at $r = 0$.

Why the Γ point and not some other k point? From the computational cost perspective, the Γ point is the best single k point to include, since it shows time-reversal symmetry, $\mathbf{k} = -\mathbf{k}$, allowing the computational cost to be halved. From an accuracy point of view the Γ point is generally not the optimum single k point to include when describing a *single* configuration, as the optimum point is dictated by the crystal symmetry of that configuration[62]. But in our MD simulations, the system should not be constrained by any specific symmetry, but rather be allowed to visit a large number of different configurations, most of will have no special point group symmetry. As such, there should be no *a priori* reason for the Γ point to be any worse a choice than any other $\mathbf{k}$ point. It is even likely to be less discriminate towards a specific symmetry than any other point.

If results are to be compared between different supercells, using only the Γ point may also be advantageous. As Castleton et al. point out[70], incomplete k-point convergence give rise to errors that in general vary with the supercell size in a 'non-rational' (non-simple) way, while such errors probably vary in a more 'rational' (simple) way for Γ point calculations. As a consequence, they recommend using either fully converged k-point sets or the Γ-point only when comparing supercells of different size.

2.10.3. Plane wave energy cutoff

While force-minimised high-symmetric configurations may be described well by a relatively small basis set, less symmetric and stressed configurations will require a larger one for a good description. During a MD simulation, the system will visit a large number of configurations of various symmetry and local stress. With a limited basis set, we expect some configurations to be better described than other. A too limited plane wave basis set may for example lead to large undulations in the electronic density being smoothed out, effectively making stressful configurations where the atoms are very close more favourable than they should be. This is just one imaginable basis set effect.

To get a slight feel for the above described basis set effect, we return to the model diffusion process described in section 2.10.2 on k points above, in which a phosphorus atoms move into a adjacent vacancy. With the number of k points fixed ($n_k = 1$), the effect of varying the energy cutoff E_{cut} on the activation energy for the model process was studied.

As a function of varying cutoff, the lower part of figure 2.5 shows the difference in energy for each increment on a logarithmic scale. As expected, the energy difference decreases towards zero as the converged value is approached. More interesting, the error from the 'very soft' potential is almost one order of magnitude lower than from the 'standard' potential for cutoffs below 300 eV. The upper figure shows the percentage error in the energy barrier, compared to converged values E_{conv} (1.622 eV for the normal GGA potential, 1.618 eV for the 'very soft' GGA potential and 1.565 eV for the LDA potential). Here we see that the 'very soft' potential works very well down to even 120 eV. This is very impressive, but since this is just a model reaction, we increase the cutoff value to 180 eV in the MD simulations to add a little more flexibility to the basis set.

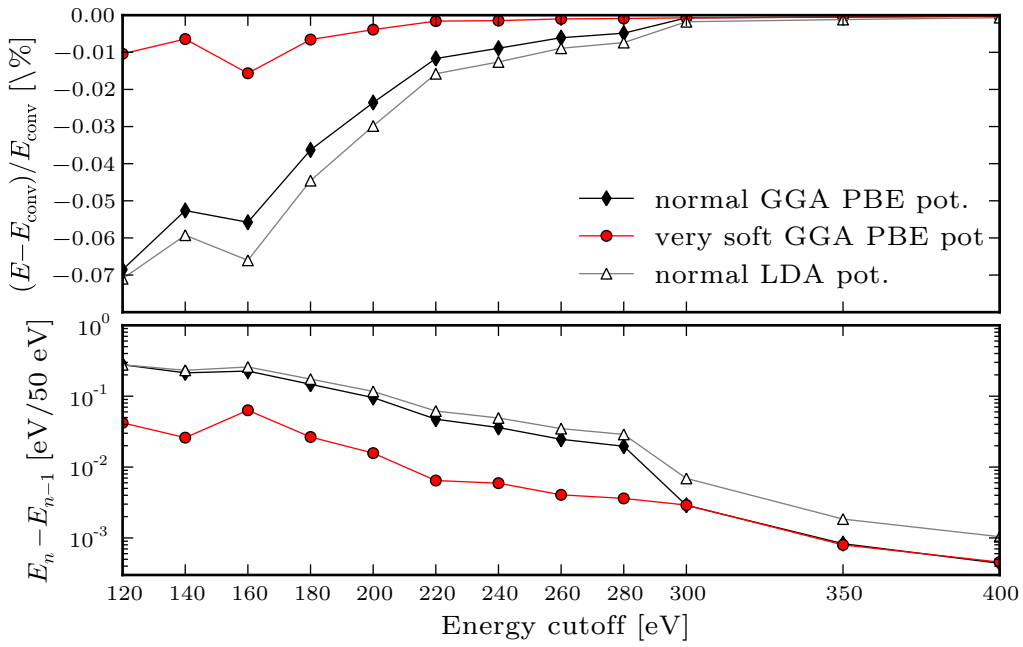
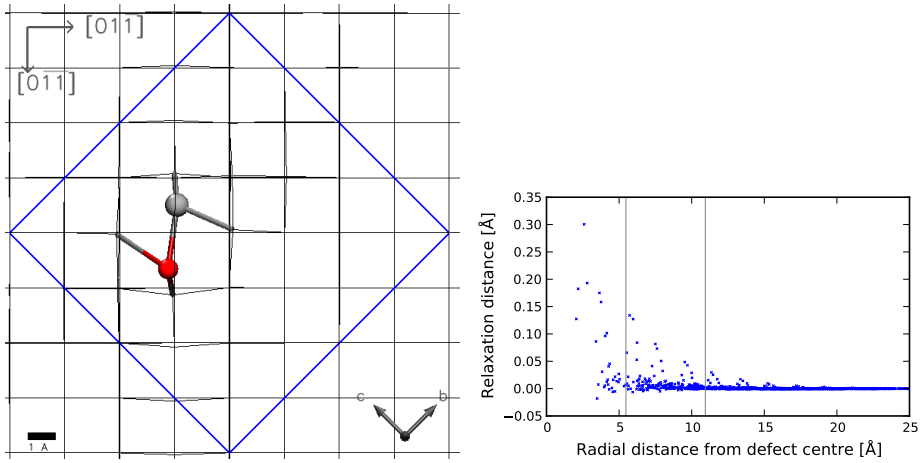


Figure 2.5.: Influence of the energy cutoff on the energy barrier for a phosphorus atom to jump into a neighbour vacancy site. *Top*: The error in barrier energy relative to the converged energy barrier value for each potential. Note that the converged values for the different potentials. The converged value is taken as the average barrier height found from calculations using 800, 900 and 1000 eV as energy cutoff. *Bottom*: Difference in barrier energy upon increase of the energy cutoff by 50 eV.

2.10.4. Cell size

A too small simulation cell will result in spurious self-interaction across the periodic boundaries, but from a computational cost perspective it was quickly found that a simulation cell larger than $2 \times 2 \times 2$ (64 atoms) would not be viable with MD simulations. While we can give no authorative answer to whether this cell is "large enough", a small test was carried out to check the "range" of a P-Si defect (one of the configurations discussed in section 4.2 below). The defect was introduced into a very large cell, a $6 \times 6 \times 6$ cell with lattice parameter 32.8 Å containing 1728 atoms, and the whole cell was relaxed.

The right part of fig. 2.6 shows the distance each of the atoms in the cell relaxed as a function of the radial distance from the defect center. The two vertical lines indicate the length of 8-atom cell (5.47 Å) and a 64-atom cell (10.74 Å). The left part of the figure shows a small part of the cell encompassing a 64-atom cell around the defect, with the size of the 64-atom cell indicated by a blue square. As we see, the defect causes some distortion even beyond the boundaries of the 64-atom cell, but the error from self-interaction is most likely not any worse than the other small errors we've already introduced.



(a) (b)

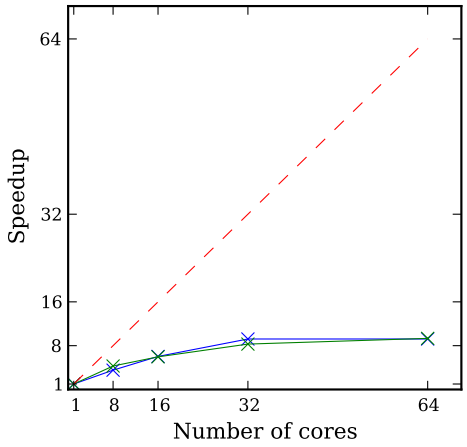
Figure 2.6.: Left) A small part of the $6 \times 6 \times 6$ unit cell (side length 32.8 Å), showing the relaxed split $\langle 110 \rangle$ Si-P pair. Indicated with a blue frame is the $2 \times 2 \times 2$ cell with side length 10.94 Å. Right) Relaxation distance for all the 1728 atoms in the cell as a function of radial distance from the defect centre.

2.10.5. Summary and parallelisation

After a somewhat lengthy discussion, the standard setup to be used to calculate the forces at each step in the molecular dynamics runs is now defined; using a $2 \times 2 \times 2$ cell with lattice parameter 10.94 Å, the "very soft" PAW PBE potential with energy cutoff

of 180 eV, and a single k point (the Γ point) with a optimised `VASP` build for Γ-only calculations.[23]

Test calculations were carried out at two different Norwegian computing clusters; the Titan cluster at University of Oslo, and the Stallo cluster at University of Tromsø, for which the results are shown in fig. 2.7. The calculations included 46656 plane waves and about 165 bands (163-168 depending on the number of cores).



Cluster	Cores n	Time t_n [s]	Speedup t_1/t_n	Efficiency $t_1/(nt_n)$
Stallo[a]	1	13.09	–	–
	8	3.71	3.6	44%
	16	2.17	6.2	38%
	32	1.42	9.3	29%
	64	1.42	9.3	14%
Titan[b]	1	16.33	–	–
	8	3.81	4.3	54%
	16	2.74	6.0	37%
	32	1.97	8.3	26%
	64	1.75	9.3	15%

[a] 2 × 2.66 GHz Intel Xeon quad core nodes.
[b] 2 × 2.3 GHz AMD Opteron quad core nodes.

Figure 2.7.: Scaling on two different clusters, Titan and Stallo, both with Infiniband connect. For a short molecular dynamics simulation parallelised over n cores, t_n is the mean clock time per step in seconds (100 steps, with the two first left out from the mean). Ideal scaling corresponds to the total CPU time nt_n being constant, and the clock time decreasing as $1/n$. The actual speedup t_1/t_n as a percentage of the ideal speedup is shown in the last column of the table. This can be considered the efficiency of computer resource utilisation.

The serial calculation time per ionic step is about 13 seconds on Stallo, and a few seconds longer on Titan. This includes on average three electronic steps per ionic step. While I find this number impressive for an *ab initio* calculation, it still means simulating a nanosecond would take 150 days. The clock time needed can be reduced by distributing the work over several CPU cores, working in parallel, but as the results shows, the calculation scales poorly. While the clock time is reduced by a factor of ten by using 64 cores, the efficiency is only about 15 %, meaning that a large amount of computational resources are wasted. Surprisingly, even parallelisation over several cores at the same node is only 45% efficient, indicating that inter-node communication is not a bottleneck.

Of course, good scaling requires a task that actually can be parallelised. A molecular dynamics simulation is not very well suited in that respect; Two ionic steps can logically not be carried out parallelly, as one depends on the other, and neither can two electronic steps. We are left parallelising the calculation for each electronic step, and with only one k point, what remains is the options to parallelise over bands (n) and plane waves (G) in (2.20).

[23] Thanks to Espen Flage-Larsen for the build!

The `VASP` code can handle parallelisation over bands and plane waves efficient up to at least 32 cores in typical cases.[71] But with electronic steps calculated in less than one clock-second, it is possible that initialisation steps and other parts of the code that can not be parallelised, starts to dominate over the parallelised parts of the code. Investigating this further could be an interesting study on its own, but here we will just accept the current scaling situation and parallelise only over eight cores.

3. Molecular dynamics simulations

In section 2.2 we used the Born Oppenheimer approximation to separate the total Hamiltonian $\hat{H}$ into a nuclei part $\hat{H}_N$ and an electronic part $\hat{H}_e$, and in the remainder of chapter 2 we only looked at the electronic one. In this chapter we will be concerned with the Hamiltonian for the N nuclei, which in the absence of any constraints can be written as a sum of the kinetic energy and the potential energy,

$$\hat{H}_N = \hat{K}_N + \hat{V}_N = \sum_{k=1}^{N} \frac{1}{m_k} \nabla_k^2 + V_N, \tag{3.1}$$

Since we will only be concerned with the nuclei Hamiltonian, I will skip the subscripts N from now on.

In contrast to the electrons, we assume that the nuclei can be treated as classical entities, following trajectories $\mathbf{r}_k(t)$ described by classical equations of motion, be it in the first-order formulation of Hamilton,

$$\frac{d\mathbf{r}_k}{dt} = \frac{\partial H}{\partial \mathbf{p}_k}, \qquad \frac{\partial H}{\partial \mathbf{r}_k} = -\frac{d\mathbf{p}_k}{dt}, \tag{3.2}$$

where motion occurs in a way as to preserve the Hamiltonian function, [1]

$$H(r, p, t) = K(p, t) + V(r, t) = \sum_{k=1}^{N} \frac{\mathbf{p}_k(t)^2}{2m_k} + V(r, t), \tag{3.3}$$

or the second-order formulation of Newton,

$$\frac{d^2}{dt^2}\mathbf{r}_k = \frac{1}{m}\mathbf{F}_k, \tag{3.4}$$

where motion is a response to an applied force.

Using the formulation of Newton, trajectories was evolved by integrating (3.4) using the velocity Verlet algorithm (section 3.2) and forces from the Hellmann-Feynman force theorem (section 2.9), as implemented in `VASP`. Temperature was maintained by the 'velocity scaling thermostat' (section 3.3.2).

[1] r, p is shorthand for $\{\mathbf{r}_k, \mathbf{p}_k\} = \{r_{kx}, p_{kx}\}$, the set of all positions and momenta. For N atoms, we have $6N$ Cartesian coordinates, but excluding movement relative to the frame, we are left with $6(N-1)$ degrees of freedom, which means we only need that many generalised coordinates. Generalised coordinates is used in statistical mechanics, but since the topic is only touched upon here, I'm not distinguishing notation between the two, although I'm not sure if that was a good idea or not.

3.1. Statistical mechanics

What is obtained from a molecular dynamics simulation, is the configuration or microstate of the system (the positions and momenta of every single atom) at any given time included in the simulation – a quantity that can not be measured experimentally. How to relate the *microscopic* configuration to *macroscopic* quantities that can be measured experimentally (*observables*), such as temperature (3.14) or diffusivities (3.32) is the subject of statistical mechanics.

