Notes on DFT Formation Energy Corrections

Ali Binai-Motlagh

August 2019

Contents

1 Introduction

In plane Wave DFT, periodic Boundary conditions (PBC) are used to perform calculations, which simplify the description of perfect crystals. The concentration of defects in real crystals are relatively small and thus ideally one would want to study the properties of a single defect in an infinite crystal; however, this is not possible with PBC. Instead this dilute limit is approximated by using large super cells, in place of the primitive unit cell of the crystal. To obtain an accurate description of point defects within this limit, one has to utilize large super cells, sometimes on the order of several thousand atoms, pushing the limits of what is possible with DFT, within the LDA and GGA approximations. To circumvent this problem, several correction schemes have been devised that allow one to perform these calculations on significantly smaller cells, by accounting for interactions between defect images and the unphysically large defect concentration that arises due to finite super cell size and PBC.

The Central quantity in the description of point defects is the defect formation energy:

$$\Delta H_f = (E_{d,q} - E_b) + q(e_{VBM} + \Delta E_f) - \sum_i n_i \mu_i + E_{corr} \tag{1}$$

Where $E_{d,q}$ is the total energy of the defect super cell in charge state q and E_b is the total energy of the bulk crystal (the pristine host). e_{VBM} is the valence band edge of the host after potential alignment (discussed in the next section), ΔE_F is the fermi level, referenced to the host valence band edge. The fermi-level is considered to be a free parameter, to account for shifts in its position, for example due to doping. n_i is the number of the species added ($n_i > 0$) or removed ($n_i < 0$) and μ_i is the corresponding chemical potential. E_{corr} represents the correction to the formation energy due to the finite size of the super cell.

2 Correction Methods

2.1 Correction of spurious Coulomb interaction

The main source of error in Formation energies obtained through the super cell method is due to the spurious electrostatic interaction between periodic images of charged defect. Note that the energy of a charged periodic system is infinite and so in performing calculations with charged cells, the $\mathbf{G} = 0$ term in the DFT total energy is removed [1]. This has the same effect as adding a uniform charge of density q/Ω, to neutralize the super cell [1]. The interaction of the defect charges with this uniform background adds further error to the formation energy.

This spurious electrostatic interaction was first addressed by Leslie and Gillian by subtracting the screened Madelung energy of an array of point charges immersed in a neutralizing background [2], originally derived

by Fuchs [3], given by the Ewald sum[1]:

$$\begin{aligned} E_{img} = E_{PC}^{iso} &= \frac{-\alpha Q^2}{2\epsilon L} \\ &= \frac{1}{\epsilon}\Bigg[\frac{2\pi}{\Omega}\sum_{\mathbf{G}\neq\mathbf{0}}\sum_i \frac{1}{G^2} q_i e^{-i\mathbf{G}\cdot\mathbf{r_i}} e^{\frac{-G^2}{4\gamma^2}} + \frac{1}{2}\sum_{i,j,k}\frac{q_i q_j}{|\mathbf{r_i}+\mathbf{R_k}-\mathbf{r_j}|}\mathrm{erfc}(\gamma|\mathbf{r_i}+\mathbf{R_k}-\mathbf{r_j}|) \\ &\quad - \frac{\gamma}{\sqrt{\pi}}\sum_i q_i^2 - \frac{\pi Q^2}{2\Omega\gamma^2}\Bigg] \end{aligned} \tag{2}$$

Where $Q = \sum_i q_i$ and γ is the ewald parameter controlling the convergence rate of the real and reciprocal space sums. $\mathbf{R_k}$ is a direct lattice vector and $\mathbf{r_i}$ and $\mathbf{r_j}$ denote the position of the ions in the $\mathbf{R_k} = \mathbf{0}$ unit cell. α is the madelung constant, dependant only on the geometry of the unit cell. Terms in the real space sum for which i = j when $\mathbf{R_k} = 0$ should be omitted.

For the case of a single ion in the unit cell this simplifies to [4]:

$$E_{PC}^{iso} = \frac{-\alpha q^2}{2\epsilon L} = \frac{-q^2}{2\epsilon}\Bigg[\sum_{\mathbf{R_i}}^{i\neq 0}\frac{\mathrm{erfc}(\gamma|\mathbf{R_i}|)}{|\mathbf{R_i}|} + \sum_{\mathbf{G_i}}^{i\neq 0}\frac{4\pi}{\Omega}\frac{\exp -\mathrm{G_i^2}/4\gamma^2}{G_i^2} - \frac{2\gamma}{\sqrt{\pi}} - \frac{\pi}{\Omega\gamma^2}\Bigg] \tag{3}$$

In general, an isotropic dielectric response is only valid for the case of cubic cells. Rurali and Cartoixà extended this point charge correction to the case of anisotropic material where the dielectric constant is replaced by the tensor $\bar{\epsilon}$, [5]:

$$\begin{aligned} E_{PC}^{aniso} =& \frac{-q^2}{2}\Bigg[\sum_{\mathbf{R_i}}^{i\neq 0}\frac{1}{\sqrt{\det\bar{\epsilon}}}\frac{\mathrm{erfc}(\gamma\sqrt{\mathbf{R_i}^T\bar{\epsilon}^{-1}\mathbf{R_i}})}{\mathbf{R_i}^T\bar{\epsilon}^{-1}\mathbf{R_i}} + \sum_{\mathbf{G_i}}^{i\neq 0}\frac{4\pi}{\Omega}\frac{\exp(-\mathbf{G_i}^T\bar{\epsilon}\mathbf{G_i}/(4\gamma^2))}{\mathbf{G_i}^T\bar{\epsilon}\mathbf{G_i}} \\ &- \frac{\pi}{\Omega\gamma^2} - \frac{2\gamma}{\sqrt{\pi det\bar{\epsilon}}}\Bigg] \end{aligned} \tag{4}$$

