Pawpyseed: A parallelized Python/C package to aid in density functional theory point defect calculations via utilities for band shifting corrections

Kyle Bystrom

Dept. of Materials Science and Engineering, University of California, Berkeley

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honors in Chemistry, under the supervision of Professor Mark Asta

Fall 2018

Abstract

Significant progress has been made recently in the automation and standardization of *ab initio* point defect calculations in the form of improved formation energy corrections for charged defects and workflow software to aid in calculation setup and post-processing. However, the task of developing, implementing, and benchmarking charge corrections for density functional theory (DFT) point defect calculations is still an open challenge. To contribute to this goal, a parallelized Python and C package called pawpyseed is developed to perform numerical analysis of DFT wavefunctions in the projector augmented wave (PAW) formalism. The utilities contained in the code can be used to perform perturbative band shifting corrections for point defect calculations. The theory and implementation of pawpyseed for this application is discussed, and other potential applications of the code are mentioned briefly in the discussion. In addition, various correction methods, including a perturbative band shifting method implemented using pawpyseed, are used to calculate the formation energies and transition levels of several point defects in silicon (phosphorous, boron, sulfur, and copper substitutionals and the single vacancy). The transition level predictions are compared to each other as well as previous experimental and theoretical data. A discussion of the correction methods is presented in the context of the studied defects, and hypotheses are presented for errors for different correction methods. Possible future developments of corrections for high-throughput point defect calculation workflows are discussed.

Contents

1 Introduction **2**
- 1.1 The Supercell Method for Point Defects . . . 2
- 1.2 The Perturbative Band Shifting Correction . . . 4
- 1.3 The Projector Augmented Wave (PAW) Method . . . 5
- 1.4 Implementing overlap operators in pawpyseed . . . 6

2 Theory **6**
- 2.1 Overlap Operators in PAW for Wavefunctions from Different Structures . . . 8
- 2.2 Mapping Pseudo Wavefunctions Between Symmetrically Identical K-points . . . 9
- 2.3 The Band-Shifting Correction for Point Defect Calculations . . . 9

3 Implementation **10**
- 3.1 Overlap of Pseudo Wavefunctions (O_0) . . . 10
- 3.2 Concentric Augmentation Spheres (O_M) . . . 11
- 3.3 Partial Waves Overlapping with Pseudo Wavefunction (O_R, O_S) . . . 11
- 3.4 Partial Wave Overlap on Non-Orthogonal Augmentation Spheres (O_N) . . . 12

4 Computational Details **12**
- 4.1 DFT Calculations . . . 12
- 4.2 Correction Methods . . . 13

5 Results **13**
- 5.1 Defect Transition Level Predictions 13
- 5.2 Understanding Energy Corrections Using Level Projections . . . 16
- 5.3 Γ-centered K-point Mesh DFT Calculations . . . 17
- 5.4 Quantitative Comparison of 100% Shift and Projection Shift . . . 17

6 Discussion **20**
- 6.1 Effectiveness of Band Filling and Band Shifting Corrections . . . 20
- 6.2 Other Applications of Pawpyseed . 21

7 Summary and Conclusions **21**

8 Acknowledgments **22**

1 Introduction

Point defects play a defining role in materials science, particularly electronics [1]. For example, a thorough understanding of phosphorous and boron doping in silicon were essential to the discovery of the p-n junction, a fundamental component of modern electronics. More recently, rapid progress has been made in halide perovskite solar cell efficiency partly because the defects in halide perovskites are unusually benign toward electronic properties [2].

Experimental studies of point defects often provide limited information about defect type and composition [3], so it is informative to determine the properties of defects using electronic structure simulations such as Kohn-Sham density functional theory (DFT) [4]. In addition, achieving high accuracy of ground state energies is critical because the concentration of a defect in a crystal, a primary property of interest, scales exponentially with the defect's formation energy. Developing high-throughput models is particularly useful for point defects because setting up and parsing point defect calculations is time-consuming and complicated to do manually. Automation can make these methods accessible to industrial scientists and therefore accelerate their research.

One of the key hurdles to high-throughput calculation of defect properties is the "band-gap problem," in which local and semi-local DFT significantly underestimate the band gaps of solids [3]. Since discrete defect levels (the single-particle states introduced by a defect in a solid) are located in the band gap, this band-gap problem results in inaccurate defect properties in simulations. One proposed solution to this problem is to use perturbation theory to shift the band edges from a DFT calculation to improve the accuracy of the predicted defect properties.

In this work, a DFT post-processing software package called pawpyseed is presented with the goal of making the implementation of perturbative band shifting corrections easier. The rest of this section gives an overview of current work in DFT point defect calculations and correction methods (1.1 and 1.2) and then provides a brief background on the challenges addressed by pawpyseed (1.3 and 1.4). The tools in pawpyseed are centered around the evaluation of overlap operator expectation values between wavefunctions from defect structures and wavefunctions from bulk structures. Section 2 describes the theoretical formalism for these overlap operators and presents a simple perturbative band shifting correction. Section 3 gives an implementation and runtime scaling analysis. Section 4 provides computational details for the DFT study of silicon point defects in this work. Section 5 presents predictions of transition levels and formation energies for these defects using several correction methods, including a newly implemented band shifting correction. Section 6 discusses the performance of different correction methods, as well as possible future applications of the pawpyseed code. Section 7 contains concluding remarks.