Statistical mechanics is concerned with statistical *ensembles*, an ensemble being a theoretical construct holding a large number of copies (sometimes infinitely many) of essentially the same system, that is; a collection of systems described by the same set of microscopic interactions, and sharing a common set of macroscopic control variables like internal energy E, volume V and number of atoms (or moles) N.[72, p. 63]

Each system in an ensemble has a different *microstate*, but all systems share the same *macrostate*. A microstate is specified by specifying the state of each individual particle in the system, while a macrostate specifies only how many particles are in each state.

In statistical mechanics, (macroscopic) observables are found from averages over microstates. For some observable A, represented by the operator $\hat{A}$, such an ensemble average can be written

$$\langle A \rangle = \sum_i P_i \langle i|\hat{A}|i\rangle = \sum_i \langle i|\hat{P}\hat{A}|i\rangle, \tag{3.5}$$

where angle brackets $\langle \cdots \rangle$ indicate an ensemble average, or expectation value, and $P_i = \langle i|\hat{P}|i\rangle$ is the probability that any given system in the ensemble will be in a microstate i, depending on what kind of ensemble we are simulating. We are not concerned with discrete states, however, but rather continuous states of coordinates q and momenta p in the $6N-6$-dimensional classical phase space for N particles. The probability operator is then replaced by a distribution function $\rho(p,q)$, and (3.5) can be shown to reduce to[2]

$$\langle A \rangle = \int \rho(r,p) A(r,p)\, dr\, dp \tag{3.6}$$

As with the probability operator, the phase space density function will depend on what kind of ensemble we study. Two types of ensembles will be introduced in section 3.1.2.

3.1.1. The ergodic hypothesis

The basic idea of molecular dynamics is that the average over *all systems* in the ensemble at a single time (3.6) can be replaced by an average over a single system in the ensemble at *all times*, that is over the time evolution of that system,

$$\bar{A} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau A(r(t), p(t))\, dt \tag{3.7}$$

along the trajectory, where τ is the observation time.

[2] See e.g. [73, pp. 13-15]. Again, generalised coordinates could be used instead.

The ergodic hypothesis has only been proved for some simple model systems such as the hard sphere gas.[74] For a general system, it can neither be proven or disproven,[72, p. 99] but it can be tested by comparing Monte Carlo results (ensemble averages) with molecular dynamics results (time averages). Its plausibility for any given system may also be discussed in terms of the energy barriers of the system, since unsurpassable energy barriers will separate phase space into regions that can not be sampled from the same initial conditions.

For the simulation of phosphorus diffusion in silicon, the ergodic hypothesis is expected to hold, but we may still run into sampling problems if we can not simulate the system for a long enough time τ. In practice, we also have to use finite time steps Δt instead of infinitesimals dt, as discussed in section 3.2 below.

3.1.2. Ensembles

The Hamiltonian (3.3), describes a system completely isolated from its surroundings. For such a system, the number of particles N, the volume V and the internal energy E are all conserved, since there is no mechanism in the model for exchanging particles, volume or energy with the surroundings. During a simulation, only configurations with the same N, V and E will thus be visited, configurations of the so-called *microcanonical* or NVE ensemble. This ensemble is characterised by the phase space probability function being constant,

$$\rho \propto \delta(E - H(r, p)), \tag{3.8}$$

with equal probability of sampling phase space points on the hypersurface $H(r, p) = E$.

Molecular dynamics samples configurations from the microcanonical ensemble, but we may well be interested in sampling ensembles with other control variables to better simulate experimental conditions. In a typical experiment, it is easier to control the temperature T than the internal energy E, the pressure P than the volume V, and often the chemical potential μ than the number of particles N. Such an μPT ensemble is the *grand canonical ensemble*.

In the case of a solid system, the volume and number of particles will generally vary little, justifying the use of the simpler NVT or *canonical ensemble*, where the phase space probability function is the Gibbs probability distribution,

$$\rho(r, p) = \frac{1}{Z} \mathrm{e}^{-\beta H(r,p)}, \qquad Z = \int \mathrm{e}^{-\beta H(r,p)} \, dr \, dp \tag{3.9}$$

Methods for carrying out a molecular dynamics simulation sampling the canonical ensemble is discussed in section 3.3 on thermostats below. It is worth noting that the constant-volume restriction may not always be justified, even for solids. If the unit cell is relaxed at zero temperature, and simulations are carried out at elevated temperatures, this will lead to an artificial pressure build-up.

Experimentally, the lattice parameter of silicon has been shown to expand by about 0.026 Å, from 5.431 Å to 5.457 Å, when the temperature is increased from 298 K to

1500 K.[12] With the quite flat minimum in the energy-lattice parameter curve (fig. 2.3) and uncertainties related to it, this effect is expected to be of less importance, but it could be interesting to investigate. Especially at temperatures close to the melting point, at least equilibrating the system using a NPT ensemble rather than a NVT ensemble might be advantageous.

3.1.3. The virial and the equipartition theorems

The virial theorem is one of the theorems connecting microscopic phase space functions with macroscopic thermodynamic observables. If x_i and x_j are generalised coordinates, then the virial theorem states that[3]

$$\left\langle x_i \frac{\partial H}{\partial x_j} \right\rangle = k_B T \delta_{ij} \tag{3.10}$$

where the brackets indicate an average over a microcanonical ensemble, which can, if the system is ergodic, be replaced by an average over time.

If we take $x_i = p_i$, a momentum component, then for a system described by a Hamiltonian (3.3), we have

$$\left\langle p_i \frac{\partial H}{\partial p_i} \right\rangle = k_B T, \tag{3.11}$$

$$\left\langle \frac{p_i^2}{2m_i} \right\rangle = \frac{1}{2} k_B T, \tag{3.12}$$

where m_i is the mass of the particle having the momentum component p_i. Summing both sides over all $3N-3$ components, results in the equipartition theorem,

$$\sum_{i=1}^{3N-3} \left\langle \frac{p_i^2}{2m_i} \right\rangle = \sum_{i=1}^{3N-3} \frac{1}{2} k_B T, \tag{3.13}$$

$$\langle K \rangle = \frac{1}{2} f k_B T, \tag{3.14}$$

where $f = 3N-3$ is the degrees of freedoms. The theorem relates the thermodynamic temperature T to the average kinetic energy $\langle K \rangle$. From the relation, we could also define an 'instantaneous temperature'

$$\mathcal{T}(t) = \frac{2K(t)}{f k_B} = \frac{2}{f k_B} \sum_{k=1}^{N} \varepsilon_k(t) = \sum_{k=1}^{N} \frac{m_k v_k(t)^2}{f k_B}, \tag{3.15}$$

that we will make use of in the next section to discuss temperature fluctuations in a system.

[3] For a proof, see e.g. [72, p. 81].

3.1.4. Maxwell-Boltzmann distribution for kinetic energies

With the Hamiltonian being separable, $H(r,p) = K(p)+V(r)$, Gibbs probability function (3.9) is separable as well; $\rho(r,p) = \rho_T(p)\rho_V(r)$. With the total kinetic energy

$$K(p) = \sum_{k,x=1}^{3N} \frac{p_{kx}^2}{2m_k} \tag{3.16}$$

from (3.3), we find

$$\rho_T(p) = \frac{\mathrm{e}^{-\beta K(p)}}{\int \mathrm{e}^{-\beta K(p)} dp} = \prod_{k,x}^{3N} \frac{\mathrm{e}^{-\beta p_{kx}^2/2m_k}}{\int \mathrm{e}^{-\beta p_{kx}^2/2m_k} dp_{kx}} = \prod_{k,x}^{3N} \rho_{\mathrm{MB}}(p_{kx}), \tag{3.17}$$

where p_{kx} is a momentum component for nucleus k, related to a velocity component by $p_{kx} = m_k v_{kx}$, so we can find the distribution of velocities

$$\rho_{\mathrm{MB}}(v_{kx}) = \frac{\mathrm{e}^{-\beta m_k v_{kx}^2/2}}{\int \mathrm{e}^{-\beta m_k v_{kx}^2/2} dv_{kx}} = \left(\frac{\beta m_k}{2\pi}\right)^{1/2} \mathrm{e}^{-\beta m_k v_{kx}^2/2}, \tag{3.18}$$

which is the Maxwell-Boltzmann distribution, anticipated by the choice of subscript. This is a Gaussian with standard deviation $\sigma = 1/\sqrt{\beta m_k} = \sqrt{k_B T/m_k}$ about $\langle v_{kx}\rangle = 0$.

Although derived in the canonical ensemble, the velocity distribution is expected to be Maxwellian in any equilibrium situation, since properties in different ensembles become equivalent in the thermodynamic limit (large N and V).[75, p. 65] With only 65 atoms we can not really claim being close to the thermodynamic limit, but we will still assume a Maxwellian distribution.

Now, for the speed $v_k = \sqrt{v_{kx}^2 + v_{ky}^2 + v_{kz}^2}$, we have

$$\rho_{\mathrm{MB}}(v_k) = 4\pi \left(\frac{\beta m_k}{2\pi}\right)^{3/2} v_k^2 \mathrm{e}^{-\beta m_k v_k^2/2}. \tag{3.19}$$

From its second and fourth moments,

$$\langle v_k^2\rangle = \int_0^\infty v_k^2 \rho_{\mathrm{MB}}(v_k)\, dv_k = \frac{3}{m_k\beta}, \tag{3.20}$$

and

$$\langle v_k^4\rangle = \int_0^\infty v_k^4 \rho_{\mathrm{MB}}(v_k)\, dv_k = \frac{15}{(m_k\beta)^2}, \tag{3.21}$$

we can find the relative variance in single-particle kinetic energy $\varepsilon_k = \frac{1}{2} m_k v_k^2$ as

$$\frac{\sigma_{\varepsilon_k}^2}{\langle\varepsilon_k\rangle^2} = \frac{\langle\varepsilon_k^2\rangle - \langle\varepsilon_k\rangle^2}{\langle\varepsilon_k\rangle^2} = \frac{\langle v_k^4\rangle - \langle v_k^2\rangle^2}{\langle v_k^2\rangle^2} = \frac{2}{3}, \tag{3.22}$$

and the relative variance in instantaneous temperature (3.15),[73, p. 127]

$$\frac{\sigma_{\mathcal{T}}^2}{\langle\mathcal{T}\rangle^2} = \frac{2}{f} = \frac{2}{3N-3}. \tag{3.23}$$

We will come back to this instantaneous temperature fluctuation in section 3.3 on thermostats.

3.1.5. The Einstein-Smoluchowski diffusion relation

Following Frenkel[73, pp. 78-79], we start with Fick's second law (1.3),

$$\frac{\partial c}{\partial t} - D\nabla^2 c = 0 \tag{3.24}$$

where $c(\mathbf{r}, t)$ is the concentration, whose second moment is

$$\langle r^2(t)\rangle \equiv \int_V r^2(t) c(\mathbf{r}, t)\, dV, \tag{3.25}$$

and normalisation is

$$\int_V c\, dV = 1. \tag{3.26}$$

We multiply (3.24) by r^2, and integrate over all space (V),

$$\frac{\partial}{\partial t}\int_V r^2(t) c(\mathbf{r}, t)\, dV = D\int_V r^2(t)\nabla^2 c(\mathbf{r}, t)\, dV \tag{3.27}$$

The left hand side is just

$$\frac{\partial}{\partial t}\langle r^2(t)\rangle, \tag{3.28}$$

while the integral on the right hand-side can be treated with partial integration and the divergence theorem. Here, S is the surface of all space, with surface element $d\mathbf{S} = \hat{\mathbf{n}} dS$, $\hat{\mathbf{n}}$ being the surface normal.

$$\begin{aligned}
\int_V r^2\nabla^2 c\, dV &= \int_V \nabla\cdot(r^2\nabla c)\, dV - \int_V \nabla r^2\cdot\nabla c\, dV \\
&= \underbrace{\oint_S r^2\nabla c\cdot d\mathbf{S}}_{0} - 2\int_V \mathbf{r}\cdot\nabla c\, dV \\
&= -2\int_V \nabla(c\mathbf{r})\, dV + 2\int c\nabla\mathbf{r}\, dV \\
&= \underbrace{-2\oint_S c\mathbf{r}\cdot d\mathbf{S}}_{0} + 2\int c\nabla\mathbf{r}\, dV \\
&= 2d\int_V c\, dV \\
&= 2d
\end{aligned} \tag{3.29}$$

since $\nabla\mathbf{r} = d$, the dimensionality. Substituting (3.28) and (3.29) into (3.27), and letting $d = 3$, we have

$$\frac{\partial}{\partial t}\langle r^2(t)\rangle = 6D. \tag{3.30}$$

Since motion at the microlevel is chaotic, Fickian diffusion is only obtained over 'long' time scales, making the relation exact in the limit $t \to \infty$,

$$D = \frac{1}{6}\lim_{t\to\infty}\frac{\partial}{\partial t}\langle r^2(t)\rangle. \tag{3.31}$$

This is the Einstein-Smulowski relation, revealed independently by Einstein (1905)[16] and Smoluchowski (1906),[17] which relates the macroscopic self-diffusion coefficient D to the slope of the microscopic mean-square displacements r^2 as a function of time.

If we have N_α atoms of a species α, the self-diffusion coefficient D_α for species α is

$$D = \frac{1}{6N_\alpha} \lim_{t\to\infty} \frac{d}{dt} \sum_{k=1}^{N_\alpha} \left\langle r_k(t)^2 \right\rangle, \tag{3.32}$$

While the relation is generally exact in the limit of infinite trajectories, we can only finite ones, in practice timespans of picoseconds or nanoseconds. This calls for the question about how long is 'long enough', a question that is not necessarily easy to answer in advance. During a simulation, we have some options, however. One way is to monitor the slope of the mean square displacement. If the slope approaches a straight line, that is a good indication for the validity of the Einstein relation.

3.2. The velocity Verlet algorithm

The trajectory $\mathbf{r}_k(t)$ of a classical particle k can in principle be found by integrating Newton's law of motion

$$\ddot{\mathbf{r}}_k(t) = \frac{1}{m_k}\mathbf{F}_k(t). \tag{3.33}$$

with $\mathbf{F}_k(t)$ the force acting on the particle (atom) at time t. No analytic expression is available for $\mathbf{F}_k(t)$ (if it was, carrying out a simulation would not be necessary), but $\mathbf{r}_k(t)$ can be found in a step-wise manner using finite-difference methods to estimate the position $\mathbf{r}_k(t+\Delta t)$ for small Δt. In principle, using a Taylor expansion to infinite order, the position can be found for any Δt, but that would require knowledge about the time derivatives of $\mathbf{r}_k(t)$ to infinite order. In practice the expansion has to be truncated at low order, requiring Δt to be small to minimize *truncation errors*. Apart from truncation errors, which will depend on the specific finite-difference method, all such methods will be prone to *round-off errors*, since a small round-off error in each step may grow very large with very many steps.