By considering the case of molecules in vacuum and decomposing the charge density as a sum of a point charge term and an extended charge density ρ_e, Makov and Payne [6], obtained a third order term to the image charge correction, which accounts for the interaction of the extended charge with the potential of the point charges:

$$E_{img} = \frac{q^2\alpha}{2L} - \frac{2\pi qQ}{3L^3} \tag{5}$$

Where Q is the second radial moment of the extended charge density:

$$Q = \int r^2\rho_c(\mathbf{r})d\mathbf{r} \tag{6}$$

For the case of defects in crystalline systems, the extended density is replaced by the defect induced density, $\Delta\rho_D = \rho_D - \rho_H$, [1]. Where ρ_D and ρ_H are the charge density of the defect and host systems

[1]For a discussion of Ewald sums see [12]

respectively. Lany and Zunger found that after potential alignment (which scales as L^{-3}), the image charge correction scales as L^{-1} rather than $aL^{-1} + bL^{-3}$ as expected by the Makov-Payne correction [1]. They also demonstrated that the third order correction scales linearly as a function of the third order, i.e. it also has a L^{-1} dependence. They explain this by demonstrating that the extended part of the defect density is caused in large by the dielectric screening response of the host crystal [1]. When a defect with charge q is added, it is screened to $\frac{q}{\epsilon}$ and the screening charge $-q(1-\frac{1}{\epsilon})$ is drawn approximately uniformly from the entire super cell. This means that beyond a certain distance from the defect, the delocalized part of the defect induced potential is approximately the constant: $\Delta\rho_D \approx \frac{q}{\Omega}(1-\frac{1}{\epsilon})$. This implies that $Q \sim L^2$, thus explaining the L^{-1} scaling of the third order correction.
Using this approximation, the image correction can be written as a multiple of the first order correction [7]:

$$E_{img} = [1 + C_{sh}(1 - \frac{1}{\epsilon})]\frac{\alpha \cdot q^2}{2\epsilon L} \tag{7}$$

C_{sh} is the shape factor, a quantity that only depends on the geometry of the cell and can be obtained using the approximate $\Delta\rho_D$ and integrating Q over a Wigner Seitz cell.

Later Freysoldt, Neugebauer, and Van de Walle proposed another scheme for correcting formation energies obtained through the DFT super cell method [8].

First let $V_{d,q}$, $V_{d,0}$ denote the electrostatic potential of the system with a charged defect and the neutral defect respectively. The change in the potential due to the addition of a charge q is given by: $V_{q/0} = V_{d,q} - V_{d,0}$ (Note that equivalently one can use the potential of the pristine cell V_b as reference i.e., consider $V_{q/b} = V_{d,q} - V_b$). And the change in potential due to a periodic set of charges is given by:

$$\tilde{V}_{q/0}(\mathbf{r}) = \frac{1}{\Omega}\sum_{\mathbf{G}\neq\mathbf{0}} V^{rec}_{q/0}(\mathbf{G})e^{\mathbf{G}\cdot\mathbf{r}} \tag{8}$$

Where $V^{rec}_{q/0}$ is the Fourier transform of $V_{q/0}$. So that the unwanted potential due to the artificial periodicity is: $\tilde{V}_{q/0} - V_{q/0}$. From this we can calculate the spurious interaction energy between the periodic images:

$$E^{inter} = \frac{1}{2}\int_\Omega (\tilde{V}(\mathbf{r})_{q/0} - V_{q/0}(\mathbf{r}))(q_d(\mathbf{r}) + n)d^3r \tag{9}$$

The total correction must also include the interaction between the background and the defect in the $\mathbf{R} = \mathbf{0}$ cell:

$$E^{intra} = \int_\Omega nV_{q/0}d\mathbf{r} \tag{10}$$

Rewrite $V_{q/0} = V^{lr}_{q/0} + V^{sr}_{q/0}$ as the sum of a short range, accounting for the microscopic screening, and a long range potential, accounting for the macroscopic screening. The long range potential is given by the screened Coulomb potential:

$$V^{lr}_{q/0}(\mathbf{r}) = \frac{1}{\epsilon}\int \frac{q_d(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}' \tag{11}$$

The total correction to the formation energy is:

$$E_{img} = E_{inter} + E_{intra} = E_q^{lat} - q\Delta_{q/0} \tag{12}$$

Where E_q^{lat} is the macroscopically screened energy of the defect charges and the neutralizing background:

$$E_{lat}^q = \int_\Omega (\frac{1}{2}(q_d + n)(\tilde{V}_q^{lr} - V_q^{lr}) + nV_q^{lr})d\mathbf{r} \tag{13}$$

Since the average electrostatic potential in DFT is set to 0 by removing the $\mathbf{G} = \mathbf{0}$ component of the sum, i.e. $\int_\Omega \tilde{V}_q^{lr} d\mathbf{r} = 0$, this simplifies to:

$$E_{lat}^q = \underbrace{\frac{1}{2}\int_\Omega q_d \tilde{V}_q^{lr} d\mathbf{r}}_{\text{Periodic Energy}} - \underbrace{\frac{1}{2}\int_\Omega q_d V_q^{lr} d\mathbf{r}}_{\text{isolated energy}} + \underbrace{\frac{n}{2}\int_\Omega V_q^{lr} d\mathbf{r}}_{\text{Background energy}} \tag{14}$$

The periodic energy can be calculated using Poisson's equation in reciprocal space, $V(\mathbf{G}) = \frac{4\pi}{\epsilon}\frac{q_d(\mathbf{G})}{G^2}$ and writing $\tilde{V}_q^{lr}$ analogously to equation 8:

$$E_{per} = \frac{2\pi}{\epsilon\Omega}\sum_{\mathbf{G}} \frac{|q_d(\mathbf{G})|^2}{G^2} \tag{15}$$