1.1 The Supercell Method for Point Defects

Most first principles studies of point defects use the supercell method: The point defect is embedded in a supercell of the bulk structure, which is repeated infinitely throughout space and modeled with DFT. This method allows point defects to be modeled in robust, high-performance plane-wave DFT codes, but it also raises several problems. First, the periodic boundary conditions cause unphysical elastic and electronic interactions of the defect with periodic images of itself (finite-size effects). Second, typical DFT functionals drastically underestimate material band gaps due to inaccurate correlation effects, resulting in unrealistic delocalization of electron density [1, 3]. The former problem is generally addressed with finite-size corrections that subtract unphysical interactions out of the total energy, but these corrections break down when electron charge is unrealistically delocalized [5]. The delocalization issue is generally circumvented using hybrid functional methods, which correct the band gap but are too computationally expensive for high-throughput applications.

In addition, the behavior of defect levels in hybrid is nontrivially influenced by mixing parameters which are difficult to tune and not always physically justified [6, 7]. Therefore, high-throughput defect calculation workflows use DFT and then attempt to correct the resulting physical inaccuracies using post-processing tools.

In the supercell method formalism, the formation energy of a defect X in a charge state q is given by [8]:

$$E^f[X^q] = E_{tot}[X^q] - E_{tot}[bulk] - \sum_i n_i \mu_i + qE_F + E_{corr} \tag{1}$$

where E_{tot} is the total energy of a DFT calculation, n_i is the change in number of each specie (chemical element) from the bulk structure to the defect structure, μ_i is the chemical potential of each specie, E_F is the Fermi level, and E_{corr} contains correction terms.

The correction for electrostatic interaction of the defect with itself is generally performed using the Freysoldt [9] method or Kumagai [10] method. The Freysoldt method is used in this study. The Freysoldt method defines

$$E_{corr} = -E_q^{lat} + q\Delta_{q/b} \tag{2}$$

E_q^{lat} is the electrostatic interaction of the defect charge with periodic images of itself, and $\Delta_{q/b}$ is a correction to the potential that removes the compensating background charge introduced in non-charge-neutral DFT calculations. (Because electrostatic energy of a periodic system diverges if each unit cell has a net charge, periodic DFT codes add a compensating background charge to make the system neutral if the number of electrons does not equal the number of protons in the system). See Freysoldt and Van de Walle [1] for details.

Another common correction is the band filling correction [3]. Because defects are localized features in a crystal, bands introduced by a defect should be dispersionless (i.e. have the same energy at each k-point). However, the periodicity of the system in the supercell method can cause dispersion in defect bands, which means shallow defects can introduce holes below the valence band maximum (VBM) and electrons above the conduction band minimum (CBM) in DFT. This cannot occur in a real system because a hole below the VBM will rise to the VBM, and an electron above the CBM will relax to the CBM. Band filling corrections, such as the one in PyCDT [8] and pymatgen [11], shift energy levels associated with defects to remove dispersion of defect levels above the CBM and below the VBM. For example, if a conduction band state of energy $E_{d,\mathbf{k}}$ is occupied at k-point $\mathbf{k}$, then the term $\omega_{\mathbf{k}}(E_{CBM} - E_{d,\mathbf{k}})$ is added to E_{corr} in Equation 1, where $\omega_{\mathbf{k}}$ is the k-point weight.

The final correction is the band shifting correction, which attempts to correct for the inaccuracy of the band gap for the generalized gradient approximation (GGA) and other semi-local functionals [3]. The general approach is to calculate accurate band edges of the bulk crystal using a higher level of theory, such as hybrid DFT, and then correct the defect energy based on the band edge shifts:

$$\Delta E_{CBM} = E_{CBM,hybrid} - E_{CBM,GGA} \tag{3}$$

$$\Delta E_{VBM} = E_{VBM,hybrid} - E_{VBM,GGA} \tag{4}$$

Hybrid indicates that the band edges were calculated using a hybrid DFT level of theory, which generally gives better band gaps than semi-local DFT. GGA indicates the band edges were calculated with the GGA functional.

The most basic band shifting correction, which will be called the "100% shift" method, has three terms to be added to E_{corr}. The first is $q\Delta E_{VBM}$, which serves as an effective shift to the Fermi level in Equation 1 since the Fermi level is referenced to the VBM [3]. The second is ΔE_{CBM} times the number of electrons in the conduction band, as these electrons are assumed to shift in energy

100% with the conduction band. Third, $-\Delta E_{VBM}$ times the number of holes in the valence bands accounts for shifting hole energy. One can imagine the third term as subtracting the energy term that would arise if electrons occupied the unfilled valence states.

The 100% shift correction is relatively rudimentary because it is possible that not all levels shift entirely with the valence or conduction band. For example, a defect level described completely by an atomic state on a defect atom should not shift with the valence or conduction bands between the GGA and hybrid levels of theory. However, due to the underestimation of the band gap in GGA, such a level may be located above the CBM or below the VBM in GGA and therefore be incorrectly shifted by the 100% shift method. In addition, states in the band gap do not get shifted, even though a shallow defect level near a band edge might shift with the band edge. A more sophisticated band shifting method based on perturbation theory has been proposed [3] and is discussed presently.

1.2 The Perturbative Band Shifting Correction

Throughout this section, Kohn-Sham single-particle states, which are the eigenfunctions of the Kohn-Sham DFT Hamiltonian, will be referred to as wavefunctions for brevity. However, it is important to note that these states are not unique and only physically significant because they generally give good approximations to single-particle ionization energies, which allows them to be treated as orbitals occupied by a single electron (or electron pair) in an effective potential. Because the eigenfunctions of a Hamiltonian form a complete basis, a defect wavefunction can be expanded in the bulk wavefunctions [3]:

$$\psi_D(\mathbf{r}) = \sum_{n,\mathbf{k}} \langle \psi_{n,\mathbf{k}} | \psi_D \rangle \, \psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{n,\mathbf{k}} A_{n,\mathbf{k}} \psi_{n,\mathbf{k}}(\mathbf{r}) \tag{5}$$