The 'prototype method' is the Euler method, which makes use of a Taylor expansion to first order around time t,

$$\mathbf{r}_k(t+\Delta t) \approx \mathbf{r}_k(t) + \mathbf{v}_k(t)\Delta t, \tag{3.34}$$

where $\mathbf{v}_k(t) = \dot{\mathbf{r}}_k(t)$ is the particle's velocity.

In the Verlet method,[76], here shown in the "velocity Verlet" variant by Swope,[77] we expand to second order about t,

$$\mathbf{r}_k(t+\Delta t) \approx \mathbf{r}_k(t) + \mathbf{v}_k(t)\Delta t + \frac{1}{2m_k}\mathbf{F}_k(t)(\Delta t)^2, \tag{3.35}$$

and about $t + \Delta t$, evolving backwards in time,

$$\mathbf{r}_k(t) \approx \mathbf{r}_k(t + \Delta t) - \mathbf{v}_k(t + \Delta t)\Delta t + \frac{1}{2m_k}\mathbf{F}_k(t + \Delta t)^2. \tag{3.36}$$

Substituting 3.35 for $\mathbf{r}_k(t + \Delta t$ in 3.36, and solving for $\mathbf{v}_k(t + \Delta t)$,

$$\mathbf{v}_k(t + \Delta t) = \mathbf{v}_k(t) + \frac{\Delta t}{2m_k}\left[\mathbf{F}_k(t) + \mathbf{F}_k(t + \Delta t)\right] \tag{3.37}$$

In the velocity Verlet algorithm, (3.35) and (3.37) are used in combination to evolve the positions and velocities simultaneously.

While the Euler method is a one-step algorithm, we see that the Verlet algorithm is a two-step algorithm, requiring information both from the current and the previous step. In the initial step, there is no previous step, so positions and velocities have to be supplied. The initial positions will be discussed in section 4.2, and the initial velocities in section 3.2.1 below.

The Verlet method requires relatively short time steps compared to more advanced methods, such as the class of predictor-corrector algorithms, but a major advantage of the Verlet method is that it exhibits little long-term drift. This is related to the method being time-symmetric and sympletic (it preserves phase-space volume), unlike the Euler method and most of the more advanced methods.[73, pp. 61-63]

3.2.1. Initial velocities

For the first step, the positions and velocities (or momenta) of all the atoms have to be specified manually. The velocities should be sampled from a Maxwell-Boltzmann velocity distribution (3.18), that is, from a normal distribution

$$f(v_{kx}) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \mathrm{e}^{-v_{kx}^2/2\sigma^2} \tag{3.38}$$

where $\sigma = 1/\sqrt{\beta m_k}$ is the width of the Gaussian and v_{kx} is a single velocity component. We can get a computer to generate a set of random (or pseudo-random) numbers ξ from a *uniform* distribution with a given range, like $[0, 1]$. To turn these into random numbers from a normal distribution, we make use of the cumulative distribution function

$$F(X) = \int_{-\infty}^{X} f(v_{kx})\, dv_{kx} = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \int_{-\infty}^{X} \mathrm{e}^{-v_x^2/2\sigma^2}\, dv_x, \tag{3.39}$$

since $\xi = F(X)$ is a uniformly distributed variable in the range $[0, 1]$. The problem of sampling $f(v_x)$ then consists of solving the equation $F(X) = \xi$ for X. In the most naïve approach, known as *inverse transform sampling*, we just find $X = F^{-1}(\xi)$. But since we don't have a closed form expression for $F(X)$, a computationally more efficient method is preferred. A common approach, which is also used in `VASP`, is the *Box-Muller transform sampling* method, which is described very well in [72, p. 101]. From two random numbers ξ_1 and ξ_2, two gaussian random numbers X and Y are generated by

$$X = \sigma\sqrt{-2\ln\xi_2}\cos(2\pi\xi_1), \tag{3.40}$$

$$Y = \sigma\sqrt{-2\ln\xi_1}\sin(2\pi\xi_2). \tag{3.41}$$

3.3. Thermostats

3.3.1. No thermostat

The Hamiltonian (3.3) describes an isolated system, in which the total energy E must necessarily be a constant of motion. Upon the start of the simulation, a temperature $T(t = 0)$ is imposed by giving the system a total kinetic energy K corresponding to that temperature. The potential energy is found by solving the electronic Hamiltonian.

By following the time evolution of such a Hamiltonian, we will sample a microcanonical ensemble. While the total energy $E = K + V$ is a constant of motion, the kinetic energy K and potential energy V themselves will fluctuate as energy is exchanged in particle collisions, and over time the kinetic and potential energy may very well each drift away from their initial values. And if they do, so does the temperature, according to (3.15). A most dramatic example of this can be seen in fig. 3.1 below, showing the temperature and total energy variation for the first 0.5 ps of a 5.0 ps simulation of a system of 62 silicon atoms and a phosphorus atom in a 64-atom cell. The system starts with an initial temperature of 1400 K, but rapidly drops to a mean temperature of 690 K. Interestingly, the system remains very stable at this temperature for the remaining 4.5 ps.

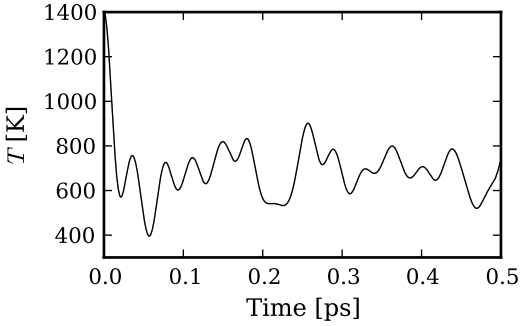


Figure 3.1.: Instantaneous temperature variation for a system with no thermostat, starting at 1400 K. To be able to see the rapid temperature drop, only the first 0.5 ps of the 5.0 ps simulation is shown. In the remaining 4.5 ps, the system remained very stable at the same temperature.

3.3.2. Ad-hoc velocity-scaling thermostat

In thermodynamics, constant temperature is achieved by putting the system in contact with an infinite heat bath, or with a *thermostat*. The simplest approach to simulating a thermostat is to just rescale the particle velocities so that the instantaneous temperature $\mathcal{T}(t)$ matches the temperature T of the thermostat. From the definition of the instantaneous temperature (3.15), we see that rescaled velocities can be introduced as

$$v_j' = \sqrt{\frac{T}{T_k}} v_j. \tag{3.42}$$

Physically, such a rescaling corresponds to a process where energy is instantaneously transferred between the heat bath and all the particles. As such, it is not the most realistic model. Also, there is no inherent clue on how often the re-scaling should be carried out. Re-scaling the temperature every single step is clearly not a good idea, since that would result in a delta function distribution instead of the Gaussian distribution with standard deviation

$$\sigma_{\mathcal{T}} = \langle \mathcal{T} \rangle \sqrt{\frac{2}{3N-3}} \tag{3.43}$$

expected for a canonical ensemble with N particles from (3.23). The number of steps between rescaling is therefore a parameter that has to be selected such as to obtain a temperature distribution similar to the temperature distribution for a canonical ensemble.[4]

In this work, the rescale periode has been set to 50. Fig. 3.2a shows the resulting instantaneous temperature and total energy variations for a 5 ps simulation of the same system simulated without a thermostat in section 3.3.1 above. In the temperature distribution, we see that the rescaling results in a strong peak at the thermostat temperature, but otherwise the shape is quite Gaussian. In particular, it's satisfying that the distribution is quite symmetric. If the temperature had dropped between each rescaling in the same way as it did in 3.1, the result would have been a very asymmetric distribution. The standard deviation of 110 K is lower than 146 K, the standard deviation for the perfect canonical distribution, (3.43), but it was considered acceptably close, and a more systematic investigation of the effect of varying the rescale period was not carried out, although it would certainly be interesting.

3.3.3. Nose-Hoover thermostat

While velocity-scaling is a simple and straightforward approach to keeping the thermodynamic temperature constant, it does not sample a canonical ensemble, but rather some canonical-like ensemble. More advanced thermostat models have therefore evolved, that accurately sample a canonical ensemble, of which the most popular approaches are those of Andersen[78] and Nose[79–81].

In Andersen-type thermostats, the temperature is controlled by stochastic collisions with the heat bath.[5] This makes the simulation non-deterministic and the interpretation of the trajectories somewhat difficult.

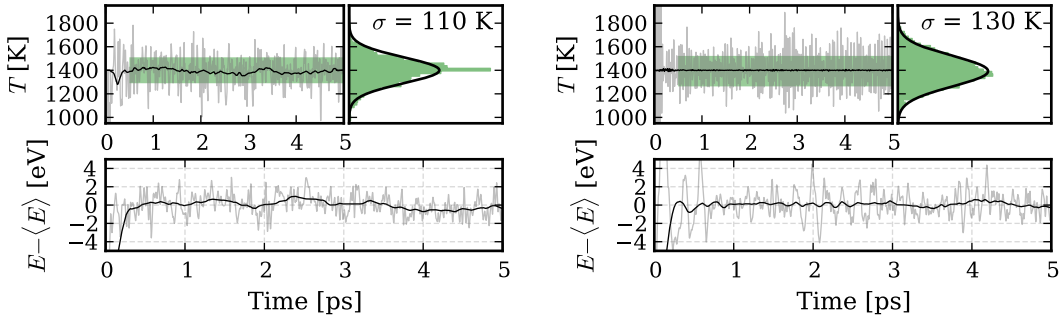
In Nose-type thermostats, a heath bath is added as a component s to the Lagrangian, forming an "extended system" coupled with the original system, from which a new Hamiltonian can be derived in the Nose-Hoover scheme. The original system is allowed to exchange energy with the thermostat through variations in s, thus breaking the energy conservation of the original system, but conserving the total energy of the system plus the thermostat. The heath bath is associated with an "effective mass" called the Nose mass, that is used to control the coupling strength. The Nose-Hoover equations of motion are

[4] In `VASP` this is the NBLOCK parameter

[5] A simple pseudo-code is given in [73, p. 129].

smooth, deterministic and time-reversible, but may lead to high-frequency temperature oscillations, and some care must be taken in the choice of the Nose mass.[82]

The simulation carried out above with no thermostat (fig. 3.1) and with velocity-scaling (fig. 3.2a), was repeated using a Nose-Hoover thermostat. As seen in figure 3.2b, the thermostat gives a near-perfect Gaussian distribution with a standard deviation of 130 K. It is worth noting that this is still lower than the expected 146 K for a perfect canonical ensemble of this size and temperature. It is possible that this is an effect of non-ergodicity, with the distribution not fully corresponding to the state-averaged distribution, or an effect of choosing a non-optimal Nose mass.



(a) Velocity-scaling **(b)** Nose-Hoover thermostat

Figure 3.2.: Instantaneous temperature distribution (above) and total energy fluctuation (below) in a simulation using velocity-scaling (left) and Nose-Hoover thermostat (right), for a system with $N = 62$, $T = 1400\,\mathrm{K}$. A 0.5 ps symmetric running mean is drawn (black) on top of the raw data (gray). The standard deviation σ is indicated by a solid, green overlay. In the upper right, the temperature distribution and its standard deviation is shown, together with a Gaussian (thick black line) with the same standard deviation and normalisation.

At this point, the Nose-Hoover thermostat may seem superior to the velocity-scaling thermostat, and it probably is. The choice of using the velocity-scaling thermostat was taken primarily based on reports of energy drift issues with the Nose-Hoover thermostat within my group. I'm not sure if this could have been related to the implementation in `VASP`, poor choice of parameters, or a more general issue with the method, and I regret that I did not investigate this further. Yet, the velocity-scaling thermostat does keep the temperature, produces a quite Gaussian distribution, and its simplicity at least makes it very transparent.

3.4. Monitoring translational order

Crystals are characterised by the presence of short- and long-range translational order, and the melting of a crystal is characterised by the loss of long-range translational order. During a molecular dynamics simulation it can therefore be useful to monitor to what degree translational order is preserved with a simple parameter.

For the current purpose we can assume all M atoms are equal, and we can then write the structure factor as[38, p. 241] a Fourier Transform

$$F_{\mathbf{G}}(\mathbf{R}^0) = \sum_{j=1}^{M} \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{R}_j^0}, \tag{3.44}$$

where $\mathbf{G} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ is a vector in the reciprocal lattice, related to a vector in the direct lattice, $\mathbf{R}^0 = u\mathbf{a}_1 + v\mathbf{a}_2 + w\mathbf{a}_3$, by $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$.

Since the structure factor is in general a complex number, it's practical to work with the modulus, which I denote $F_{\mathbf{G}}^0 = |F_{\mathbf{G}}(\mathbf{R}^0)|$. Now, for a real structure with atoms at positions $\mathbf{R}$, in general different from the ideal lattice sites $\mathbf{R}^0$ at any given time, a simple translational order parameter can be devised from the ratio of moduluses $F_{\mathbf{G}}/F_{\mathbf{G}}^0$;[6]

$$\lambda_{\mathbf{G}} = \frac{1}{F_{\mathbf{G}}^0}\left|\sum_{j=1}^{M} \mathrm{e}^{-i\mathbf{G}\cdot\mathbf{R}_j}\right| = \frac{1}{F_{\mathbf{G}}^0}\sqrt{\left[\sum_{j=1}^{M}\cos(\mathbf{G}\cdot\mathbf{R}_j)\right]^2 + \left[\sum_{j=1}^{M}\sin(\mathbf{G}\cdot\mathbf{R}_j)\right]^2}, \tag{3.45}$$

using any $\mathbf{G}$ vector that gives rise to a non-zero structure factor $F_{\mathbf{G}}^0$. This parameter is unity for a perfect crystal, and positive but small, of order $\sqrt{M}$, for a completely disordered system. The exact value will depend on the choice of $\mathbf{G}$ vector(s).

For the diamond structure, we have a conventional (cubic) unit cell with unit cell vectors of length a aligned to the cartesian unit vectors and a basis of eight atoms. The smallest $\mathbf{G}$ vectors that give rise to a non-zero $F_{\mathbf{G}}^0$ are the [111], [220], [311] and [400] vectors, with $h^2 + k + ^2 + l^2 = 3$, 8, 11 and 16, respectively. For $h + k + l$ even and divisable by four, $F_{\mathbf{G}}^0 = 8$ (the number of atoms), while for $h + k + l$ odd, $F_{\mathbf{G}}^0 = 8/\sqrt{2}$.

To illustrate some aspects of the order parameter 3.45, we start with a perfect silicon crystal, for which the order parameter is unity. We then randomly displace all atoms so that the mean radial distance between an atom and its associated lattice site $\langle|\mathbf{R} - \mathbf{R}_0|\rangle$ is a fixed value. The order parameter is calculated and averaged over a large number of different random configurations with the same $\langle|\mathbf{R} - \mathbf{R}_0|\rangle$. The left part of fig. 3.3 shows the result on the translational order parameter (3.45) using different $\mathbf{G}$ vectors. As we see, the shorter $\mathbf{G}$ vectors are less sensitive to small displacements, such as vibrations. The right part of the figure shows the effect of the system size. The shaded part is the standard deviation. In each case, the parameter approaches a value of about $\sqrt{M}$ for a disordered system.