The isolated energy can be derived similarly (assuming a spherically symmetric charge distribution):

$$E_{iso} = \frac{1}{\pi\epsilon}\int_0^\infty dg|q_d(g)|^2 \tag{16}$$

Using a point charge distribution for q_d one obtains the first order madlung energy term of the previous schemes. $\Delta_{q/0}$ is a potential alignment like term that can be shown to be equivalent to the third order term in the Makov Payne correction when a point charge distribution is used for q_d, [9]:

$$\Delta_{q/0} = \frac{1}{\Omega}\int_\Omega V_{q/0}^{sr}(\mathbf{r})d\mathbf{r} \tag{17}$$

$V_{q/0}^{sr}$ can be calculated using:

$$V_{q/0}^{sr} = \tilde{V}_{q/0} - \tilde{V}_q^{lr} - C \tag{18}$$

$\tilde{V}_{q/0}$ can be obtained through DFT calculations and $\tilde{V}_q^{lr}$ using Ewald summation.

The undetermined shift, C, can be obtained by looking at a planar averaged plot of $\tilde{V}_{q/0} - \tilde{V}_q^{lr}$. Assuming that $V_{q/0}^{sr}$ goes to 0 far from the defect, $\tilde{V}_{q/0} - \tilde{V}_q^{lr}$ should plateau in this region, this is the value of C. The plateau also serves as a check for the assumptions of the method i.e. if it is observed, then we have successfully separated the potential into short and long range components [8].

Picking a correction scheme: In 2012 Kosma *et al.*[9] conducted a comparison of the different correction schemes. They computed formation energies for successively larger cells and fitted the scaling law

$aL^{-1} + bL^{-3} + c$ to the data. The infinite cell size limit, $L = \infty$, of this scaling law was used as the benchmark for comparing the different correction schemes. They found the FNV correction outperformed all other schemes.

Dielectric constant: if the system containing the defect is allowed to relax, one should use the static dielectric constant, ϵ_0, otherwise the high frequency limit of the dielectric constant, ϵ_∞ should be used.

2.2 Potential Alignment

In plane wave DFT, the divergent average electrostatic potential is set to 0 by convention [7]. This is done by removing the $\mathbf{G} = \mathbf{0}$ term in the Fourier expansion of the potential [7]. This has the same effect as adding a uniform background neutralizing charge to the cell. The removal of this term means that the potential, and therefore the KS eigenvalues are determined up to an undetermined constant [7]. While the total energy of a charge neutral system can be compared between DFT calculations, the total energy of charged systems suffers from the same arbitrary shift [7]. This convention therefore results in an error in the formation energy of charged defects.
This arbitrary shift is corrected for by matching up the undetermined constant for the total energy of the defect, $E_{d,q}$with that of the valence band maximum of the host, e_{VBM}. To account for this shift, take the difference between the potential of the pristine bulk and defect systems:

$$\Delta V(\mathbf{r}) = V_{d,q}(\mathbf{r}) - V_b(\mathbf{r}) \tag{19}$$

And average it over a region far enough from the defect such that the potential of the defect cell replicates that of the bulk (modulo the undetermined shift):

$$\Delta V_A = \Delta V(\mathbf{r})|_{far} \tag{20}$$

This average accounts for the undetermined shift. The corresponding energy, $q\Delta V_A$, added to the formation energy, is called the "potential alignment" correction [7].

The averaging can be done in many different ways. The simplest would be to perform a planer average of the data in the xy plane and pick the value along z that is furthest from the defect [10]. This can however lead to problems:
The electrostatic potential is highly oscillatory due to the presence of deep wells at the location of each atom. Far enough away from the defect, where the position of atoms in the defect and pristine cell correspond, this strong oscillation is canceled out when one takes the difference between the potential of the defect and pristine cells. In some cases however, the atoms shift considerably after geometry optimization, even in very large super cells, and this cancellation does not occur, leading to a highly oscillatory potential [10]. This problem may be circumvented by averaging out the oscillations. Another way, as proposed by lany and Zunger in their correction method, would be to use atomic sphere averaged potentials and average the difference between the host and defect cell at sites far from the defect [7].

Whatever the chosen scheme maybe, if the averaging is not performed far enough away from the de-

fect, then it is possible that some of the shift captured by ΔV_A is caused by the electrostatic interaction between defects, not just the error caused by the 0 potential convention [10]. This can lead to double counting if the part of the electrostatic interaction present in ΔV_A was already accounted for in E_{corr}. Note that the FNV scheme does not suffer from this double counting problem because the electrostatic potential of the model charge is subtracted during alignment.

Some authors have shown that both potential alignment and electrostatic interaction corrections are necessary to obtain rapid convergence of ΔH_f with super cell size [1]. Others argue that if the electrostatic interaction correction is performed accurately, then there may be no need for potential alignment [4].

Typically, the potential used for this alignment procedure is the total electrostatic potential due to the electrons and ions, given by:

$$V_H(\mathbf{r}) + V_{loc}(\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \sum_i V^i_{loc}(|\mathbf{r} - \mathbf{r}'|) \tag{21}$$

Where V_H is the hartree potential due to the electron density $n(\mathbf{r})$ and V_{loc} is the potential due to the nuclei and core electrons (i.e. the local component of the pseudo-potential).

2.3 Band filling Corrections

Another source of error in ΔH_f due finite size is Moss-Burstein-type band-filling, which arises as a result of unphysically large defect concentration. For shallow donors, the correction is given by [11]:

$$\Delta E_{bf} = -\sum_{n\mathbf{k}} \eta_{n,\mathbf{k}} (e_{n,\mathbf{k}} - \tilde{e}_{CBM}) \quad \text{for } e_{n,\mathbf{k}} > \tilde{e}_{CBM} \tag{22}$$

Where $\tilde{e}_{CBM}$ is the host conduction band minimum after potential alignment. $\eta_{n,\mathbf{k}}$, $e_{n,\mathbf{k}}$ are the band occupations and energies respectively.