ψ_D is a defect wavefunction, and $\psi_{n,\mathbf{k}}$ are evaluated in the pristine bulk. Physically, ψ_D is a single state, as opposed to a band of states in k-space, because the defect is not periodic and therefore does not have dispersion. However, it is evaluated at different k-points in the supercell method, so the supercell method defect levels $\psi_{D,\mathbf{k}}$ will be used to ground the band shifting correction in a practical computational framework. Since all wavefunctions at different k-points are orthogonal,

$$\psi_{D,\mathbf{k}}(\mathbf{r}) = \sum_{n} A_{n,\mathbf{k}} \psi_{n,\mathbf{k}}(\mathbf{r}) \tag{6}$$

If the single-particle energy level of $\psi_{D,\mathbf{k}}(\mathbf{r})$ in DFT is $e^0_{D,\mathbf{k}}$, first-order perturbation theory can be used as suggested by Lany and Zunger to calculate a corrected energy [3].

$$e_{D,\mathbf{k}} = e^0_{D,\mathbf{k}} + \langle \psi_{D,\mathbf{k}}(\mathbf{r}) | \, \Delta H \, | \psi_{D,\mathbf{k}}(\mathbf{r}) \rangle \tag{7}$$

Here, ΔH is a correction term that extrapolates from the DFT picture to the true quasiparticle energies. Assuming the diagonal elements of ΔH in the basis of the bulk wavefunctions are small, Equation 7 can be expanded:

$$e_{D,\mathbf{k}} = e^0_{D,\mathbf{k}} + \sum_{n} |A_{n,\mathbf{k}}|^2 \langle \psi_{n,\mathbf{k}}(\mathbf{r}) | \, \Delta H \, | \psi_{n,\mathbf{k}}(\mathbf{r}) \rangle \tag{8}$$

A rough but potentially useful approximation to ΔH is a "band shifting" operator that shifts the energy of a bulk conduction band by ΔE_{CBM} (Equation 3) and the energy of a bulk valence band by ΔE_{VBM} (Equation 4). This band shifting operator therefore shifts the energy levels of a defect system calculated in GGA toward the more accurate hybrid energy levels based on how they project onto host bands.

It is useful to define the "proportion valence" $v_{D,\mathbf{k}}$ and "proportion conduction" $c_{D,\mathbf{k}}$ as:

$$v_{D,\mathbf{k}} \equiv \sum_{n \in VB} |\langle \psi_{D,\mathbf{k}} | \psi_{n,\mathbf{k}} \rangle|^2 \tag{9}$$

$$c_{D,\mathbf{k}} \equiv \sum_{n \in CB} |\langle \psi_{D,\mathbf{k}} | \psi_{n,\mathbf{k}} \rangle|^2 \tag{10}$$

Here, VB is the set of valence bands and CB is the set of conduction bands. Using these definitions, Equation 8 can be expressed simply:

$$e_{D,\mathbf{k}} = e^0_{D,\mathbf{k}} + c_{D,\mathbf{k}} \Delta E_{CBM} + v_{D,\mathbf{k}} \Delta E_{VBM} \tag{11}$$

This perturbative method assumes that the difference in exchange-correlation (XC) energy of an electron in a single-particle state between hybrid and GGA is similar in both the bulk and defect structures. A second assumption made for the simplified perturbative method above is that this change in XC energy is equal for all valence bands and all conduction bands.

This correction is of interest because it has been proposed in multiple cases [1, 3, 12] but does not have a standard formalism or open-source implementation. This is partly because the accurate evaluation of the overlap terms $\langle \psi_{D,\mathbf{k}} | \psi_{n,\mathbf{k}} \rangle$ is computationally difficult in plane-wave DFT with the projector-augmented wave (PAW) method, which is the most commonly used basis set for modern periodic solid calculations. Understanding why this is requires a basic introduction to PAW wavefunctions, which is outlined in the next subsection.

1.3 The Projector Augmented Wave (PAW) Method

It is ideal to use plane waves of the form $e^{i\mathbf{k}\cdot\mathbf{r}}$ as a basis set for the wavefunctions of periodic crystals because they can be manipulated quickly using fast Fourier transforms (FFTs), leading to fast algorithms for Hamiltonian diagonalization and potential calculation. However, the atomic valence states of atoms have frequency components of tens of thousands of electron volts (eV) near the nucleus because the electrostatic potential is large and discontinuous. Describing these valence states with plane waves would lead to a prohibitively large basis set, so a transformation T is introduced which maps a pseudo (PS) wavefunction, which is a smooth sum of plane waves, to an all electron (AE) wavefunction, which describes the true eigenfunction of the KS Hamiltonian [13].

$$T = 1 + \sum_a \sum_l \sum_m \sum_\epsilon (|\phi_{alm\epsilon}\rangle - |\widetilde{\phi}_{alm\epsilon}\rangle) \langle \widetilde{p}_{alm\epsilon}| = 1 + \sum_i (|\phi_i\rangle - |\widetilde{\phi}_i\rangle) \langle \widetilde{p}_i| \tag{12}$$

$\phi_{alm\epsilon}$ are all electron (AE) partial waves, $\widetilde{\phi}_{alm\epsilon}$ are pseudo (PS) partial waves, $\widetilde{p}_{alm\epsilon}$ are projector functions, a are the site indices of each atom in the structure, and l, m, and ϵ specify a spherical harmonic and energy quantum number which uniquely specify a partial wave at a given atomic site. Projector functions are localized within a cutoff radius r_c around the nucleus of an atom, and $\widetilde{\phi}_{alm\epsilon} = \phi_{alm\epsilon}$ outside an augmentation radius r_a. AE partial waves form a basis for the atomic valence states inside r_c. PS partial waves form a basis for the PS wavefunction inside r_c. Projector functions determine how the PS wavefunction maps to the AE wavefunction. A summation over i (or j, as below) represents a summation over a, l, m, and ϵ. For further details on the PAW method, including the physical significance and construction of the partial waves and projector functions, see Blochl's original paper [13] and Kresse and Joubert's paper relating ultrasoft pseudopotentials and PAW [14]. For the purpose of this work, the primary concern is the form of PAW wavefunctions and how operators are evaluated in this formalism, rather than the exact method by which the PAW datasets are constructed and how the Hamiltonian is diagonalized.