Fig. 3.4 shows an example of melting from an actual MD run with 64 Si atoms and one P atom at 1600 K, where the system melted after being in a semistable, crystalline state for almost 29 ps, with 3.4a showing the translational order parameter using three different $\mathbf{G}$ vectors, and 3.4b showing the change in potential energy that also accompanies a melting (there is no change in kinetic energy); about 0.3 eV per atom. In macroscopic units, this corresponds to a change in internal energy $\Delta U = 29\,\mathrm{kJ\,mol^{-1}}$, about half the

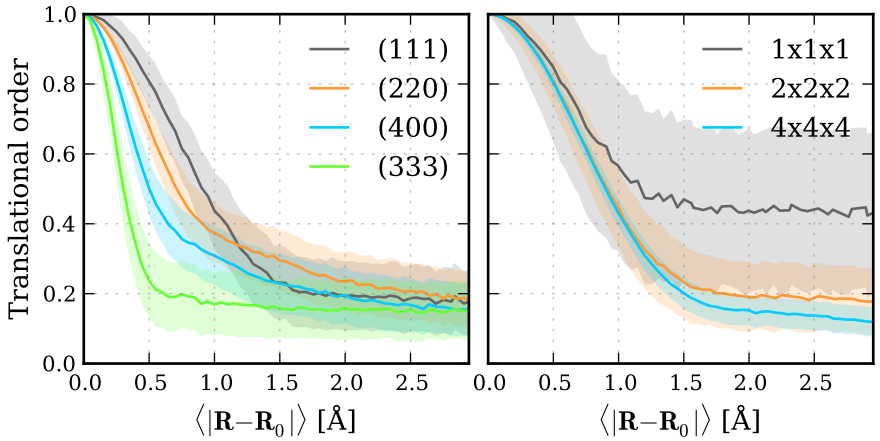
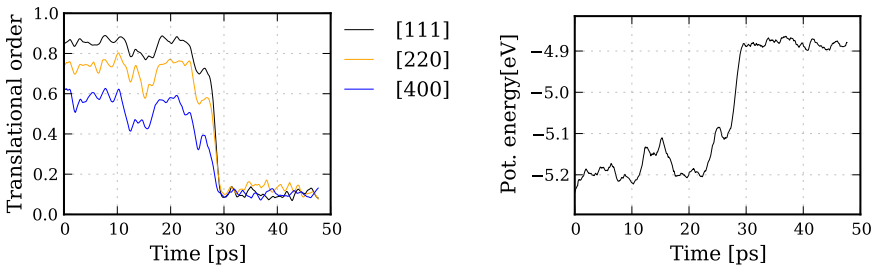


Figure 3.3.: Translational order parameters λ as a function of mean random displacement from lattice sites.



(a) Translational order parameter λ from (3.45).

(b) Potential energy per atom

Figure 3.4.: Translational order and potential energy per atom for a system with 65 atoms at 64 lattice sites (one interstitial) at 1600 K which melts after slightly less than 29 ps.

experimental enthalpy of fusion $\Delta_{\text{fus}}H$ of pure silicon,[7]

An even simpler test than (3.45), introduced by Verlet in 1968, just ignores the imaginary part of (3.45).[85][75, p. 117] This works best with $\mathbf{G}$ vectors that give completely real structure factors for the ideal structure $\mathbf{R}^0$, such as [400], which Verlet used;

$$\lambda' = \frac{1}{F^0_{\mathbf{G}}} \sum_{j=1}^{M} \cos\left(\mathbf{G}_{400} \cdot \mathbf{R}_j\right) = \frac{1}{M} \sum_{j=1}^{M} \cos\left(4\pi u\right) \tag{3.46}$$

This parameter is unity for the perfect lattice, and fluctuate about zero for a completely disordered system. In 1968 the computationally simpler form of (3.46) over (3.45) may have been important, but today it is generally not so. However, there is another interesting distinction between the two tests. While (3.45) is origin-independent, (3.46) is origin-dependent. If all atoms are translated $d/4$ in each direction, the parameter drops to zero, even though the translational order is still perfect. Therefore a comparison of the two tests can act as a quick test to determine if the whole system is drifting or not.

A comparison between the two order parameters is shown in figure 3.5a for a system with 64 silicon atoms and one phosphorus atom at 1400 K (the experimental melting point of pure silicon is 1685 K). Note that the extra phosphorus atom introduces a local distortion into the crystal, but this is quite short-ranged, so the overall translational order of the crystal is still largely preserved, as reflected in the order parameters.

3.4.1. Handling drift of the reference frame

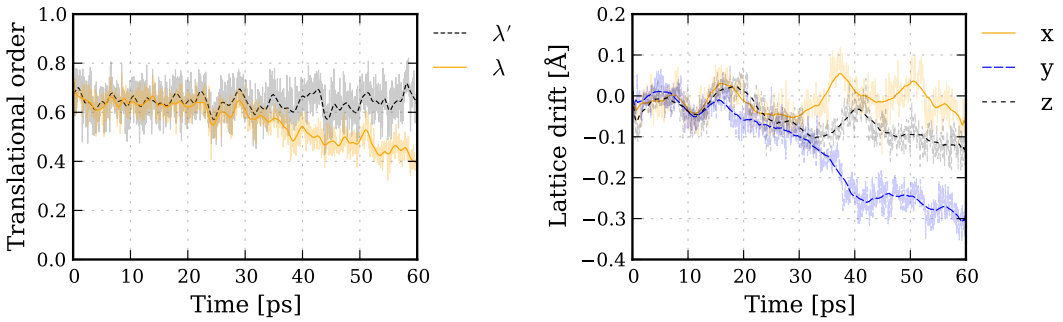
Comparing the two translational order parameter functions, it's seen that the whole system drifts over time in this particular run, although the shift is not easily quantified from this information alone. There is nothing inherently unphysical about the drift, as the total momentum is conserved and there is no drift in the mass centre. If an atom moves 10 Å to the right in a cell with 101 equal atoms, momentum may be conserved in different ways: by another atom moving 10 Å to the left, by two atoms moving 5 Å to the left, or by all the other 100 atoms moving 0.1 Å to the left. Thus the effect is to some extent an effect of the small cell, as a small cell will have a small inertia, making it more vulnerable for concerted movement of all the atoms than a larger cell.

Whether the drift causes a problem for analysis and needs to be subtracted depends on whether a fixed reference frame is expected. Having a fixed reference frame becomes important when comparing the system to another one. In our case we're interested in comparing the system to a reference lattice – the perfect diamond structure, so a method to quantisise the drift was devised: For any given timestep we first assign each atom to its nearest lattice site and then calculate the drift as the directional median distance

[6] From microfiche F.25, available online at http://www.ccl.net/cca/software/SOURCES/FORTRAN/allen-tildesley-book/f.25.shtml See [83, p. 171] for discussion.

[7] To convert to macroscopic units, we multiply by the Avogradro constant, and convert electronvolts to kilograms: $0.3\,\text{eV} \times 6.022 \times 10^{23}\,\text{mol}^{-1} \times 1.60 \times 10^{-22}\,\text{kg eV}^{-1} = 29\,\text{kJ mol}^{-1}$. Enthalpy change $\Delta H = \Delta U + p\Delta V$, but ΔU is normally the dominant term. Experimentally, $\Delta_{\text{fus}}H = 50.2\,\text{kJ mol}^{-1}$.[84, p. 6.130].

between an atom and its lattice site.[8] The result is a vector $\boldsymbol{\delta}(t)$ that can be subtracted from the trajectory or added to the reference lattice after appropriate smoothing over lattice vibrations and minor wiggles; a symmetric running mean over 4 ps was found to work well. Fig. 3.5b shows the components of the drift vectors before (semitransparent lines) and after (solid lines) applying such a smoothing.



(a) Time evolution of translational order parameters λ' from (3.45) and λ from (3.46)

(b) Lattice drift in each direction

Figure 3.5.: Translational order and lattice drift for a system with 65 atoms at 64 lattice sites (one interstitial) at 1400 K.

[8] Note that since we have added an extra atom, there will in general be two atoms assigned to the same site, but if there is no lattice drift the atoms will still be centred about lattice sites.

4. Results and discussion

4.1. Introduction

All calculations were carried out using a 64-atom silicon unit cell with an extra phosphorus atom added. When a name is necessary, the system will be referred to as $Si_{64}P$. All simulations started from a relaxed structure with phosphorus at an interstitial position. The rationale for this choice was given in section 1.3.3.

4.2. Initial structures

Two trajectories with even very minor differences in initial properties (structure and/or velocities), can easily diverge exponentially from each other due to Lyapunov instability.[73, p. 72] Trajectory averages, on the other hand, are relatively stable with respect to such minor differences. However, since many different interstitial and interstitialcy positions have been reported in literature (section 1.3), we will look slightly into the stability of the different structures, to make sure we start from a quite stable initial position.

Structures were prepared and relaxed with phosphorus at the tetrahedral site, the hexagonal site, the "classic" bond-centred site (with C_{3v} symmetry) and the split $\langle 110 \rangle$ site. Special interest was taken in the split $\langle 110 \rangle$ site, since it has been identified as the most stable site in a recent nudged elastic band study.[28] Since there is some freedom in defining the site, several structures were prepared with slight variations to see if they would result in the same minimum-energy structure or not. All relaxations were carried out in a two-step manner using the conjugate gradient method, requiring forces on the atoms less than 0.05 eV/Å.

The tetrahedral site (fig. 4.1a), where the phosphorus atom is surrounded by four nearest neighbours, turned out less stable than anticipated from purely geometrical considerations. On the first relaxation attempts, it relaxed into the more stable hexagonal position (fig. 4.1b), with phosphorus surrounded by six neighbours. It was eventually possible to relax the tetrahedral configuration, but this indicates that the minimum is a very shallow one. Energetically, the tetrahedral site is less favourable than the hexagonal and split $\langle 110 \rangle$ sites by about 1 eV.

The bond-centred $\langle 111 \rangle$ site (fig. 4.1c) is the energetically least favourable of the investigated sites, due to the large distortion effect it has on the neighbouring silicon atoms. In

the relaxed structure, each of the two silicon atoms has moved 0.9 Å away from its original lattice site. It is interesting that the original symmetry is preserved, even though the relaxation was carried out without use of symmetrisation, illustrating the general importance of starting relaxations from slightly asymmetric structures.

The figures 4.1d and 4.1e show the two lowest-energy variants of the split ⟨110⟩ configuration, both close to 0.1 eV lower in energy than the hexagonal configuration. While the two variants are practically equal energetically, there is considerable variation in bond lengths, illustrating the relative freedom available in this configuration.

An interesting case is fig. 4.1f, showing a structure that emerged from relaxing a split ⟨110⟩ variant, and which I will refer to as just the "triangle" configuration. A relaxed split ⟨110⟩ configuration is already shifted in the [110] direction with respect to the shared lattice site, with Si closer to the lattice site than P. The triangle configuration can be viewed as a result of further shifting the pair in the same direction. It can be seen that the structure in figs. 4.1e and 4.1f are not very different, and 4.1e may be viewed as an intermediate structure between 4.1d and 4.1f.

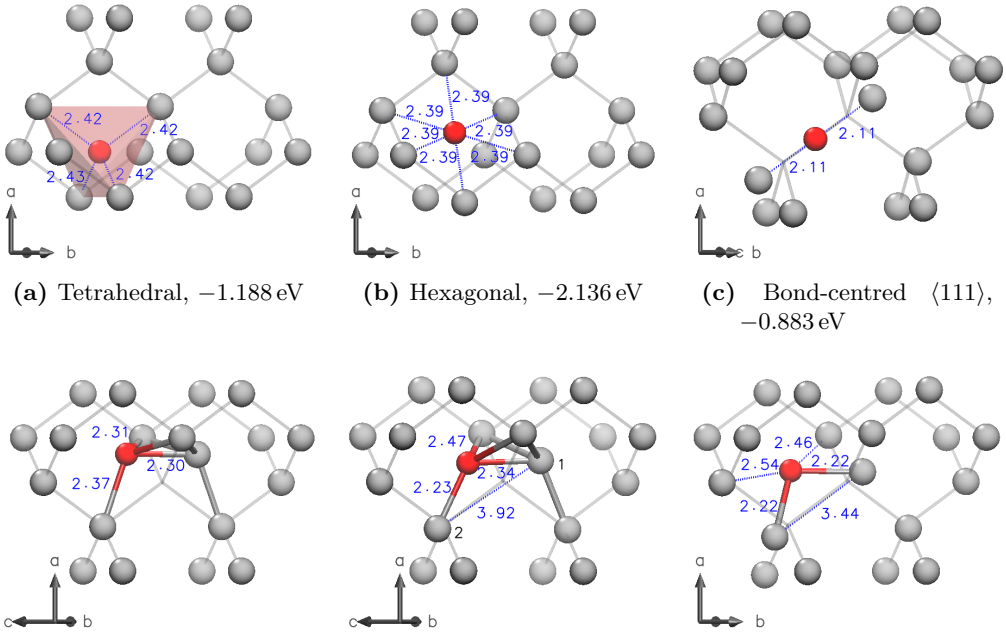
At the point of identifying the triangle structure, I had not yet seen such a structure before, and it was therefore quite exciting to later find a very similar structure reported as the lowest-energy configuration in a NEB study by Harrison *et al.*, referred to as a bond-centre site,[29] a name I will not use, however, since it conflicts with the classic bond-centre site with C_{3v} symmetry (fig. 4.1c). In the structure found by Harrison, the two P-Si bonds are 2.19 Å long, with an angle of 107°, while the structure in fig. 4.1f has P-Si bonds 2.22 Å long, with an angle of 101°. This agreement is satisfying.

Comparing the total energies of the relaxed structures, the triangle configuration is the lowest-energy configuration. This is in agreement with the NEB study by Harrison,[29] while the NEB study by Liu[28] identified a split ⟨110⟩ configuration as minimum. Bond lengths or angles were not reported by Liu, so it's not possible to compare the structure with the structures found here.

4.3. Trajectories

In total, 25 trajectories were simulated at 14 temperatures, ranging from 700 K to 1700 K. In addition, an attempt was made running a simulation at 1750 K, but this system melted almost immediately. At temperatures in the range 1600 to 1700 K, the system could easily melt, but could also remain in a crystalline state for tens of picoseconds. Excluding the melted parts of the trajectories, the total simulated time was 1036 ps.

Note that simulations were started either from a split ⟨110⟩ configuration (fig. 4.1e) or from a hexagonal configuration (fig. 4.1b), not the triangle configuration (fig. 4.1f), due to initial mistakes in assigning the correct minimum-energy structure. Looking at the trajectories, it is however apparent that when P starts in the hexagonal position, it quickly moves into a split ⟨110⟩-like structure, as exemplified in fig. 4.3c → 4.3d, and we could expect the same to happen for the triangle configuration, but it is a bit unfortunate that it was not tested.