For shallow acceptors, the correction is given by [11]:

$$\Delta E_{bf} = \sum_{n,\mathbf{k}} w_{\mathbf{k}} (1 - \eta_{n,\mathbf{k}}) (e_{n,\mathbf{k}} - \tilde{e}_{VBM}) \quad \text{for } e_{n,\mathbf{k}} < \tilde{e}_{VBM} \tag{23}$$

Considering all sources of error, the total correction to the formation energy is:

$$E_{corr} = E_{img} + q\Delta V_A + \Delta E_{bf} \tag{24}$$

2.4 The Band Gap Problem

Beyond errors introduced due to super cell finite size effects, another important source of uncertainty inherent to DFT is the Band gap problem. These can be addressed through post LDA/GGA methods such as DFT+U and GW [1]. A simpler approach would be to use the experimental band gap of the material. There are several possibilities in this case [13]:

- Rigidly shift the conduction band to match the experimental band gap of the material, leaving defect states unchanged.
- Shift shallow donors along with the conduction band.
- Compute the formation energy as a function of a parameter λ (e.g. the energy cutoff) that affects the band gap, then the correction to the formation energy is given by extrapolating to the experimental value:

$$H_f(\lambda) = H_f(\lambda_0) + \frac{\partial H_f}{\partial E_g}|_{\lambda_0}\Big(E_g(\lambda) - E_g(\lambda_0)\Big) = H_f(\lambda_0) + \delta E$$

  Where $E_g(\lambda) = E_g^{expr}, E_g(\lambda_0) = E_g^{DFT}$, $H_f(\lambda)$ is the band gap corrected formation energy and $H_f(\lambda_0)$ is the initial formation energy.

Position of the band edges especially in narrow band gap material can have a profound effect on the predicted doping behaviour [14]. For example, considering intrinsic defect formation energies in PbTe using different levels of theory, A. Goyal *et al.* found that the only level of theory that reproduces experimental findings, required a combination of hybrid functionals with spin orbit coupling, along with G_0W_0 to shift the band edges. They found that the main difference between the different levels of theory was in the position of the VBM and CBM relative to each other and Vacuum [14].

3 Documentation

The corrections.py package contains several methods for manipulating crystal structures and computing the formation energy corrections.
To start, one has to first create a crystal structure. This is done by creating a cell object. The constructor can be called in 2 ways; By providing a file specifying the structure and atoms or by directly passing a list of atoms and lattice vectors:

```
import numpy as np
from corrections import *

Si = cell('Si_struct.txt')

atom1 = atom('Si', [0, 0, 0])
atom2 = atom('Si', [0.25, 0.25, 0.25])
atoms = [atom1, atom2]

lattice = np.array( [[0.5, 0, 0.5], [0, 0.5, 0.5], [0.5, 0.5, 0]] )

Si = cell(atoms = atoms, lattice = lattice)
```

```
Si.alat = 5.4
print(Si)
```

The cell object expects the lattice as a 3×3 matrix with the primitive lattice vectors as its columns. Atoms are stored using the atom object constructor. This takes the name of the atom as the first parameter and its position in crystal coordinates as the second. When created from a text file, the cell object expects the following format:

```
5.4
0.5  0  0.5
0  0.5  0.5
0.5  0.5  0
Si  0  0  0
Si  0.25  0.25  0.25
```

The first line specifies the lattice parameter and can be omitted, the next 3 lines specify the lattice vectors **as rows**. The lines following that, specify the atoms in the cell and their coordinates. When a cell object is printed, the output will have the above format as well.
Now that we have imported the crystal structure, we can manipulate the structure and compute the corrections:

```
#creates a 2x2x2 super cell
Si.super(2, 2, 2)

thr = 100 #energy cutoff (eV) for reciprocal space sum
gamma = 5 #ewald parameter
eps = np.identity(3) * 12.1 #dielectric tensor (isotropic)
charge = 2 #charge of the defect

#computes the madelung energy of an array of point charges
#in a neutralizing background as given by equation 3
E1 = Si.firstO(thr, gamma, charge,eps=np.identity(3))

#Computes Lany and Zunger's image charge correction
E2 = Si.LZ_img(thr, gamma, charge, eps)
```

Note that to ensure that the sum has converged, it may be useful to tune the ewald parameter as well as the reciprocal space cutoff of the firstO function and plot the result:

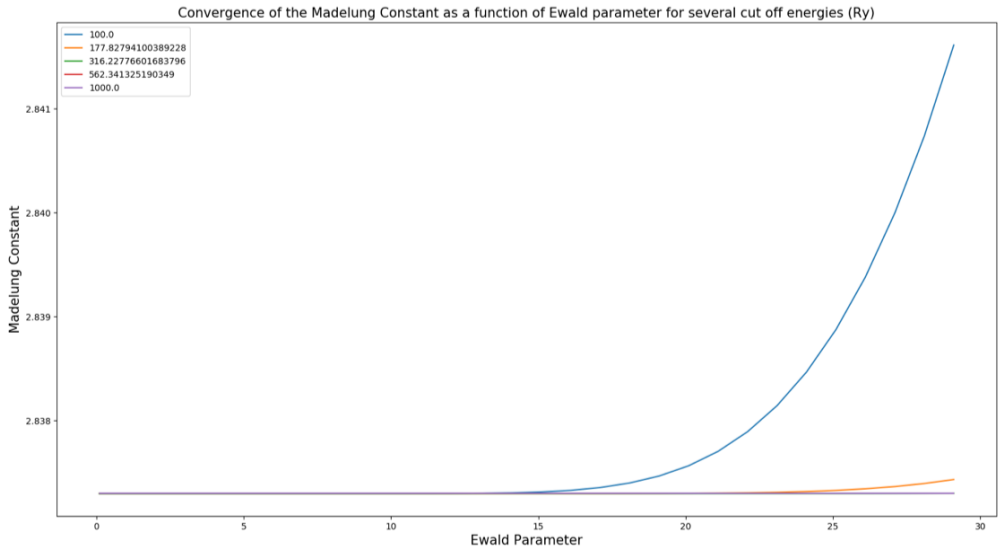

Figure 1: Madelung Constant of an FCC lattice as a function of the ewald parameter, plotted for several energies (eV).