When T is applied to a PS wavefunction $|\widetilde{\psi}_{n\mathbf{k}}\rangle$, the AE wavefunction $|\psi_{n\mathbf{k}}\rangle$ is recovered.

$$|\psi_{n\mathbf{k}}\rangle = |\widetilde{\psi}_{n\mathbf{k}}\rangle + \sum_{a,l,m,\epsilon} \left(|\phi_{alm\epsilon}\rangle - |\widetilde{\phi}_{alm\epsilon}\rangle\right) \langle \widetilde{p}_{alm\epsilon}|\widetilde{\psi}_{n\mathbf{k}}\rangle \tag{13}$$

To evaluate operators, one defines a pseudo operator $\widetilde{A}$ for each operator A such that $\langle\psi| A |\psi\rangle = \langle\widetilde{\psi}| \widetilde{A} |\widetilde{\psi}\rangle$. Because $|\psi\rangle = T|\widetilde{\psi}\rangle$, one can write [13]:

$$\widetilde{A} = T^{\dagger} A T \tag{14}$$

One can then plug Equation 12 into Equation 14 to find

$$\widetilde{A} = [1 + \sum_i |\widetilde{p}_i\rangle \left(\langle\phi_i| - \langle\widetilde{\phi}_i|\right)] A [1 + \sum_j (|\phi_j\rangle - |\widetilde{\phi}_j\rangle) \langle\widetilde{p}_j|] \tag{15}$$

$$\begin{aligned}\widetilde{A} =& A + \sum_i |\widetilde{p}_i\rangle \left(\langle\phi_i| - \langle\widetilde{\phi}_i|\right) A + \sum_j A(|\phi_j\rangle - |\widetilde{\phi}_j\rangle) \langle\widetilde{p}_j| \\ &+ \sum_i \sum_j |\widetilde{p}_i\rangle \left(\langle\phi_i| - \langle\widetilde{\phi}_i|\right) A (|\phi_j\rangle - |\widetilde{\phi}_j\rangle) \langle\widetilde{p}_j|\end{aligned} \tag{16}$$

When the operator A is local, then $\sum_j |\widetilde{\phi}_j\rangle \langle\widetilde{p}_j| = 1$, which reduces Equation 16 to a simpler form for local operators [13]:

$$\widetilde{A} = A + \sum_i \sum_j |\widetilde{p}_i\rangle \left(\langle\phi_i| A |\phi_j\rangle - \langle\widetilde{\phi}_i| A |\widetilde{\phi}_j\rangle\right) \langle\widetilde{p}_j| \tag{17}$$

1.4 Implementing overlap operators in pawpyseed

At first glance at Equation 17, it seems that the overlap operator should be evaluated as any local operator is evaluated in PAW. However, the basis of a PAW wavefunction is structure dependent because the partial waves and projector function are dependent on the atomic composition and positions. For the unusual application described here, where the wavefunctions for which the overlap operator is desired belong to two different structures, it is necessary to evaluate Equation 16 in full, where i and j sum over the sites of the two different structures. Figure 1 illustrates how the augmentation regions interact for the evaluation of overlap operators of wavefunctions from different structures.

Another difficulty that arises from evaluating overlap operators is that DFT codes generally reduce the number of k-points sampled using symmetry operations. In general, defect structures will have lower symmetry than bulk structures, so the wavefunctions for the two structures will be evaluated at different sets of k-points. It is therefore necessary to extrapolate wavefunctions at one k-point from wavefunctions at a symmetrically identical k-point in order to perform all the desired overlap operator evaluations.

The theory section addresses these two problems so that pawpyseed can implement the evaluation of the overlap operators required for the perturbative band shifting correction.

2 Theory

This section develops the overlap operator formalism and mapping between symmetrically identical k-points used in pawpyseed. In addition, a simple perturbative band-shifting correction is presented.