(a) Tetrahedral, −1.188 eV **(b)** Hexagonal, −2.136 eV **(c)** Bond-centred ⟨111⟩, −0.883 eV

(d) split ⟨110⟩, −2.214 eV **(e)** split ⟨110⟩, −2.215 eV **(f)** "triangle", −2.384 eV

Figure 4.1.: Relaxed structures with P in interstitial/interstitialcy positions. The light-grey bond-framework is the perfect silicon structure with no defects, while the spheres are the positions of the relaxed atoms, with phosphorus in red and silicon in grey. Bond lengths are given in Ångstrøm, and should be compared to the equilibrium Si-Si bond length in this unit cell, 2.37 Å, not the experimental bond length of 2.35 Å (see section 2.10.1). The energies are total energies relative to the total energy of the 64-atom Si-cell without P. They are not formation energies or bond-energies, and are included only as a relative measure to compare the configurations.

In this section, two specific trajectories will be discussed in some detail, before the more statistically oriented discussion that follows in the next sections. The first trajectory is a 30 ps simulation at 1000 K, in which a single diffusive jump takes place. This trajectory shows many features that are quite typical of most simulations. The second example, at 1300 K, is a more rare case in term of diffusion distance, with phosphorus migrating about 10 Å during the 90 ps simulation time.

4.3.1. Example 1

Figure 4.2 shows some snapshots and plots from a trajectory at 1000 K, where phosphorus starts in a split ⟨110⟩ position, sharing the site labelled 2 with a Si atom. During the first 18 ps, phosphorus moves between two split positions, centred about site 1 (fig. 4.2d) and 2 (fig. 4.2c), respectively, and the intermediate position is similar to the "triangle configuration". Such a movement is seen in many of the simulations, especially at low temperatures, with Si-P-Si triangle-like structure moving in a quite correlated way back and forth between two lattice sites. In this example, the mean Si-P-Si angle is about 112°.

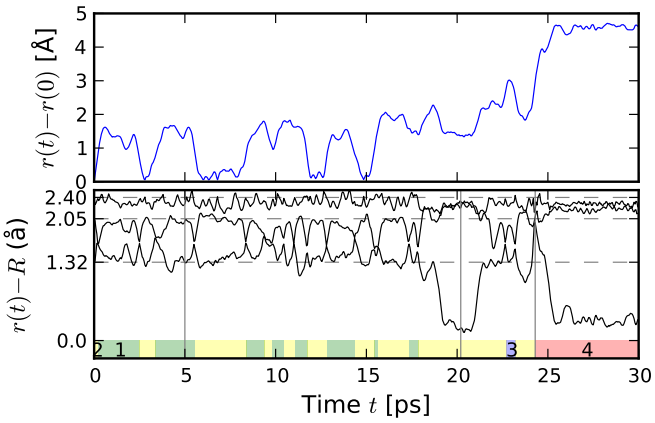
The stability of the Si-P-Si structure is eventually broken, with P first jumping into lattice site 2 (fig. 4.2e), pushing the Si atom into forming a Si-Si pair, which after a few picoseconds returns to push P through a hexagonal site (fig. 4.2f) into a new lattice site 4 (fig. 4.2g).

The complete phosphorus trajectory is shown in fig. 4.2b, while fig. 4.2a shows its displacement $r(t)$ (above), and its distance to the three nearest lattice sites (below), with dashed lines indicating the distances in the initial structure, and the closest lattice site at any given time is indicated by the colour-coded band at the bottom. Vertical grey lines indicate the times at which the snapshots in fig. 4.2c–4.2g were taken.

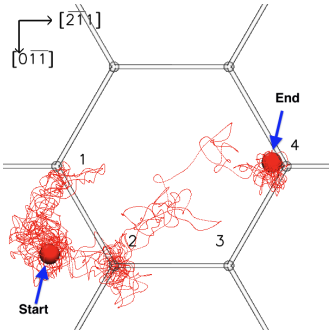
The lattice site distance plot illustrates quite well the jumping back and forth that takes place in the first 18 ps. Interestingly, P is never stable in the middle position, the "triangle" position. It is possible that this is a bias arising from starting in the split ⟨110⟩, but it seems a bit unlikely for the bias to remain over such a long timespan, since the Si-P-Si system is far from isolated.

4.3.2. Example 2

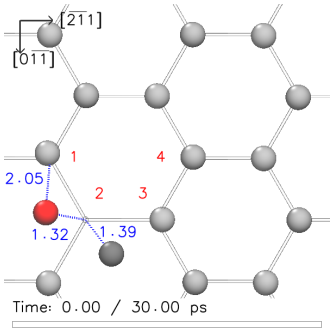
Figure 4.3 shows some snapshots and plots from a trajectory at 1300 K. Here, phosphorus starts in a hexagonal position (fig. 4.3c), but immediately (in less than 200 fs) enters a split ⟨110⟩-like position (fig. 4.3d). This is typical for the trajectories that started from the hexagonal configuration. Interestingly, some of the Si-P-Si dynamic seen in the previous example, can also be seen here during the first 5 ps, indicating that this type of dynamics is not an effect of the initial structure. During this timespan, the mean angle for the Si-P-Si structure is about 115°.



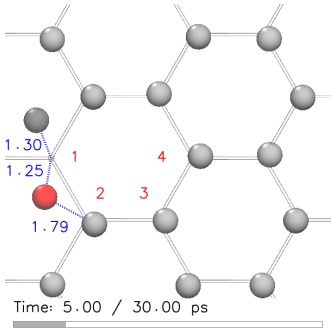
(a) Displacement $r(t)$ and distance to nearest lattice site R for the P atom.



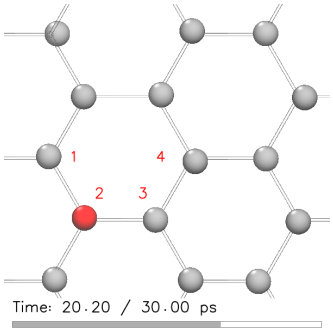
(b) The full P trajectory



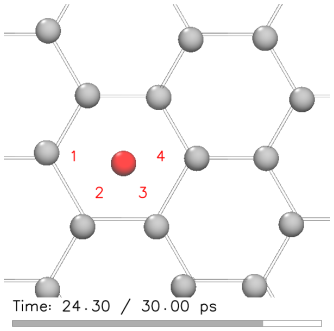
(c) Relaxed initial structure, with P at split ⟨110⟩ site.



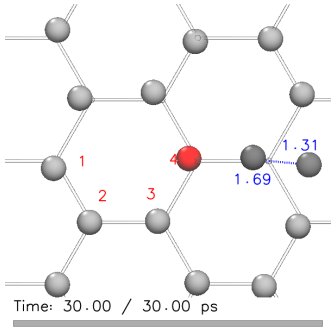
(d) Si-P pair sharing site 1.



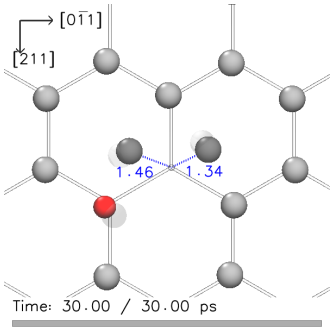
(e) P at site 2. A Si-Si pair shares the adjacent site in the plane above (not depicted).



(f) P at intermediate position before jumping to site 4.

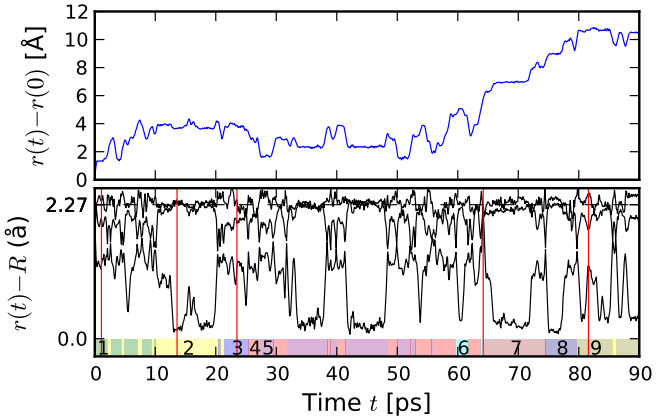


(g) P at site 4. Si-Si pair sharing adjacent site.

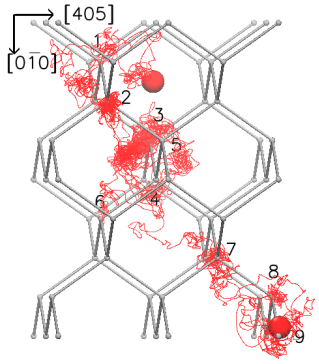


(h) Relaxed final structure. Note: along $[1\bar{1}\bar{1}]$ direction.

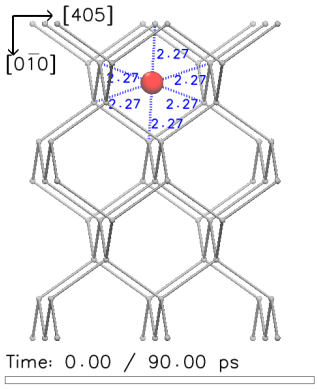
Figure 4.2.: 30 ps simulation at 1000 K. (c-g) shows 5 selected configurations viewed along the $[\bar{1}1\bar{1}]$ direction, indicated on (a) by vertical grey lines. (h) shows the final structure after geometry relaxation. Interstitial silicon atoms are drawn with a darker shade than silicon atoms at lattice sites. Vibrations have beeen smoothed out by a 0.5 ps symmetric mean.



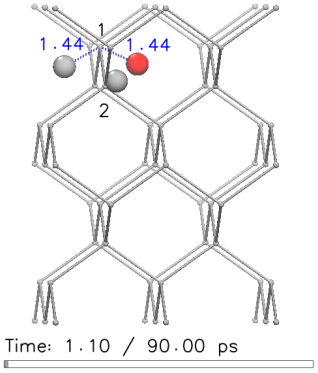
(a) Displacement $r(t)$ and distance to nearest lattice site R for P atom.



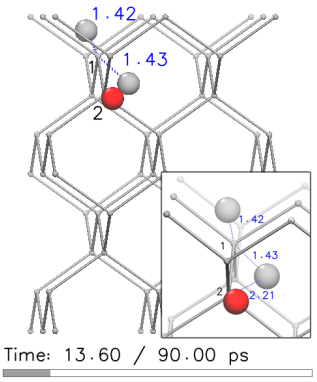
(b) The full P trajectory



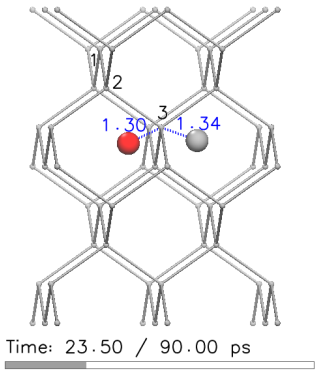
(c) Relaxed initial structure, with P at hexagonal site.



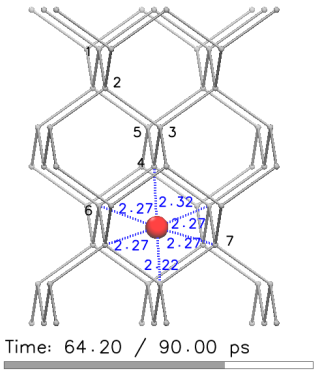
(d) P-Si sharing site 1



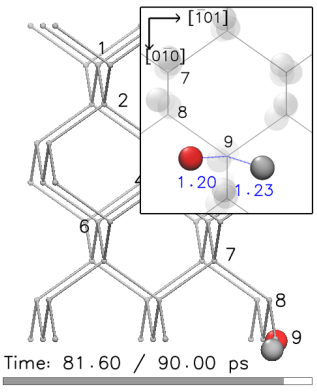
(e) P at site 2. Si-Si sharing site 1



(f) P-Si sharing site 3



(g) P at hexagonal site on its way from site 6 to 7



(h) P-Si split at site 9

Figure 4.3.: 90 ps simulation at 1300 K. (c-f) shows 6 selected configurations, indicated on (a) by vertical grey lines. Only interstitial silicon atoms are shown.

After 5 ps, the P atom makes a very short visit to lattice site 1 before returning to a rotated split ⟨110⟩-like configuration. At 9 ps it approaches the Si atom at lattice site 2, and effectively changes place with it, through a process that involves many steps and takes about 10 ps. In the lattice site distance plot in fig. 4.3a it is seen that the P atom first enters the lattice site at about 12 ps (fig. 4.3e), but then retracts. This is part of the process where P and Si change places. At about 18 ps, the change of places is complete, and the P atom can diffuse further in the same direction.

I will not attempt to describe the remainder of the trajectory, since the dynamics is quite complex, but I have included a snapshot showing a split ⟨110⟩-like configuration (fig. 4.3f), which is seen all the time, and also an intermediate hexagonal position (fig. 4.3f), which the P atom rapidly passes through from one site to another, just as in the previous example.

It should perhaps be stressed that these snapshots are just snapshots, and should thus not be taken as particularly representative for phosphorus diffusion in silicon in general. Studying single trajectories may provide some intuitive familiarity with the system, but with such complex dynamics there is also inherent a great danger of seeing what you expect to see.

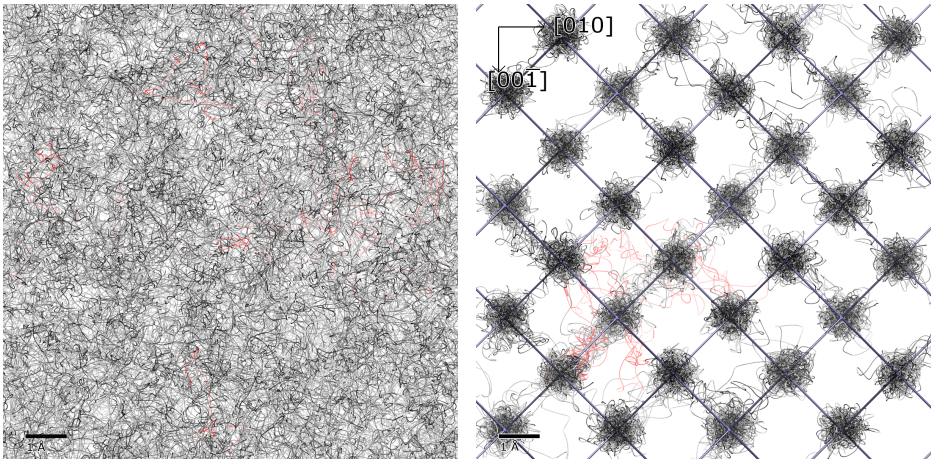
4.4. Melting

The lowest temperature at which the system melted was 1600 K (1330 °C), making this an upper limit for the melting point of the system. It is likely that the melting point is lower though, since a system near its melting point can remain in a metastable state for a long time, even on a macroscopic timescale. That no melting was seen at 1550 K may be due to the temperature being below the melting point, or just as well due to metastability.