As it can be seen from the graph, the madelung constant converges to a value of 2.8373 and the sum converges more rapidally for smaller values of the ewald parameter.

The package also contains methods for the other corrections:

```
#Computes band filling correction, given the VBM and CBM
#of the host and a file containing occupations, weights
#and energies from the DFT calculation
bandfill('file.txt', VBM, CBM)

#Reads the potential, obtained from a DFT calculation
#into a 3d array. Expects the file in the .cube format
readPotential('potFile.cube')

radii = [3] * 216 #assuiming the cell contains 216 atoms
n = 10 #grid density for integration

#Performs a sphere average of the potential about
#the position of each atom
Si.atomSphereAverage('potFile.cube' raddi, n)
```

'file.txt' for the bandfilling correction can be obtained using the **getOCC.sh** scraper by specifying the name of the scf output file as its first parameter.
The potential in the cube file format can be obtained using the post processing code **pp.x**. *radii* is a list containing the radius of the sphere about each atom used for the spherical averaging. The averaging is

performed by linearly interpolating the potential and evaluating the integral:

$$\frac{\int_0^{2\pi}\int_0^{\pi}\int_0^{R} \mathrm{r}^2 \sin(\theta) V(r,\theta,\phi) dr d\theta d\phi}{4\pi \mathrm{r}^3/3} \tag{25}$$

about each atom.

The above package implements the Lany and Zunger corrections. To use the FNV correction run the **sxdefectalign** executable [13].

When performing structural relaxation on large super cells it is best to use Davidson's diagonalization algorithm as it is faster than conjugate gradients. However, in my experience, it is much more unstable and prone to crashes. The **getIonRelax.sh** script can be used to extract final ionic coordinates from a relaxation calculation or to quickly restart a failed one using the latest coordinates before the crash.

4 Notes on DFT calculations with Quantum Espresso

4.1 DFT: A short summary

The many body Schrodinger equation is given by:

$$\left[-\sum_l \frac{\hbar^2}{2M_l} \nabla^2_{\mathbf{R_l}} - \sum_k \frac{\hbar^2}{2m_k} \nabla^2_{\mathbf{r_k}} + V(r,R) \right] \Psi(r,R,t) = i\hbar \frac{\partial \Psi(r,R,t)}{\partial t}$$

$$V(r,R) = \underbrace{-\sum_{k,l} \frac{Z_l e^2}{|\mathbf{r_k} - \mathbf{R_l}|}}_{e^- - \text{ion}} + \underbrace{\frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r_i} - \mathbf{r_j}|}}_{e^- - e^-} + \underbrace{\frac{1}{2} \sum_{l \neq n} \frac{Z_l Z_n e^2}{|\mathbf{R_l} - \mathbf{R_n}|}}_{\text{ion-ion}}$$

where $r \equiv (\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n}) \equiv (e^- s\text{positions}), R \equiv (\mathbf{R_1}, \mathbf{R_2}, ..., \mathbf{R_N}) \equiv (\text{ion positions})$

Storage requirements makes numerical solutions impossible for most problems of interest (8 e^- on a $10 \times 10 \times 10$ grid requires storage of 1000^8 complex number or $\sim 10^{17}$TeraBytes). Therefore approximations are needed to make the problem tractable.
Born-Oppenheimer approximation (M >> m), yields the following equation for the electronic wave function:

$$\left(-\sum_k \frac{\hbar^2}{2m_k} \nabla^2_{\mathbf{r_k}} + V(r,R) \right) \psi_R(r) = E(R) \psi_R(r) \tag{26}$$

Calculated for a given static combination of ion positions. Global minimum of the potential energy surface E(R), determines the ground state configuration.
Density functional theory maps this 3(n + N) dimensional problem to a 3 dimensional one through the Honenbrg-Kohn theorem [15]:

Theorem 1 *The ground state Energy is a unique functional of the electron density* $n(\mathbf{r})$ *which itself is uniquely determined by V(r, R).*

The density of N $e^{-}s$ is given by:

$$n(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r_2}, ..., \mathbf{r_N})|^2 d\mathbf{r_2}...d\mathbf{r_N} \tag{27}$$

Kohn and Sham suggested to write the electron density in terms a set of non-interacting $e^{-}s$ with wave functions ϕ_i whose density is equal to that of the actual system [16]:

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2, \quad \text{where } \langle \phi_i | \phi_j \rangle = \delta ij \tag{28}$$

The energy functional has the general form:

$$E[n(\mathbf{r})] = T[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] + \int n((r)) V_{ext}(\mathbf{r}) d\mathbf{r} \tag{29}$$

$$= \frac{-\hbar^2}{2m} \sum_i \int \langle \phi_i | \nabla^2 | \phi_i \rangle \, d\mathbf{r} + \frac{e^2}{2} \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + \int n((r)) V_{ext}(\mathbf{r}) d\mathbf{r} \tag{30}$$