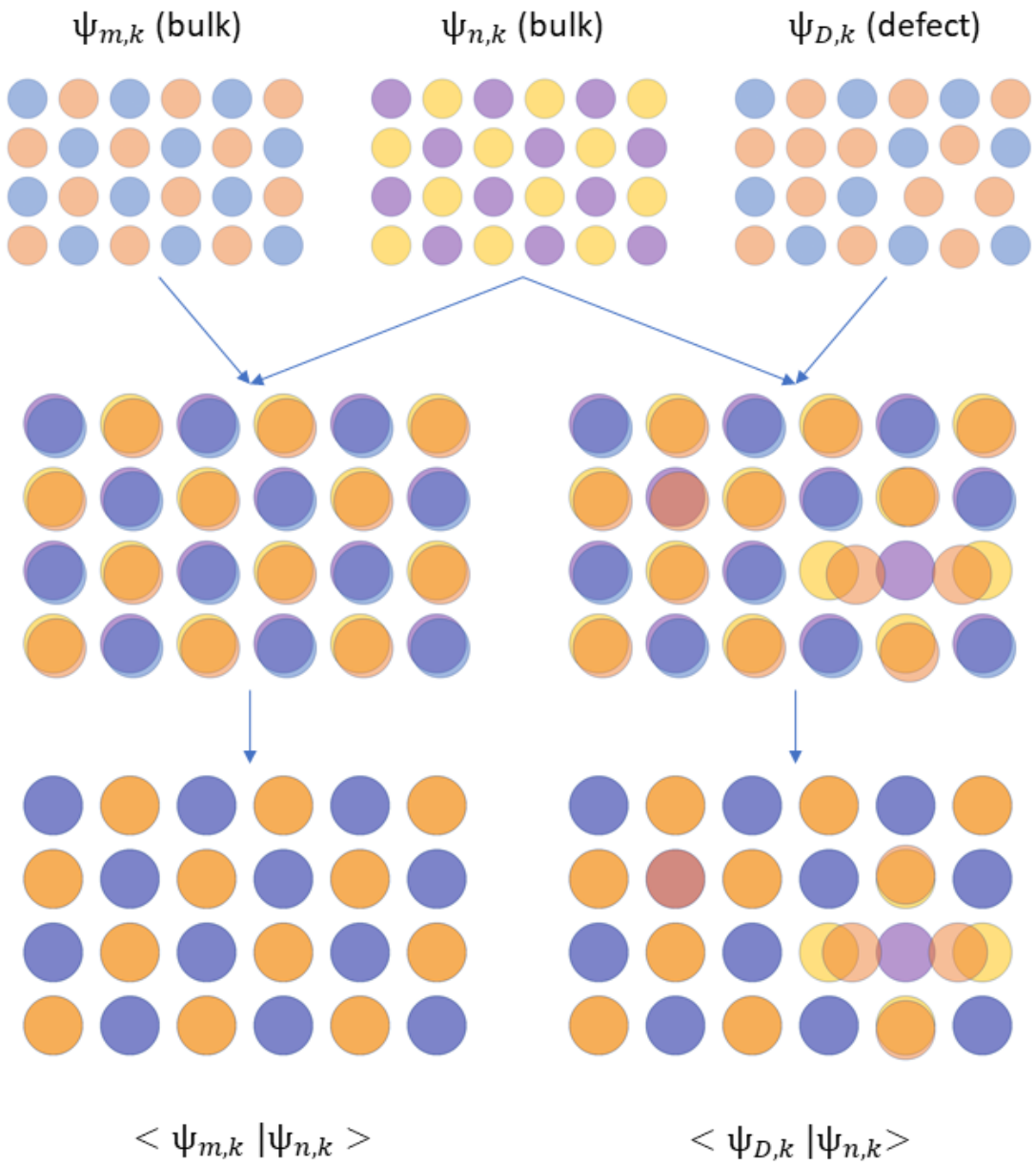


Figure 1: A schematic diagram of the augmentation regions for a crystal in 2D. The colored regions represent the augmentation regions of a wavefunction, which are defined by the crystal structure. Orange and yellow are one element, and purple and blue are a different element. On the left, two bulk states are projected onto each other, and the augmentation regions overlap exactly. This means that integrals of the partial waves can be calculated on simple radial grids. On the right, a defect state is projected onto a bulk state. As shown, augmentation regions now overlap nonconcentrically, which prevents easy radial integration, and augmentation regions overlap with the interstitial region, which requires that the high-frequency AE partial waves be projected onto the smooth PS wavefunction.

2.1 Overlap Operators in PAW for Wavefunctions from Different Structures

The following section derives an equation for the overlap operator between one Kohn-Sham single particle state of one structure R and one Kohn-Sham single particle state of another structure S, where R and S share a common lattice and the DFT PS wavefunctions are constructed with the same plane-wave basis set in the PAW formalism. The goal of this derivation is to present this overlap operator in a form convenient for computation.

Starting with Equation 16 for any single-particle operator in the PAW formalism and replacing A with unity gives the overlap operator (note that the summation over i is for structure R and the summation over j is for structure S):

$$\begin{aligned}
\langle\psi_{Rn_1\mathbf{k}}|\psi_{Sn_2\mathbf{k}}\rangle &= O_0 + O_1 + O_2 + O_3 \\
O_0 &= \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle \\
O_1 &= \sum_i \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\widetilde{p}_i\rangle \left(\langle\phi_i| - \langle\widetilde{\phi}_i|\right)|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle \\
O_2 &= \sum_j \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\left(|\phi_j\rangle - |\widetilde{\phi}_j\rangle\right)\langle\widetilde{p}_j|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle \\
O_3 &= \sum_i \sum_j \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\widetilde{p}_i\rangle \left(\langle\phi_i| - \langle\widetilde{\phi}_i|\right)\left(|\phi_j\rangle - |\widetilde{\phi}_j\rangle\right)\langle\widetilde{p}_j|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle
\end{aligned} \tag{18}$$

It is important to simplify the calculation of the other terms in equation 18 as much as possible because the calculation can be computationally expensive, and the number of necessary calculations for projecting onto an entire basis set can scale with the number of sites times the size of the basis set. One major simplification is that if a site a in structure R and site b in structure S have the same species and position, a and b will only have overlapping augmentation regions with each other and no other sites. Then, defining O_{1a} as the summation over on-site terms for the identical sites a and b in O_1 (and using like definitions for O_{2a} and O_{3a}):

$$O_{1a} + O_{2a} + O_{3a} = \sum_{l,m} \sum_{\epsilon_1} \sum_{\epsilon_2} \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\widetilde{p}_{alm\epsilon_1}\rangle \left(\langle\phi_{alm\epsilon_1}|\phi_{alm\epsilon_2}\rangle - \langle\widetilde{\phi}_{alm\epsilon_1}|\widetilde{\phi}_{alm\epsilon_2}\rangle\right)\langle\widetilde{p}_{alm\epsilon_2}|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle$$

which is the local operator solution derived by Blochl. All three terms must be evaluated in full for the other sites, but terms in O_3 where i and j correspond to sites with non-overlapping augmentation spheres vanish. Therefore, if M_{RS} is the set of identical sites in the structures R and S, N_R and N_S are the sets of sites in R and S not in M_{RS}, and N_{RS} is the set of *pairs* of sites not in M_{RS} with overlapping augmentation regions, then