Figure 4.4a shows an example of a melted system at 1600 K; the lines are 20 ps trajectories of the 65 atoms in the system, starting after the system melted from its original crystalline state. It is evident that no long scale order is present, and to be sure the lattice order parameter (3.45) is ~ 0.1. If we were to follow a single atom on its journey, however, we would find some local order, that can be revealed from e.g. the radial pair distribution function, discussed in section 4.4.1 below.

The difference to the crystalline state, exemplified with 20 ps trajectories in fig. 4.4b, is of course striking. The example is from a simulation at 1700 K, the highest temperature for which the system remained in a crystalline state for a considerable time. Yet, while the system depicted here could melt at any moment, the trajectories are not qualitatively different from the trajectories obtained at a temperature below the melting point. The "nests" are larger due to the higher kinetic energy, but they are still centred about points, and the points are part of a lattice, indicated by the grid for which the lattice order parameter (3.45) is > 0.8, using $\mathbf{G}_{111}$.

Techniques for determining the melting point from a molecular dynamics simulation are discussed e.g. by Yoo.[86] A direct approach involves starting from a system where half



(a) Liquid state at 1600 K (kinetic energy 204 meV/atom)

(b) Crystalline state at 1700 K (kinetic energy 216 meV/atom)

Figure 4.4.: Two 20 ps trajectories of the $Si_{64}P$ system, viewing along the 001 direction, at the same scale; the scale bar in the lower left is 1 Å.

the cell is crystalline, and the other half is liquid. Using a NPH ensemble, the two parts may even start at different temperatures, and then equilibrate. Using this approach, Yoo found a melting point for pure silicon $T_m = 1540 \pm 50$ K, and also noted that other DFT-based studies have underestimated the melting point.[86]

With pure silicon experimentally melting at 1687 K, and with only a very modest reduction when phosphorus at doping concentrations is added,[20] our upper limit of 1600 K indicates that the melting point is underestimated in this work as well.

4.4.1. Short-scale order in the melt

Upon melting, long-scale order is lost, while some short-scale order remains. This can be seen in the radial pair distribution function $g(r)$, defined as the ratio of the average local atomic (or molecular) density $\rho(r)$ at a distance r from an arbitrary atom (or molecule), to the bulk density ρ of the liquid;[87]

$$\rho(r) = g(r)\rho \tag{4.1}$$

For any given atom in a simulation, the function can be found from counting the number of atoms $N(r, r+\Delta r)$ within surrounding shells of thickness Δr, and average the numbers over the timespan of the simulation. The shell thickness Δr must be chosen thin enough to capture the structure of the function, while thick enough to provide sufficiently large sampling populations for statistically reliable results. A value near 1/40 the average interatomic distance is often satisfactory.[75, p. 263] We have

$$g(r) = \frac{\langle N(r, r+\Delta r)\rangle}{\rho V(r, \Delta r)} \tag{4.2}$$

where $V(r, \Delta r) = \frac{4}{3}\pi[(r+\Delta r)^3 - r^3)]$ is the volume of a shell with radius r and thickness Δr. The function falls to zero for $r \to 0$ due to repulsive forces between atoms, and approaches unity for $r \to \infty$ quite quickly, since there is no long range order in liquids.

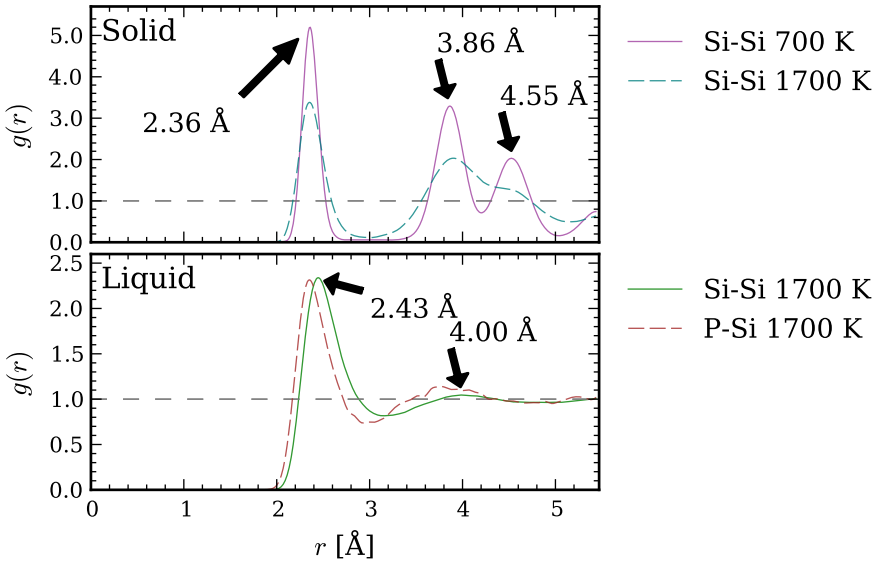


Figure 4.5.: Si-Si radial pair distribution function for two simulations of a solid system at different temperatures (upper), and comparison of the Si-Si and P-Si radial pair distribution functions for the same liquid system (lower), time-averaged over 10 ps in both cases. The annotated peaks in the lower figure refer to the Si-Si function. Note also the different scales on the ordinate axes in the upper and lower figure.

The upper part of fig. 4.5 shows $g(r)$ for the crystalline system at the lowest and highest temperatures simulated. In an ideal, homogeneous liquid, the radial pair distribution function would be independent of the choice of atom, but the addition of phosphorus gives rise to a slight inhomogeneity and dependence of selection. To avoid selection bias, and improve statistics at the same time, the function was therefore averaged over all the 64 silicon atoms. At 700 K, the nearest neighbour peak is centred at 2.36 Å, the second nearest neighbour peak at 3.86 Å, and the third-nearest neighbour peak at 4.55 Å.[1] At 1700 K, the distribution is more smoothed out due to larger vibrations, but otherwise the same structure can be seen.

The lower part of the figure compares the P-Si and Si-Si radial pair distribution functions for a system in the liquid state at 1700 K. The P-Si function is more jagged than the Si-Si function, since the sampling population for the P-Si function is only 1/64 the sampling population of the Si-Si-function. Apart from that, the functions are similar, with the P-Si function shifted slightly to the left to the Si-Si function.

Table 4.1 compares the peaks in the Si-Si pair correlation function found here with values found in the literature for pure silicon.

[1] For simplicity, the function was only plotted to a radius 5.94 Å, half the side length of the unit cell.

Only the first two peaks are included, since the third peak is not fully resolved in this work. While the first peak compares quite well with literature values, the discrepancy in the second peak is considerable, much larger than what should be expected just from the addition of a phosphorus atom. One possible reason for the discrepancy is that the current work was carried out using the canonical (NVT) ensemble, not allowing for volume change. Experimentally, melting of silicon is associated with a density increase of about 10 %.[88] In their MD studies, Stich *et al.* used a density $2.59\,\mathrm{g\,cm^{-3}}$, about 10 % denser than the density of our system; $2.32\,\mathrm{g\,cm^{-3}}$.

Source	Temp. [K]	1. peak [Å]	2. peak [Å]	Coordination no.
This work	1700	2.43	4.00	5.5 or 6.1[a]
Waseda (1975)[89][b]	1733	2.50	3.78	6.4
Stich (1991)[90][c]	1800	2.46	3.70	6.5
Stich (1996)[91][d]	1700	2.47	3.64	6.2
Ansell (1998)[92][b,e]	1603	2.42	3.27 3.95	5.6 ± 0.5
Ansell (1998)[92][b]	1829	2.46	3.45	6.4 ± 0.5

[a] See discussion in text
[b] X-ray diffraction study
[c] Ab initio Car-Parinello MD at LDA-level
[d] Ab initio Car-Parinello MD at spin-polarised GGA-level
[e] supercooled

Table 4.1.: Comparison of silicon radial pair distribution function features (units are Å)

Coordination numbers are found from $g(r)$ by

$$n(r_1, r_2) = 4\pi\rho \int_{r_1}^{r_2} g(r) r^2 \, dr, \tag{4.3}$$

In particular, the integration carried out from zero to the first minimum, gives what is known as the number of nearest neighbours $n_1(r') = n(0, r')$.[88, p. 7] The integration process is illustrated in fig. 4.6. As expected, this yields approximately four nearest neighbours when integrating to the deep minimum found for a system in the crystalline state. For the shallow minimum found in a liquid state, however, there is considerable disambiguity in how to evaluate the integral. In fig. 4.6, it was evaluated to the first minimum in $g(r)r^2$, for which $n_1(3.04) = 5.5$. Evaluated to the first minimum in $g(r)$ instead, we obtain $n_1(3.18) = 6.1$. With this kind of sensitivity, it's hard to make much sense about the absolute numbers, but relatively, the numbers found here are lower than other values found in the literature, table 4.1, which may again be due to the differing density.

The radial pair distribution function is interesting in part because it's readily obtainable from experiment, using x-ray diffraction. A molecular dynamics simulation provides, in addition, the microscopic data that makes up the function. In fig. 4.7, the distance between a silicon atom and its 16 nearest neighbours (4 in 'first shell', 12 in 'second shell') is plotted for each of the 64 silicon atoms, with each distance plotted as a small dot. The data is from a 50 ps simulation at 1600 K, for which the system remained crystalline for about 30 ps. It can be seen that the first 'band' (in black), arising from the four nearest

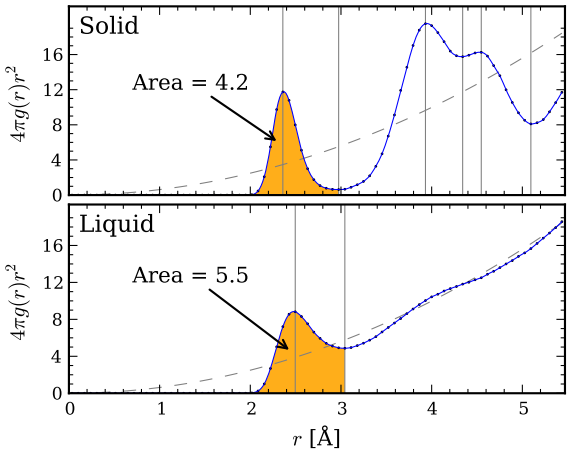


Figure 4.6.: Coordination numbers $n(r')$ from integrating $4\pi\rho g(r)r^2$ for a Si_6P system at 1600 K in solid (above) and liquid (below) phase. The parabolic function $4\pi r^2\rho$ is shown by dashed lines.

neighbours, is somewhat preserved upon melting. This is of course the first peak in the radial pair distribution function. The disambuigity in defining the integration limit in (4.3), discussed above, is also made clear here.

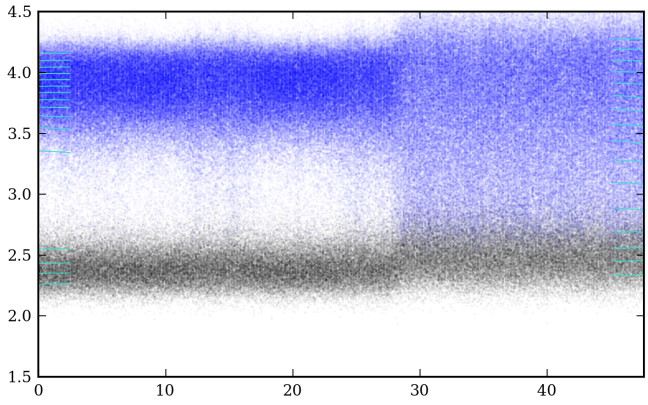


Figure 4.7.: Nearest neighbour distance for the 16 nearest neighbours, for each of the 64 silicon atoms, with black dots for the nearest 4 atoms, and blue dots for the next 12. From a simulation at $T = 1600\,\mathrm{K}$ which melted at ~ 30 ps. Average distances are drawn as lines at the edges only.

4.5. Diffusion

Table 4.2 lists phosphorus and silicon mean displacements $\langle \Delta r \rangle = \left\langle \sqrt{\Delta r^2} \right\rangle$ and diffusivities $D = \langle \Delta r^2 \rangle / 6t$, for all the simulations carried out. The angle brackets indicate

averages over atoms of the same type (1 for phoshorus, 64 for silicon), $\Delta r^2 = [\mathbf{r}(t) - \mathbf{r}(0)]^2$ are mean square displacements, and t is either the end of the simulation, or the point at which the system melted, whichever came first, since only diffusion in the solid state is considered here.

Temp (K)	t (ps)	Δr_P (Å)	$\langle \Delta r_\mathrm{Si} \rangle$ (Å)	D_P (10^{-6} cm^2/s)	D_Si (10^{-6} cm^2/s)
700	30.0	2.4	0.1	3.32	0.02
800	30.0	2.1	0.1	2.56	0.05
900	28.5	1.2	0.1	0.78	0.02
1000	30.0	4.6	0.2	11.55	0.07
1100	30.0	1.0	0.1	0.56	0.03
1200	30.0	1.3	0.1	0.95	0.05
1300	30.0	1.8	0.2	1.78	0.09
1300	90.0	10.5	0.3	20.38	0.04
1400	60.0	2.9	1.1	2.34	0.90
1400	51.6	3.4	0.5	3.78	0.46
1400	80.0	3.6	1.3	2.71	0.72
1400	90.0	1.9	1.3	0.70	0.81
1450	60.0	4.3	1.2	5.20	1.21
1500	60.0	8.3	0.8	19.15	0.79
1500	80.0	5.1	1.3	5.38	1.03
1500	60.0	6.6	0.8	11.92	0.63
1550	15.0	2.3	0.4	5.64	1.06
1550	60.0	4.3	1.0	5.23	0.77
1600	9.1*	5.1	1.4	46.64	5.15
1600	28.9*	7.1	1.9	29.36	3.63
1600	20.0	2.2	0.2	3.93	0.24
1600	28.4*	4.6	1.8	12.56	3.02
1650	8.5*	3.3	1.4	21.62	5.50
1700	6.2*	1.7	1.2	7.61	6.27
1700	19.8	1.3	0.6	1.32	1.03

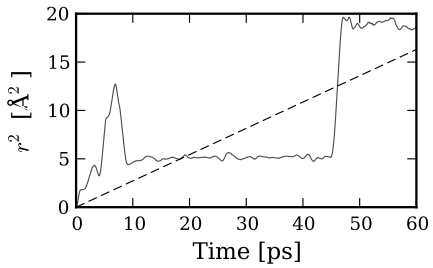
Table 4.2.: Mean displacements and diffusion coefficients for phosphorus (P) and silicon (Si) in the simulations carried out. The times t are the durations of each simulation, or the duration the system remained crystalline (systems that melted are indicated by *).