Where $V_{ext} = \sum_i \frac{Z_i e^2}{|\mathbf{r} - \mathbf{R_i}|}$ is the potential due to ions, acting on the e^{-}. E_H is the Hartree energy, associated with $e^{-} - e^{-}$ interaction (mean field approximation). E_{xc} is the exchange-correlation energy which accounts for the non-classical interactions. Its exact form is unknown and must be approximated. The most popular approximations, striking a balance between accuracy and time are the Local Density approximation (LDA) (E_{xc} depends only on $n(\mathbf{r})$ (local) and is approximated using exchange-correlation energy of a uniform $e^{-}gas$) and the generalized gradient approximation (GGA) (E_{xc} also depends on the gradient of the density $|\nabla n(\mathbf{r})|$). While these approximations yield accurate total energies and structures, they suffer from several problems:

- Self interaction from the Hartree potential is not fully canceled by V_{XC} since exchange is approximated in DFT (unlike HF). This leads to problems when dealing with systems that are strongly correlated (e.g. magnetic systems).
- Underestimation of band gap by roughly a factor of 2, attributed to the lack of discontinuity in E_{XC} at integer e^{-} numbers (due to non-cancellation of self-interaction, fractional e^{-} occupations are favoured).

Minimizing the energy w.r.t to $\{\phi_i(\mathbf{r})\}$, i.e. $\frac{\delta E[n((r))]}{\delta \psi(\mathbf{r})} = 0$, yields the Kohn-Sham (KS) equation:

$$\underbrace{\left(\frac{-\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) \right)}_{\hat{H}_{KS}} \phi_i(\mathbf{r}) = E_i \phi_i(\mathbf{r}) \tag{31}$$

The KS equation is typically solved self-consistently:

1. Make an initial guess for the density n($\mathbf{r}$), any positive function normalized to the total number of e^- works but an intelligent choice yields faster convergence: e.g. sum of atomic densities: $\sum_i n_i(\mathbf{r} - \mathbf{R_i})$, $\mathbf{R_i}$ = position of nuclei.

2. Evaluate V_{KS} with the guess. Note that it may be faster to obtain the Hartree potential by solving Poisson's equation instead: $\nabla^2 V_H = -4\pi n(\mathbf{r})$.

3. Write the ϕ_i in terms of some basis set: $\phi(\mathbf{r}) = \sum_i c_i f_i(\mathbf{r})$. In which case the Kohn-Sham equation becomes: $\sum_j (H_{ij} - E_{ks} S_{ij}) c_j = 0$ where $H_{ij} = \langle \phi_i | \hat{H}_{KS} | \phi_j \rangle$ are the matrix elements of $\hat{H}_{KS}$ and $S_{ij} = \langle \phi_i | \phi_j \rangle$ is the positive definite overlap matrix.
   A popular basis set, used by Quantum Espresso, ideal for periodic systems are plane waves: $f_i(\mathbf{r}) = e^{\mathbf{G_i} \cdot \mathbf{r}}$, which are orthogonal $S_{ij} = \delta_{ij}$ and can take advantage of FFTs.
   Then solve for the n lowest eigenstates of $\hat{H}_{KS}$ through diagonalization, typically solved iteratively so that the entire Hamiltonian does not have to be stored, e.g. using conjugate gradients. This is the most time consuming part of the calculation.

4. Compute the new density using the eigenstates

5. Stop if $|E_i - E_{i-1}| < \epsilon_E$ or $\int |n_i - n_{i-1}| d\mathbf{r} < \epsilon_n$ (some user defined threshold). Otherwise make a new guess for the density and go to 2. To avoid numerical instability, rather than using the newly computed density as the guess, a linear combination of previously computed densities is used: e.g. $n_{new,in} = \beta n_{out} + (1 - \beta) n_{in}$. This algorithm is guaranteed to converge for small enough β while when simply using n_{new}, convergence is rarely achieved.

Pseudo Potentials: The wave functions are highly oscillatory and contain many nodes near the core, this means that a large number of plane waves are required to accurately represent ψ. This is circumvented by only considering valence $e^- s$ (core e^- typically do not participate in bonding). That is, by replacing the Coulomb potential, V_{ext} with an effective ("pseudo") potential that approximates the potential felt by the valence e^- due to the ion and core $e^- s$.

The following sections will outline how to use the Plane Wave DFT code, Quantum Espresso for different types of calculations, using Silicon as an example.

4.2 SCF Calculation

In Quantum Espresso the program `pw.x` is used to self-consistently solve the KS equations. Quantum espresso input files are divided into 2 components: **NAMELISTS** and **INPUT_CARDS**. There are 3 required **NAMELISTS**:

- **&CONTROL**: Used to specify the type of calculation and I/O related variables.
- **&SYSTEM**: Used to specify the system, e.g. the number of atoms (nat), number of types of atoms (ntyp), the bravais lattice (ibrav) and the plane wave energy cutoff (ecutwfc).
- **&ELECTRONS**: Used to specify the details of the algorithm used to solve the KS equation self consistently. e.g. the diagonalization technique (diagonalization), energy convergence threshold

for self-consistency (conv_thr) and how to generate the density for an iteration from previous ones (mixing_mode, mixing_beta).

These should be terminated with a '/'.

There are 3 required **INPUT_CARDS:**

- **ATOMIC_SPECIES**: Specify the name, mass and the pseudo potential file (located in pseudo_dir) for each specie involved in the calculation
- **ATOMIC_POSITIONS**: Coordinates of each atom in the unit cell. This can for example be in Cartesian or crystal coordinates. The example input lists all the possible options.
- **K_POINTS**: Used to specify the coordinate and weight of each k point used in Brillouin Zone sums. These can be manually specified in Cartesian or crystal coordinates or automatically: by specifying the number of grid points along each direction, the code will automatically generate a Monkhorst-Pack grid.