$$\langle\psi_{Rn_1\mathbf{k}}|\psi_{Sn_2\mathbf{k}}\rangle = O_0 + O_M + O_R + O_S + O_N \tag{19}$$

$$O_M = \sum_{i,j\in M_{RS}} \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\widetilde{p}_i\rangle \left(\langle\phi_i|\phi_j\rangle - \langle\widetilde{\phi}_i|\widetilde{\phi}_j\rangle\right)\langle\widetilde{p}_j|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle \tag{20}$$

$$O_R = \sum_{i\in N_R} \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\widetilde{p}_i\rangle \left(\langle\phi_i| - \langle\widetilde{\phi}_i|\right)|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle \tag{21}$$

$$O_S = \sum_{j\in N_S} \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\left(|\phi_j\rangle - |\widetilde{\phi}_j\rangle\right)\langle\widetilde{p}_j|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle \tag{22}$$

$$O_N = \sum_{i,j\in N_{RS}} \langle\widetilde{\psi}_{Rn_1\mathbf{k}}|\widetilde{p}_i\rangle \left(\langle\phi_i| - \langle\widetilde{\phi}_i|\right)\left(|\phi_j\rangle - |\widetilde{\phi}_j\rangle\right)\langle\widetilde{p}_j|\widetilde{\psi}_{Sn_2\mathbf{k}}\rangle \tag{23}$$

2.2 Mapping Pseudo Wavefunctions Between Symmetrically Identical K-points

Since changing the basis of a lattice (atoms and atomic positions) can change the space group, and because DFT calculations reduce the k-point sampling space based on symmetry operations, it is important for a code which calculates overlap operators of wavefunctions from different structures to be able to derive a wavefunction at one k-point from a wavefunction at a symmetrically identical k-point. Two k-points $\mathbf{k}$ and $\mathbf{k}'$ are symmetrically identical if $\mathbf{k}' = \Theta\mathbf{k}$, where Θ is the non-translation component of a space group operation $R = T\Theta$ of the crystal, where T is the translation. For nonmagnetic systems, they are also symmetrically identical if related by time inversion ($\tau : t \to -t$).

Because R commutes with H, the above condition guarantees that

$$H\psi_{n\mathbf{k}} = E_{n\mathbf{k}}\psi_{n\mathbf{k}} \tag{24}$$

$$HR\psi_{n\mathbf{k}} = E_{n\mathbf{k}}R\psi_{n\mathbf{k}} \tag{25}$$

Since the k-point of $R\psi_{n\mathbf{k}}$ is $\mathbf{k}'$, the eigenfunctions at $\mathbf{k}'$ can be specified as:

$$\psi_{n\mathbf{k}'} = R\psi_{n\mathbf{k}} \tag{26}$$

Next, the wavefunctions are expressed as a sum of plane waves:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sqrt{\frac{1}{V}}e^{i\mathbf{k}\cdot\mathbf{r}}\sum_{\mathbf{G}} C_{n,\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot\mathbf{r}} \tag{27}$$

Plugging Equation 27 into Equation 26 and taking $T = \Delta r$ to be the translational component of R, a condition on the plane-wave constants can be derived as shown:

$$\begin{aligned}
e^{i\mathbf{k}\cdot\mathbf{r}}\sum_{\mathbf{G}} C_{n,\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot\mathbf{r}} &= e^{i\boldsymbol{\Theta}\mathbf{k}\cdot(\Theta\mathbf{r}+\Delta\mathbf{r})}\sum_{\mathbf{G}} C_{n,\Theta\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot(\Theta\mathbf{r}+\Delta\mathbf{r})} \\
\sum_{\mathbf{G}} C_{n,\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot\mathbf{r}} &= e^{i\Theta\mathbf{k}\cdot\Delta\mathbf{r}}\sum_{\mathbf{G}} C_{n,\Theta\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot(\Theta\mathbf{r}+\Delta\mathbf{r})} \\
\sum_{\mathbf{G}} C_{n,\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot\mathbf{r}} &= e^{i\Theta\mathbf{k}\cdot\Delta\mathbf{r}}\sum_{\mathbf{G}} C_{n,\Theta\mathbf{k},\Theta\mathbf{G}}e^{i\Theta\mathbf{G}\cdot(\Theta\mathbf{r}+\Delta\mathbf{r})} \\
\sum_{\mathbf{G}} C_{n,\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot\mathbf{r}} &= \sum_{\mathbf{G}} C_{n,\Theta\mathbf{k},\Theta\mathbf{G}}e^{i\mathbf{G}\cdot\mathbf{r}}e^{i\Theta(\mathbf{k}+\mathbf{G})\cdot\Delta\mathbf{r}} \\
C_{n,\mathbf{k},\mathbf{G}} &= e^{i\Theta(\mathbf{k}+\mathbf{G})\cdot\Delta\mathbf{r}}C_{n,\Theta\mathbf{k},\Theta\mathbf{G}}
\end{aligned} \tag{28}$$

This simple relationship allows a pseudo wavefunction at $\mathbf{k}$ to be quickly mapped to a pseudo wavefunction at $\Theta\mathbf{k}$. Similarly, if time reversal symmetry holds:

$$C_{n,\mathbf{k},\mathbf{G}} = C^*_{n,-\mathbf{k},-\mathbf{G}} \tag{29}$$

The complex conjugation occurs because time reversal is an antilinear operator.