There are several issues with calculating diffusivities from the simulation data shown in the above table. One is the small magnitudes of the displacements. With phosphorus starting in an interstitial site, a displacement of $\sim 2.3\,$Å constitutes only a movement from the interstitial site to the neighbouring lattice site (see fig. 4.1). Such a displacement, which was the largest displacement that took place in many of the simulations, cannot by itself be related to macroscopic diffusion.

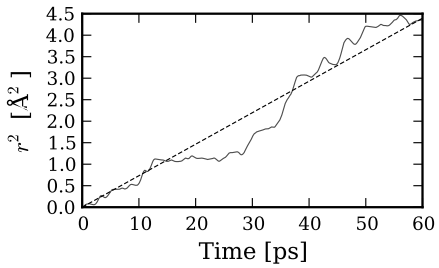
Related to this issue, is the issue that the relation $D = \Delta r^2/6t$ requires $r^2(t)$ to be linear, but with few jumps taking place, $r^2(t)$ is generally not very linear. Fig. 4.8a illustrates this, showing a quite representative simulation, where phosphorus first jumps from the

initial interstitial site to the neighbouring lattice site, and then later to another lattice site. While a linear fit can be made, the slope depends quite strongly on the (random) choice of endpoint. If the simulation for instance had ended after 50 ps, not after 60 ps, the slope would have been steeper. Several more jumps would be needed to reliably assign D from $\langle r^2(t)\rangle$.

As seen in fig. 4.8b, the statistics for silicon is better, since the diffusion accumulates over all the 64 atoms. This is why calculating the self-diffusion coefficient is a more tractable problem than calculating the diffusion coefficient of a dopant.



(a) Phosphorus (Pearson's correlation coefficient is 0.71)



(b) Silicon (Pearson's correlation coefficient is 0.97)

Figure 4.8.: Mean square displacement $\langle \Delta r^2 \rangle$ of phosphorus and silicon from a simulation at 1450 K. Please note the different scales of the ordinate axes.

Figure 4.9 is an Arrhenius plot with the diffusivities from table 4.2 plotted on a log scale as a function of the inverse temperature. The least square fit for the silicon diffusivities is better than the fit for the phosphorus diffusivities, due to the better statistics. From the fits, an activation energy of 0.17 eV was found for phosphorus, and 0.58 eV for silicon. The values are not terribly far off values reported in the literature, but the fits are not particularly good, and it is clear that longer simulations, or more simulations at each temperature, would be needed to somehow reliably determine migration energies.

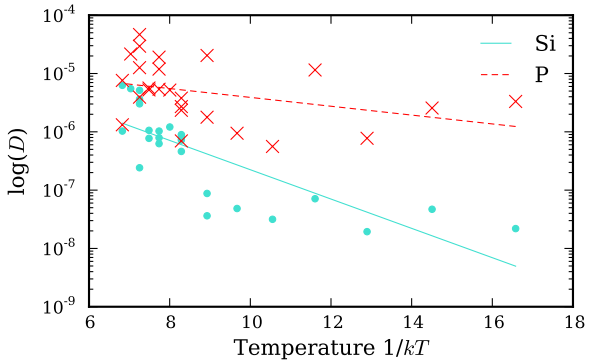


Figure 4.9.: Arrhenius plot of diffusivities of phosphorus and silicon, with attempted least square fits. Pearson's correlation coefficient is -0.36 for phosphorus and -0.79 for silicon. From the slopes of the fitted lines, the activation energy is 0.17 eV for phosphorus and 0.58 eV for silicon.

4.6. Jump analysis

Molecular dynamics provides very detailed information about the movements of the simulated atoms, but the quite chaotic motion that results, especially at temperatures close to the melting point, can be hard to interpret directly. An attempt was therefore made to analyse the data in terms of atomic jumps between lattice sites. To do this, we must first define what is meant by an atom sitting at a lattice site.

In the lattice site distance plots 4.3a and 4.2a for the example trajectories in section 4.3, we calculated the distance between an atom and its nearest lattice site, and the nearest site was identified by an index. In doing this, we calculated the distance between the given atom and all the 64 sites in the ideal silicon lattice, and corrected for drift using the method described in section 3.4.1.

Having obtained the distance to the nearest lattice site, we must define some threshold for what it means for an atom to sit on a lattice site. Figure 4.10 shows an example result from an automated procedure for identifying when the atom sits at a lattice site with threshold 0.5 Å, and also for identifying melting using the translational order parameter. Using this procedure, figure 4.11 shows the time the P atom spends at lattice sites in all the simulations using the given threshold.

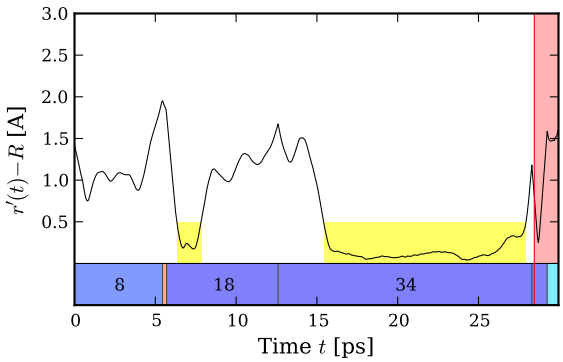


Figure 4.10.: Distance to the nearest lattice site for the P atom in a simulation at $T = 1400$ K, with shaded (yellow) regions indicating a distance less than the threshold distance (0.5 Å). The nearest lattice site is identified by an index. The shaded (red) region to the right indicates melting, and should be excluded from analysis.

More interesting perhaps, we can find the amount of time τ the P atom remains in an interstitial position, before moving to a substitutional site, kicking out a Si atom. This has been tabulated in table 4.3. At the lowest temperatures, phosphorus remains in interstitial positions for the whole simulation, but at temperatures of 900 K and higher, phosphorus almost always enters a lattice site during the timespan of the simulation. This is an activated process, so an attempt was made to determine the activation energy from the slope of $1/\langle\tau\rangle$ as a function of $1/k_BT$, shown in fig. 4.12, leaving out the simulations where the P atom remained in interstitial positions during the entire simulation, and also the simulation at 1450 K, since only one simulation was carried out at this temperature. The slope of the fitted line corresponds to an activation energy of 0.26 eV, but since

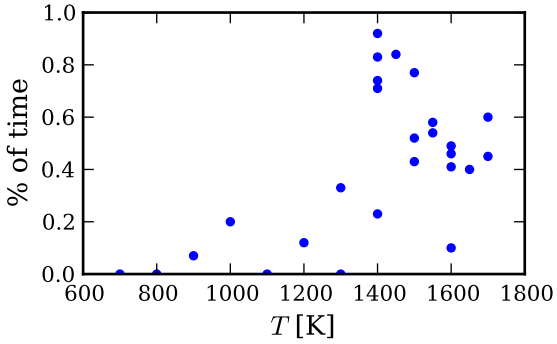


Figure 4.11.: Average percentage of the simulation time the P atom spends at lattice sites (threshold 0.5 Å).

simulations started from different initial configurations, its possible that the value is a mean value for slightly different mechanisms.

If a reliable value could be obtained, using data from more simulations with the same initial configuration, an interesting next question would be how the value compares to the corresponding time for a silicon atom.

T [K]	τ [ps]	T [K]	τ [ps]	T [K]	τ [ps]	T [K]	τ [ps]
700	> 30	1300	> 30	1500	13.25	1600	6.35
800	> 30	1400	3.75	1500	4.60	1600	6.70
900	22.65	1400	37.90	1500	7.90	1650	4.35
1000	19.35	1400	4.50	1550	6.00	1700	1.15
1100	> 30	1400	6.10	1550	6.85	1700	10.30
1200	14.35	1400	9.60	1600	2.45		
1300	12.75	1450	2.30	1600	5.80		

Table 4.3.: Time τ in ps before the P atom jumps from the initial interstitial site to a substitutional site.

Finally, it was attempted to quantify the total number of jumps in the simulations. This turned out not to be very straightforward though.

Here, an atom A is defined to *sit* on a lattice site X when (a) X has been the nearest lattice site to A for more than $t_{\min}$ ps, and (b) the distance between X and A has been less than $x_{\max}$ (not necessarily for any long time). Having defined what it means to *sit* on a lattice site, an atom *jumps* simply when it changes from sitting at one site to another.

Two different attempts were carried out for calculating the number of jumps for the single phosphorus atom. In the first "liberal" attempt, the distance criterion is ignored altogether, but the time criterion is set to 0.3 ps. This should suffice to avoid counting particularly long vibrations as jumps, but still be short enough to avoid missing jumps if multiple jumps are made in series. In the second "conservative" attempt, a distance

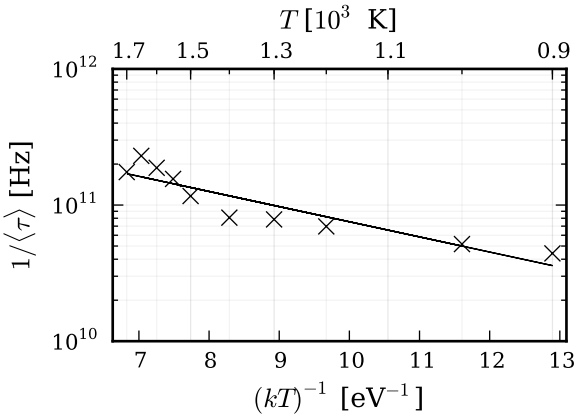


Figure 4.12.: Arrhenius plot for interstitial to lattice site jump The Pearson correlation coefficient is -0.92.

criterion of 0.3 Å is imposed. This should filter out most cases where the atom jumps from one interstitial position to another.

Table 4.4 shows the results of applying the two attempts for all the runs, illustrating that very different results are obtained with different jump definitions.

Temp. [K]	Duration [ps]	Liberal estimate [a] Jumps	Liberal estimate [a] Jump frequency [/ps]	Conservative estimate [b] Jumps	Conservative estimate [b] Jump frequency [/ps]
700	30	23	0.767	0	0.000
800	30	20	0.667	0	0.000
900	28	19	0.668	2	0.070
1000	30	19	0.633	1	0.033
1100	30	24	0.800	0	0.000
1200	30	13	0.433	2	0.067
1300	90	43	0.478	12	0.133
1300	30	19	0.633	0	0.000
1400	90	10	0.111	3	0.033
1400	80	4	0.050	1	0.013
1400	160[c]	27	0.169	5	0.031
1400	60	11	0.183	4	0.067
1400	51	34	0.659	0	0.000
1450	60	5	0.083	2	0.033
1500	60	20	0.333	4	0.067
1500	80	15	0.188	1	0.013
1500	60	26	0.433	5	0.083
1550	15	5	0.333	0	0.000
1550	60	19	0.317	5	0.083
1600	47	35	0.734	5	0.105
1600	20	4	0.200	1	0.050
1600	30	10	0.333	3	0.100
1600	49	42	0.852	5	0.101
1650	15	9	0.600	1	0.067
1700	19	9	0.454	1	0.050

[a] In the liberal estimate, $t_{\min} = 0.3$ ps and $x_{\max} = \infty$

[b] In the conservative estimate, $t_{\min} = 0.5$ ps and $x_{\max} = 0.3$ Å

[c] This run was carried out using a timestep three times larger than the other runs (3 fs vs. 1 fs). It may not be directly comparable to the other runs.

Table 4.4.: Jumps and jump frequencies of phosphorus in the simulations carried out.

4.7. Final discussion

The use of *ab initio* molecular dynamics for studying systems in the solid state is still quite rare, due to the high computational cost involved. This work was started without knowing *a priori* how long time spans could be simulated, nor how long time spans would be needed to provide reliable statistical results. Use of new PAW potentials and parameter optimisation showed that only 13 CPU-seconds was needed per ionic step for a 65-atom system. This led to the generation of about one nanosecond of simulation data using a quite modest amount of computation time.

Yet a good estimate of the migration energy could not be obtained. This might be due to simulation times being too short, but perhaps equally well because it's not really possible to sample just one mechanism in molecular dynamics. Once the P atom makes a move to a substitutional site, we are generally not sampling the migration mechanism anymore.

Molecular dynamics is therefore best suited for finding the overall diffusion parameters, the same found from experiment, but an activation energy of > 3 eV is out of reach when using an *ab initio* potential, at least for diffusion of dopants, where a large number of atoms needs to be simulated, while only data from a single atom (the dopant) can be used for analysis.

Simulating the overall diffusion process should, however, be perfectly possible using an empirical potential. A parametrisation for phosphorus in silicon does probably not exist, but once such a parametrisation is developed (or found), for instance on the Tersoff form, the simulation data obtained in this work could be used to test many aspects of the potential.

Considered the difficulty of isolating a single mechanism in a molecular dynamics, the analysis of the initial interstitial-to-substitutional time τ appeared to be quite successful. With a more carefully designed setup (equal initial structures) and more, but shorter simulations per temperature, it is possible that similar analyses could provide reliable jump frequencies or activation energies for some jump mechanisms. This could in turn be useful as input to Kinetic Monte Carlo simulations.

Bibliography

1. Sahli, B. & Fichtner, W. Ab initio molecular dynamics simulation of self-interstitial diffusion in silicon. *Phys. Rev. B* **72,** 245210 (Dec. 2005).
2. Stillinger, F. H. & Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B* **31,** 5262–5271 (1985).
3. Tersoff, J. New empirical model for the structural properties of silicon. *Phys. Rev. Lett.* **56,** 632–635 (Feb. 1986); Tersoff, J. Empirical interatomic potential for silicon with improved elastic properties. *Phys. Rev. B* **38,** 9902–9905 (Nov. 1988); Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Phys. Rev. B* **39,** 5566–5568 (Mar. 1989).
4. Powell, D., Migliorato, M. A. & Cullis, A. G. Optimized Tersoff potential parameters for tetrahedrally bonded III-V semiconductors. *Phys. Rev. B* **75,** 115202 (11 Mar. 2007).
5. Posselt, M., Gao, F. & Bracht, H. Correlation between self-diffusion in Si and the migration mechanisms of vacancies and self-interstitials: An atomistic study. *Phys. Rev. B* **78,** 035208 (2008).
6. Schroder, D. K. & Meier, D. L. Solar Cell Contact Resistance – A Review. *IEEE Trans. Electron Dev.* **ED-31,** 637–647 (1984).
7. Bentzen, A., Holt, A., Christensen, J. S. & Svensson, B. G. High concentration in-diffusion of phosphorus in Si from a spray-on source. *J. Appl. Phys.* **99,** 064502 (2006).
8. Fahey, P. M., Griffin, P. B. & Plummer, J. D. Point defects and dopant diffusion in silicon. *Rev. Mod. Phys.* **61,** 289–384 (Apr. 1989).
9. Hu, S. M. Nonequilibrium point defects and diffusion in silicon. *Mater. Sci. Eng. R* **13,** 105–192 (Oct. 1994).
10. Pichler, P. *Intrinsic point defects, impurities, and their diffusion in silicon* (Springer, 2004).
11. Smakula, A. & Kalnajs, J. Precision determination of lattice constants with a Geiger-counter x-ray diffractometer. *Phys. Rev.* **99,** 1737–1743 (1955).
12. Okada, Y. & Tokumaru, Y. Precise determination of lattice parameter and thermal expansion coefficient of silicon between 300 and 1500 K. *J. Appl. Phys.* **56,** 314 (1984).
13. Fourier, J. *The Analytical Theory of Heat* (1822); Fourier, J. *The Analytical Theory of Heat* trans. by Freeman, A. (Dover, 1955); Adiutori, E. F. *Mech. Eng.* 30 (Aug. 2005).