To slove the KS equation execute: `pw.x < scf.in > scf.out`
scf.in:

```
&CONTROL
   calculation = 'scf'
   prefix      = 'Si'
   outdir      = './'
   pseudo_dir  = './'
/
&SYSTEM
   ! ibrav:  0 = free lattice,   1 = PC,   2 = FCC,   3 = BCC, ...
   ibrav      = 0
   celldm(1)  = 1
   nat        = 1, ntyp = 1
   ecutwfc    = 30.0
   nbnd       = 12
/
&ELECTRONS
   conv_thr        = 1d-7
   mixing_beta     = 0.7
   diagonalization = 'david'
/
CELL_PARAMETERS { alat }
   0.5   0.0   0.5
   0.0   0.5   0.05
   0.5   0.5   0.0
ATOMIC_SPECIES
```

```
! atomLabel   atomMass   atomPseudoPotential
   Si          28.086     Si.pz-vbc.UPF

ATOMIC_POSITIONS { alat | bohr | angstrom | crystal | crystal_sg }
 ! atomLabel   x   y   z
    Si        0.0  0.0  0.0
    Si        0.25 0.25 0.25

K_POINTS { tpiba | automatic | crystal | gamma | tpiba_b | crystal_b | tpiba_c | crystal_c }
```

$\underbrace{5\ \ 5\ \ 5}_{\text{Grid Points}}\quad \underbrace{0\ \ 0\ \ 0}_{\text{Grid Offset}}$! If automatic is used

The quantity of interest (e.g. total energy) must be converged w.r.t ecutwfc and k-grid density, exact value is dependent on the pseudo-potential (PP) used. Note that the absolute value of energy is not physically meaning full (dependent on PP), differences are important, and these typically converge more quickly.
For spin-polarized, co linear calculations add `nspin=2, starting_magnetization(i)=0` and for non-co linear add `noncolin=.true., lspinorb=.true.` (Need fully relativistic PP) to `SYSTEM`.
For metallic systems the Brillouin Zone integrals are over functions that are discontinuous at the fermi-level and very dense k-grids are needed to adequately approximate them. This problem is circumvented by replacing the step function by a smooth function, i.e. smear the T = 0K Fermi-Dirac distribution. To do so add the lines:

```
occupations = 'smearing'
smearing    = 'Gaussian' ! Step function replaced by 1/2(1 - erf(x / a))
degauss     = 0.01 ! smearing parameter(a), larger = smoother
```

to the `SYSTEM` name list. Other options for smear include using the finite temperature fermi-Dirac or Methfessel Paxton (expand step function in terms of a complete orthogonal basis):

$$f_0(x) = \frac{1}{2}(1 - erf(x)) \tag{32}$$

$$f_N(x) = \sum_{n=1}^{N} \frac{-1^n}{n!4^n\sqrt{\pi}} H_{2n-1}(x)e^{-x^2} \quad H_n = n^{th} \text{ Hermite polynomial} \tag{33}$$

The value of the smearing parameter must be tested by "performing several runs at different σ for increasingly dense k-point grids, until a suitable k-point grid and σ are found yielding satisfactorily converged results."
Another option is to use the linear tetrahedron method (Break up the BZ into tetrahedra and linearly interpolate the integrand within each), specified through `occupations = 'tetrahedra'`. This yields the best k-grid convergence for energy but it is not variational. So use this for Energy/DOS calculations and regular smearing for optimization.

4.3 Structural Relaxation

To perform ionic relaxation, change calculation to 'relax' and add the following card:

```
&ions
    ion_dynamics = 'bfgs'
/
```

3 binary numbers beside the atomic coordinate can be used to constrain motion during optimization. For example:

```
    ATOMIC_POSITIONS { alat }
    Si        0.0  0.0  0.0 1 1 0 ! motion constrained to xy plane
    Si        0.25 0.25 0.25 0 0 0 ! Atom is fixed
```

To optimize the cell parameters, in addition to atomic positions, change calculation to 'vc-relax' and add the card to the scf text:

```
&CELL
    cell_dynamics = 'bfgs'
    cell_dofree   = 'Volume' ! Used to constrain lattice vectors
                             ! In this case Volume changes but angles remain fixed
/
```

4.4 Band Structure

Band structure calculation is divided into 3 parts:

- pw.x < scf.in > scf.out
- pw.x < Sbands.in > Sbands.out
- bands.x < bands.in > bands.out

First perform a self consistent calculation then perform a non-self consistent calculation by changing calculation to 'bands' and changing K_POINTS to:

```
K_POINTS crystal_b
5
-0.25  0.25  -0.5  50 !W
 0.0   0.5    0.0  50 !L
 0.0   0.0    0.0  50 !Gamma
-0.5   0.0   -0.5  50 !X
-0.25  0.25  -0.5     !W
```

This specifies a path in reciprocal space along which to compute the eigenvalues. To visualize the structure and 1st BZ of the reciprocal lattice we can use the program **Xcrysden**:
`xcrysden --pwi scf.in` or `xcrysden --pwo scf.out`

The program can also be used to select a path in reciprocal space. Also checkout `https://www.materialscloud.org/work/tools/seekpath`.
bands.x is used to format the nscf output in a manner that can be easily plotted.
bands.in:

```
&BANDS
    prefix  = 'Si'
    outdir = './'
    filband = 'Si_bands'
/
```

This writes 2 files: `Si_bands.gnu`, which can be plot with: `gnuplot < Si_bands.gnu`, and `Si_bands` which can be formatted using the `plotband.x` program.
Position of high symmetry points along the specified path is contained in `bands.out`.