2.3 The Band-Shifting Correction for Point Defect Calculations

For the purpose of benchmarking, a very basic perturbative band shifting method is presented here and referred to as the "Projection Shift." Because the perturbative band shifting correction adjusts single-particle energy levels, it is nontrivial to decide the correction to the total energy, even for a simplified correction like Equation 11. For this work, it is assumed that the shifted single-particle energies $e_{n,\mathbf{k}}$ given by Equation 11 only hold if $e_{n,\mathbf{k}}$ is in the band gap (i.e. it is

a defect level). Otherwise, the energy shifts entirely with the set of bands in which it is located (valence or conduction). This is reasonable because Equation 11 should only be applied to defect levels, but all of the bands shift in energy going from the GGA to hybrid level of theory.

To avoid counting energy shifts from occupied valence bands, a reference energy $N_{VB,X,0}\Delta E_{VBM}$ is chosen, where $N_{VB,X,0}$ is the number of electrons in the valence band of the neutral charge state of the defect. Note that this reference energy does not affect transition levels because transition levels are only dependent on the energy difference between charge states. This method prevents large double counting errors for the Coulomb energy while accounting for the energy shifts of the valence band, conduction band, and defect levels.

$$\begin{aligned} E_{corr,shift} = &- N_{VB,X,0}\Delta E_{VBM} + \sum_{n\in BG} \Delta e_n \\ &+ \sum_{n\in VB,\mathbf{k}} \omega_{\mathbf{k}} f_{n,\mathbf{k}} \Delta E_{VBM} + \sum_{n\in CB,\mathbf{k}} \omega_{\mathbf{k}} f_{n,\mathbf{k}} \Delta E_{CBM} \\ \Delta e_n = &\sum_{\mathbf{k}} \omega_{\mathbf{k}} f_{n,\mathbf{k}} (c_{n,\mathbf{k}} \Delta E_{CBM} + v_{n,\mathbf{k}} \Delta E_{VBM}) \end{aligned} \tag{30}$$

$\omega_{\mathbf{k}}$ are k-point weights and $f_{n,\mathbf{k}}$ are occupations. BG, VB, and CB are the band gap, valence bands, and conduction bands at the hybrid level of theory, respectively, and bands are assigned to them after their average energy over all k-points and spins is shifted by Δe_n. The averaging is performed because defect levels do not have dispersion in the true physical band structure. Spin polarization can be taken into account but does not significantly affect the results for the set of defects studied. This method was used for benchmarking the effectiveness of the perturbative band shifting correction.

3 Implementation

Pawpyseed is written in Python and C. All user interface is written in Python, and computationally expensive tasks are performed by using the ctypes package to call functions from a C library compiled during installation. The Math Kernel Library is used for Fast Fourier transforms (FFTs), and OpenMP is used to implement shared memory parallelization. Pseudo wavefunctions are read from VASP WAVECAR files using a method based on the WaveTrans program [15], and data about augmentation regions can be read from VASP pseudopotential (POTCAR) files. In addition to the core Python library, pawpyseed relies heavily on NumPy and SciPy [16] (for numerical operations), pymatgen [11] (for storing and manipulating structures, reading VASP input and output files, and performing symmetry analysis), and Matplotlib [17] (for visualization).

The rest of this section presents the numerical methods used to calculate overlap operators in the form of Equation 19. Table 1 gives detailed runtime scaling for each portion of the routine.

3.1 Overlap of Pseudo Wavefunctions (O_0)

The pseudo wavefunction is a summation of plane waves,

$$\widetilde{\psi}_{n\mathbf{k}}(\mathbf{r}) = \sqrt{\frac{1}{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{G}} C_{n\mathbf{kG}} e^{i\mathbf{G}\cdot\mathbf{r}} \tag{31}$$

so the overlap between two pseudo wavefunctions per unit cell can be written as

$$O_0 = \langle \widetilde{\psi}_{n_1\mathbf{k}_1} | \widetilde{\psi}_{n_2\mathbf{k}_2} \rangle = \delta_{\mathbf{k}_1,\mathbf{k}_2} \sum_{\mathbf{G}} C^*_{n_1\mathbf{k}_1\mathbf{G}} C_{n_2\mathbf{k}_2\mathbf{G}} \tag{32}$$

Table 1: Runtime scaling functions for each component of the code and definitions for shorthand symbols to express runtime. Approximate scales with the number of electrons are also shown.

Computational Task	Θ	Frequency*
O_0	$BKSW \sim n^2$	per band
O_M	$BKSNP \sim n^2$	per band
O_R and O_S	$BKSNP \sim n^2$	per band
O_N	$BKSNP \sim n^2$	per band
$\langle \widetilde{p}_i \vert \widetilde{\psi}_{n\mathbf{k}} \rangle$	$BKSF\log(F) \sim n^2\log(n)$	per structure
$(\langle \phi_i \vert - \langle \widetilde{\phi}_i \vert) \vert \widetilde{\psi}_{n\mathbf{k}} \rangle$	$BKSF\log(F) \sim n^2\log(n)$	per structure
spherical Bessel transform partial waves	$EPG\log(G) \sim 1$	per structure
projections for overlapping aug. spheres	$NP^2G\log(G) \sim n$	per structure pair

Symbol	Definition
B	number of bands
E	number of elements
F	size of FFT grid
G	size of partial wave radial grid
K	number of k-points
N	number of sites**
P	number of projector functions
S	number of spin states
W	number of plane waves
n	number of electrons (approximate scaling)

*The frequency refers to how often the routine is called. "Per band" indicates that the routine runs once every time a band from one structure is projected onto all the bands of a basis structure. "Per structure" indicates a setup routine used to set up the wavefunctions for a structure. "Per structure pair" is a setup routine run once for each pair of structures for which the overlap operators are to be calculated.

**Number of sites flexibly refers to the number of sites relevant to the calculation, which worst-case scales with the total number of sites in the structure. For example, calculating O_M and O_N only require the sites in sets M_{RS} and N_{RS}, respectively.