14. Ohm, G. S. *Die galvanische Kette, mathematisch bearbeitet* <http://www.ohm-hochschule.de/bib/textarchiv/Ohm.Die_galvanische_Kette.pdf> (visited on 26/04/2011) (Berlin: T. H. Riemann, 1827); Ohm, G. S. *The galvanic circuit investigated mathematically* English translation. Trans. by Francis, W. (New York : Van Nostrand, 1969).

15. Fick, A. Über diffusion. *Ann. Phys.* **94,** 59 (1855); Fick, A. On liquid diffusion. *Phil. Mag.* **10,** 30–38 (1855). Abstracted by the author from the German original in Ann. Phys. Reprinted in J. Membrane Sci. 100, 33.

16. Einstein, A. Über die vond der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **17,** 549–560 (1905); Einstein, A. in *Investigations on the theory of the Brownian motion* (ed Fürth, R.)trans. by Cowper, A. D. (Dover, 1926, reprinted in 1956). ISBN: 0-486-60304-0. <http://users.physik.fu-berlin.de/~kleinert/files/eins_brownian.pdf> (visited on 27/04/2011).

17. Von Smoluchowski, M. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. Phys.* **326,** 756–780 (1906).

18. Tilley, R. J. D. *Understanding solids* (Wiley, 2008). Reprint of the 2004 edition.

19. Predel, B. in *SpringerMaterials - The Landolt-Börnstein Database* (ed Madelung, O.) chap. Group IV Physical Chemistry (). doi:10.1007/10542753_2380.

20. Safarian, J. & Tangstad, M. Phase diagram study of the Si-P system in Si-rich region. *J. Mater. Res* **26,** 1494–1503 (2011).

21. Tannenbaum, E. Detailed analysis of thin phosphorus-diffused layers in p-type silicon. *Solid-State Electron.* **2,** 123 (1961).

22. Solmi, S. & Nobili, D. High concentration diffusivity and clustering of arsenic and phosphorus in silicon. *J. Appl. Phys.* **83,** 2484–2490 (1998).

23. Bentzen, A. *Phosphorus Diffusion and Gettering in Silicon Solar Cells* PhD thesis (University of Oslo, 2006).

24. Armigliato, A., Nobili, D., Servidori, M. & Solmi, S. SiP precipitation within the doped silicon lattice, concomitant with phosphorus predeposition. *J. Appl. Phys.* **47,** 5489–5491 (1976).

25. Solmi, S. *et al.* Dopant and carrier concentration in Si in equilibrium with monoclinic SiP precipitates. *Phys. Rev. B* **53,** 7836–7841 (Mar. 1996).

26. Pearson, G. L. & Bardeen, J. Electrical properties of pure silicon and silicon alloys containing boron and phosphorus. *Phys. Rev.* **75,** 865–883 (1949).

27. Car, R., Kelly, P. J., Oshiyama, A. & Pantelides, S. T. Microscopic Theory of Impurity-Defect Reactions and Impurity Diffusion in Silicon. *Phys. Rev. Lett.* **54,** 360 (1985).

28. Liu, X. Y., Windl, W., Beardmore, K. M. & Masquelier, M. P. First-principles study of phosphorus diffusion in silicon: Interstitial- and vacancy-mediated diffusion mechanisms. *Appl. Phys. Lett.* **82,** 1839–1841 (2003).

29. Harrison, S. A., Edgar, T. F. & Hwang, G. S. Interstitial-mediated mechanisms of As and P diffusion in Si: Gradient-corrected density-functional calculations. *Phys. Rev. B* **74,** 195202 (Nov. 2006).

30. Ural, A., Griffin, P. B. & Plummer, J. D. Fractional contributions of microscopic diffusion mechanisms for common dopants and self-diffusion in silicon. *J. Appl. Phys.* **85,** 6440–6446 (1999).

31. Seitz, F. On the theory of diffusion in metals. *Acta Cryst.* **3,** 355–363 (1950).

32. Haddara, Y. M., Folmer, B. T., Law, M. E. & Buyuklimanli, T. Accurate measurements of the intrinsic diffusivities of boron and phosphorus in silicon. *Appl. Phys. Lett.* **77,** 1976–1978 (2000).

33. Christensen, J. S., Radamson, H. H., Kuznetsov, A. Y. & Svensson, B. G. Phosphorus and boron diffusion in silicon under equilibrium conditions. *Appl. Phys. Lett.* **82,** 2254–2256 (2003).

34. Naganawa, M. *et al.* Accurate Determination of the Intrinsic Diffusivities of Boron, Phosphorus, and Arsenic in Silicon: The Influence of SiO_2 Films. *Jpn. J. Appl. Phys.* **47,** 6205–6207 (2008).

35. Sholl, D. & Steckel, J. A. *Density functional theory : a practical introduction* (Wiley, 2009).

36. Parr, R. G. & Weitao, Y. *Density-Functional Theory of Atoms and Molecules (International Series of Monographs on Chemistry)* (Oxford, 1994).

37. Kantorovich, L. *Quantum Theory of the Solid State: An Introduction* (Kluwer, 2004).

38. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, 2004).

39. Kohanoff, J. *Electronic Structure Calculations for Solids and Molecules* (Cambridge, 2006).

40. Cramer, C. J. *Essentials of Computational Chemistry* 2nd ed. (Wiley, 2004).

41. De Broglie, L. *Recherches sur la théorie des quanta* PhD thesis. <http://www.ensmp.fr/aflb/LDB-oeuvres/De_Broglie_Kracklauer.pdf> (1924).

42. Styer, D. F. *Statistical Mechanics* http://www.oberlin.edu/physics/dstyer/StatMech/book.pdf. 2007.

43. Sauer, G. & Kuhn, W. On the stability of the hydrogen atom. *Eur. J. Phys.* **3,** 155–158 (1982).

44. Lovelace, C. Practical theory of three-particle states. I. Nonrelativistic. *Phys. Rev.* **135,** B1225–B1249 (1964).

45. Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **20,** 457 (1927).

46. Horsfield, A. P., Bowler, D. R., Fisher, A. J., Todorov, T. N. & Montgomery, M. Power dissipation in nanoscale conductors: Classical, semi-classical and quantum dynamics. *J. Phys. Condens. Matter* **16,** 3609–3622 (2004).

47. Thomas, L. The calculation of atomic fields. English. *Proc. Cambridge Philos. Soc.* **23,** 542–548 (1927).

48. Fermi, E. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. *Z. Phys.* **48,** 73–79 (1928).

49. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136,** B864 –B871 (1964).

50. Levy, M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. USA* **76,** 6062–6065 (1979).

51. Levy, M. Electron densities in search of Hamiltonians. *Phys. Rev. A* **26,** 1200–1208 (1982).

52. Sutton, A. P. *Electronic Structure of Materials* (Oxford University Press, 1993).

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

54. Dirac, P. Note on exchange phenomena in the Thomas atom. *P. Camb. Philos. Soc.* **26,** 376–385 (1930).

55. Gunnarsson, O. & Lundqvist, B. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B* **13,** 4274–4298 (1976).

56. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77,** 3865 (1996); Perdew, J. P., Burke, K. & Ernzerhof, M. Errata for Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **78,** 1396 (1997); Zhang, Y. & Yang, W. Comment on "Generalized Gradient Approximation Made Simple". English. *Phys Rev Lett* **80,** 890–890 (1998); Perdew, J., Burke, K. & Ernzerhof, M. Perdew, Burke, and Ernzerhof Reply. *Phys. Rev. Lett.* **80,** 891 (1998).

57. Perdew, J. P. & Yue, W. Accurate and Simple Density Functional for the Electronic Exchange Energy - Generalized Gradient Approximation. *Phys. Rev. B* **33,** 8800 (1986).

58. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50,** 17953 (Dec. 1994).

59. *Private communication*

60. Kresse, G. & Joubert, J. From ultrasoft pseudopotentials to the projector augmented wave method. *Phys. Rev. B* **59,** 1758 (1999).

61. Segall, M. *Slides from CASTEP workshop Durham, 6th–13th December 2001, lecture 6* <http://cmt.dur.ac.uk/sjc/Castep_Lectures2/lecture06.PDF>.

62. Baldereschi, A. Mean-Value Point in the Brillouin Zone. *Phys. Rev. B* **7,** 5212–5215 (June 1973).

63. Chadi, D. J. & Cohen, M. L. Special Points in the Brillouin Zone. *Phys. Rev. B* **8,** 5747–5753 (Dec. 1973).

64. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13,** 5188–5192 (June 1976).

65. Hellmann, H. *Einführung in die Quantenchemie* 285 (Franz Deuticke, Leipzig, 1937).

66. Feynman, R. P. Forces in Molecules. *Phys. Rev.* **56,** 340–343 (Aug. 1939).

67. Mohr, P. J., Taylor, B. N. & Newell, D. B. CODATA recommended values of the fundamental physical constants: 2006. *Rev. Mod. Phys.* **80,** 633–730 (2008).

68. Sugino, K.-i. & Okazaki, A. Low-temperature thermal expansion of silicon. *Jpn. J. Appl. Phys.* **29,** 700–705 (1990).

69. Haas, P. *et al.* Systematic investigation of a family of gradient-dependent functionals for solids. *Phys. Rev. B* **81,** 125136 (2010).

70. Castleton, C. W. M., Höglund, A. & Mirbt, S. Density functional theory calculations of defect energies using supercells. *Modelling Simul. Mater. Sci. Eng.* **17,** 084003 (2009).

71. Kresse, G., Marsman, M. & Furthmüller, J. *VASP the GUIDE* Computational Physics, Faculty of Physics, Universität Wien (2009). <http://cms.mpi.univie.ac.at/marsweb/>.

72. Tuckerman, M. E. *Statistical Mechanics : Theory and Molecular Simulations* (Oxford University Press, 2010).

73. Frenkel, D. *Understanding Molecular Simulation : from algorithms to applications* (Academic Press, 1996).

74. Ford, J. The Transition from Analytic Dynamics to Statistical Mechanics. *Adv. Chem. Phys.* **24,** 155–185 (1973).

75. Haile, J. M. *Molecular dynamics simulation : elementary methods* (Wiley, 1992).

76. Verlet, L. Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Phys. Rev.* **165,** 201–214 (Jan. 1968).

77. Swope, W. C. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **76,** 637–649 (1982).

78. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72,** 2384–2393 (1980).

79. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81,** 511 (1984).

80. Nosé, S. Constant Temperature Molecular Dynamics Methods. *Prog. Theor. Phys. Suppl.* **103,** 1–46 (1991).

81. Bylander, D. M. & Kleinman, L. Energy fluctuations induced by the Nosé thermostat. *Phys. Rev. B* **46,** 13756–13761 (Dec. 1992).

82. Hünenberger, P. H. Thermostat Algorithms for Molecular Dynamics Simulations. *Adv. Polym. Sci.* **173,** 105–149 (2005).

83. Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids* (Oxford : Clarendon Press, 1987).

84. *CRC Handbook of Chemistry and Physics* 85th ed. (ed Lide, D. R.) (CRC Press, 2005).

85. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159,** 98 (1967).

86. Yoo, S., Xantheas, S. S. & Zeng, X. C. The melting temperature of bulk silicon from ab initio molecular dynamics simulations. *Chem. Phys. Lett.* **481,** 88–90 (Oct. 2009).

87. Kirkwood, J. G. in *Theory of liquids* (ed Alder, B. J.) (Gordon and Breach, 1968).

88. Waseda, Y. *The structure of non-crystalline materials* (McGraw-Hill, 1980).

89. Waseda, Y. & Suzuki, K. Structure of molten silicon and germanium by X-ray diffraction. *Z. Phys. B* **20,** 339–343 (1975).

90. Štich, I., Car, R. & Parrinello, M. Structural, bonding, dynamical, and electronic properties of liquid silicon: An ab initio molecular-dynamics study. *Phys. Rev. B* **44,** 4262–4274 (1991).

91. Štich, I., Parrinello, M. & Holender, J. M. Dynamics, spin fluctuations, and bonding in liquid silicon. *Phys. Rev. Lett.* **76,** 2077–2080 (1996).

92. Ansell, S., Krishnan, S., Felten, J. J. & Price, D. L. Structure of supercooled liquid silicon. *J. Phys.: Condens. Matter* **10,** L73–L78 (1998).

A. Code

A.1. VMD plugins

For visualisation purposes, the two programs `VESTA` (Visualization for Electronic and STructural Analysis) and `VMD` (Visual Molecular Dynamics) have been used, both free for non-commercial use. As the name suggest, `VMD` was developed for visualising trajectories, while `VESTA` only supports still structures. But there are other major differences between the programs. While `VMD` originates from the biophysics community, `VESTA` comes from the crystallographic community. `VESTA` is therefore better suited for viewing crystals, and I found that I especially missed a visualisation of, and alignment to, crystal vectors in `VMD`. `VMD` comes with embedded Tcl, so I wrote a plugin first mimicking `VESTA`'s Orientation dialogue, and then adding display of crystal vectors, view vectors and some extras. The plugin is about 2000 codelines, includes a GUI and is quite usable.

In addition, two small plugins were developed, for introducing a cubic clipbox, which can later be rotated with the molecule, and a progressbar like the one shown in fig. 4.2 and 4.3, which can be useful when producing movies.

The plugins can be downloaded from

https://github.com/danmichaelo/crystallography
https://github.com/danmichaelo/statusindicator
https://github.com/danmichaelo/clipbox

A.2. Python package for working with trajectory data

For analysing trajectory data, a python package of about 5000 code lines was developed to parse and work with trajectory information from the `vasprun.xml` files generated by `VASP`. Trajectory data is stored internally as NumPy-arrays, making it easy to work with. Motion across periodic boundaries can be unwrapped, and trajectories can then be exported as vtf-files, which can be read by newer versions of `VMD`'s vtfplugin. This solves the slightly irritating issue with atoms jumping from one side to the other of the unit cell when viewing trajectories. Parts of the python package are far from complete however, and may not work at all, but other parts should work very well. It can be downloaded from

https://github.com/danmichaelo/oppvasp