4.5 DOS

First perform a scf calculation, then perform a non-self consistent calculation, increasing the k-grid density and setting `calculation = 'nscf'`:

- pw.x < scf.in > scf.out
- pw.x < nscf.in > nscf.out
- dos.x < dos.in > dos.out

Finally run dos.x with the input:
dos.in:

```
&DOS
prefix = 'Si'
outdir = './'
fildos = 'Si.dos' ! contains the D(E) data
emin   = -20, emax = 20, DeltaE = 0.02 ! energy range for D(E)
/
```

To obtain projected dos use: `profwfc.x < pdos.in > pdos.out`
pdos.in:

```
&PROJWFC
prefix = 'Si'
outdir = './'
emin   = -10, emax = 20, DeltaE = 0.02
ngauss = 0, degauss = 0.01
filpdos = 'Si' ! prefix of output files containing pDos(E)
/
```

4.6 Phonon Calculations

To obtain the phonon dispersion/DOS, follow the following steps:

- pw.x < scf.in > scf.out ! Get KS wave functions (phonon calculations require an accurate description of the ground state so decrease **conv_thr**)
- ph.x < ph.in > ph.out ! Calculate inter-atomic force constants on a uniform **q** grid using DFPT.
- q2r.x < q2r.in > q2r.out ! Post processing tool; Fourier transforms results to real space
- matdyn.x < matdyn.in > matdyn.out ! Interpolates dynamical matrix from q2r.x at arbitrary **q**

ph.in

```
&INPUTPH
prefix = 'Si'
outdir = './'
tr2_ph = 1d-14 ! self consistency threshold
ldisp  = .true. ! calculate on a uniform grid.
nq1 = 4, nq2 = 4, nq3 = 4 ! grid points
amass(1) = 28.0855 ! atomic mass
fildyn = 'Si.dyn' ! file containing the dynamical matrix
/
```

q2r.in

```
 &INPUT
 fildyn = 'Si.dyn'
 flfrc  = 'Si.fc' ! real space force constants
 zasr = 'simple' ! Acoustic sum rule
 /
```

matdyn.in (DOS)

```
&INPUT
asr   = 'simple' ! Acoustic sum rule
flfrc = 'Si.dyn'
dos   = .true.
fldos = 'Si.phdos' ! file containing phonon dos
nk1 = 40, nk2 = 40, nk3 = 40
/
```

matdyn.in (Dispersion)

```
 &INPUT
 asr   = 'simple'
 flfrc = 'Si.dyn'
 flfrq = 'Si.frq'   ! w(q) file
 q_in_band_form   = .true.
 q_in_cryst_coord = .true.
 /
 5
 -0.25  0.25  -0.5  50 !W
  0.0   0.5    0.0  50 !L
  0.0   0.0    0.0  50 !Gamma
 -0.5   0.0   -0.5  50 !X
 -0.25  0.25  -0.5     !W
```

q_in_band_form allows you to specify high symmetry points as in a band structure calculation (specify end points and the number of points in between rather than explicitly specifying each **q** point between the high symmetry points).

4.7 Dielectric tensor

First perform a scf calculation then perform a $\mathbf{q} = \mathbf{0}$ phonon calculation:

- pw.x < scf.in > scf.out
- ph.x < ph_epsil.in > ph_epsil.out

ph_epsil.in:

```
&inputph
outdir='./'
prefix='Si'
tr2_ph=1d-14
epsil=.true.
ldisp=.false.
amass(1)=28.086
fildyn='Si.dyn'
/
0 0 0
```

4.8 Extracting results from SCF calculation:

pp.x is a post processing tool that uses the output of pw.x to calculate a specified property and output it in a format suitable for plotting: `pp.x < pp.in > pp.out`

pp.in:

```
&INPUTPP ! calculates desired property
outdir   = './'
prefix   = 'Si'
plot_num = 0 !  specifies what to plot: 0 = charge density, 1 = total potential
             !  2 = local ionic potential V_bare, 11 = the V_bare + V_H potential
filplot = 'Si_rho.out' ! file that contains desired quantity
/
&PLOT !formats output for plotting
filepp(1) = 'Si_rho.out' ! file containing the quantity to be formatted
iflag = 3 ! 3D plot
output_format = 6 ! cube file format
fileout = 'Si_rho.cube'
nx = 128, ny = 128, nz = 128
/
```

The '.cube' output can the be visualized for example using Vesta. For potential Alignment, use `plot_num = 11`.

References

[1] S. Lany, A. Zunger, Phys. Rev. B 78 (2008) 235104.

[2] M. Leslie, M.J. Gillan, J. Phys. C: Solid State Phys. 18 (1985) 973.

[3] K. Fuchs, Proc. R. Soc. A 151, 585 (1935).

[4] Y. Kumagai and F. Oba, Phys. Rev. B 89, 195205 (2014).

[5] R. Rurali and X. Cartoixa‘, Nano Lett. 9, 975 (2009).

[6] G. Makov, M.C. Payne, Phys. Rev. B 51 (1995) 4014.

[7] S. Lany and A. Zunger, Modell. Simul. Mater. Sci. Eng. 17, 084002 (2009).

[8] C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev.Lett. 102, 016402 (2009).

[9] H.P. Komsa, T. T. Rantala, and A. Pasquarello, Phys. Rev. B 86, 045112 (2012).

[10] T. R. Durrant, S. T. Murphy, and A. L. Shluger, J. Chem. Phys. 149, 024103 (2018).

[11] A. Goyal, P. Gorai, H. Peng, S. Lany and V. Stevanović, Comput. Mater. Sci. 130 (2017).

[12] J.M. Ziman, Principles of the theory of solids, 2nd ed. (Cambridge University Press, Cambridge, 1972), 37-41

[13]C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, Rev. Mod. Phys. 86 (2014)

[14] A. Goyal, P. Gorai, E.S. Toberer, and V. Stevanović, NPJ COMPUT. MATER. 42 (2017)

[15] P. Honenberg, and W. Kohn, Phys. Rev. 136 (1964)

[16] W. Kohn, and L. J. Sham, Phus. Rev. 140 (1965)