3.2 Concentric Augmentation Spheres (O_M)

Integrals of the type $\langle \widetilde{\psi}_{Rn_1\mathbf{k}} | \widetilde{p}_{alm\epsilon_1} \rangle$ between a pseudo wavefunction and projector function are evaluated using a real-space FFT grid, as with VASP with LREAL=TRUE [14, 18–21]. Integrals of the type $\langle \phi_{alm\epsilon_1} | \phi_{alm\epsilon_2} \rangle$ between partial waves are evaluated by simple radial integration. This is possible because the augmentation spheres are concentric, so the spherical harmonics for the partial waves are orthonormal.

3.3 Partial Waves Overlapping with Pseudo Wavefunction (O_R, O_S)

These integrals require projecting a smoothly varying pseudo wavefunction onto a rapidly varying AE partial wave. Performing such projections in reciprocal space is computationally expensive because of the large frequency components. Taking advantage of the orthogonality of plane waves, this projection can be done in real space. Since a plane wave can be expanded around an arbitrary

origin in space (to a phase factor) using Rayleigh expansion:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l j_l(kr) Y_l^{m*}(\hat{\mathbf{k}}) Y_l^m(\hat{\mathbf{r}}) \tag{33}$$

Partial waves can be Fourier transformed into reciprocal space by evaluating overlap integrals with spherical Bessel functions. This is done using the $O(NlogN)$ NUMSBT algorithm developed by Talman [22]. Then, all frequency components greater than the 1-dimensional FFT grid density can be set to 0 because those plane-waves are orthogonal to any component of the FFT grid. The "filtered" partial waves can then be transformed back into real space, also using the NUMSBT algorithm. This results in smooth partial waves for which $(\langle\phi_i| - \langle\widetilde{\phi}_i|)\,|\widetilde{\psi}_{n\mathbf{k}}\rangle$ can be evaluated in real space.

3.4 Partial Wave Overlap on Non-Orthogonal Augmentation Spheres (O_N)

The O_N term appears similar to the O_M term, except that the integrals $(\langle\phi_i| - \langle\widetilde{\phi}_i|)(|\phi_j\rangle - |\widetilde{\phi}_j\rangle)$ contain partial waves centered at different sites. However, transforming these partial waves into reciprocal space using the spherical Bessel transform allows the overlap integrals to be evaluated using Equation 47 in Talman's NUMSBT paper [22]. Evaluating this equation requires the Gaunt coefficients, which are calculated and stored using SymPy [23].

4 Computational Details

4.1 DFT Calculations

All DFT calculations were performed using the VASP code [18–21] with the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) functional [24, 25] and the PAW method [13, 14]. For hybrid functional bulk calculations, the Heyd-Scuseria-Ernzerhof (HSE) method [26] was used with the standard HSE06 [27] tuning parameters. All k-point grids used were Monkhorst-Pack (MP) grids [28], and all calculations were performed with symmetry off to allow the local environment of the defect to relax out of its initial symmetry.

The boron, phosphorus, copper, and sulfur substitutionals and single vacancy were embedded in a 216-atom supercell of silicon using PyCDT [8]. PyCDT was also used to generate charge states for analysis. Energy convergence tests were performed with the number of k-points. For the total cohesive energy of the bulk crystal, a difference of 0.01 eV (5×10^{-5} eV/atom) was found between a 2x2x2 and 3x3x3 k-point mesh for the bulk supercell, so a 2x2x2 k-point mesh was used. The energy cutoff was set to 520 eV. An ionic relaxation and total energy calculation were performed for each defect and charge state. The static dielectric constant was also calculated for the bulk silicon crystal using VASP and found to be 12.52. For the silicon vacancy, even with symmetry turned off, the structure does not relax to a stable state because it is caught in a metastable minimum. The final symmetry is different for each charge state, but as an example, the neutral defect has a tetragonally distorted structure with D_{2d} symmetry [29]. To attain the correct structure, the positions of the four atoms nearest the vacancy site are perturbed manually to break the structure from the metastable minimum and break symmetry before the calculation is run. This process can also be done in an automated fashion and therefore is compatible with a high-throughput computing workflow. To calculate DFT and hybrid band edges, a DFT calculation was run on a single 2-atom silicon primitive cell with a 7x7x7 k-point mesh, once with the GGA functional and once with the HSE06 functional.

4.2 Correction Methods

The formation energies and transition levels were calculated using PyCDT [8]. Using the PyCDT utilities in pymatgen [11], several combinations of charge corrections were tested for prediction of formation energies and transition levels, which are elaborated below. The new code presented here, pawpyseed, was used to perform the simplified perturbative band shifting correction of Equation 30. Visualization of results was performed using PyCDT. All of the correction schemes tested are described below. Note that for all schemes that involve band filling or band shifting, the GGA band edges are determined from the primitive cell DFT calculation with a 7x7x7 MP mesh because the bulk supercell calculation with 2x2x2 mesh does not sample the VBM and CBM.

Freysoldt Correction Only No band shifting or band filling is attempted for this method. Only the electrostatic Freysoldt Correction is used. All other correction schemes add terms to this one.

Freysoldt and Band Filling In addition to the Freysoldt correction, the band filling correction is used to shift electrons in the conduction band down to the (GGA) conduction band minimum and holes in the valence band to the (GGA) valence band maximum.

Freysoldt, Band Filling, 100% Shift On top of the Freysoldt and Band Filling corrections, the 100% shift method discussed in the introduction is used.

Freysoldt, Band Filling, Projection Shift This correction makes use of pawpyseed to attempt a more sophisticated band shifting correction. Equation 30 is used as described in Section 2.3 to correct each energy level and extrapolate these single-particle corrections to a total energy correction.

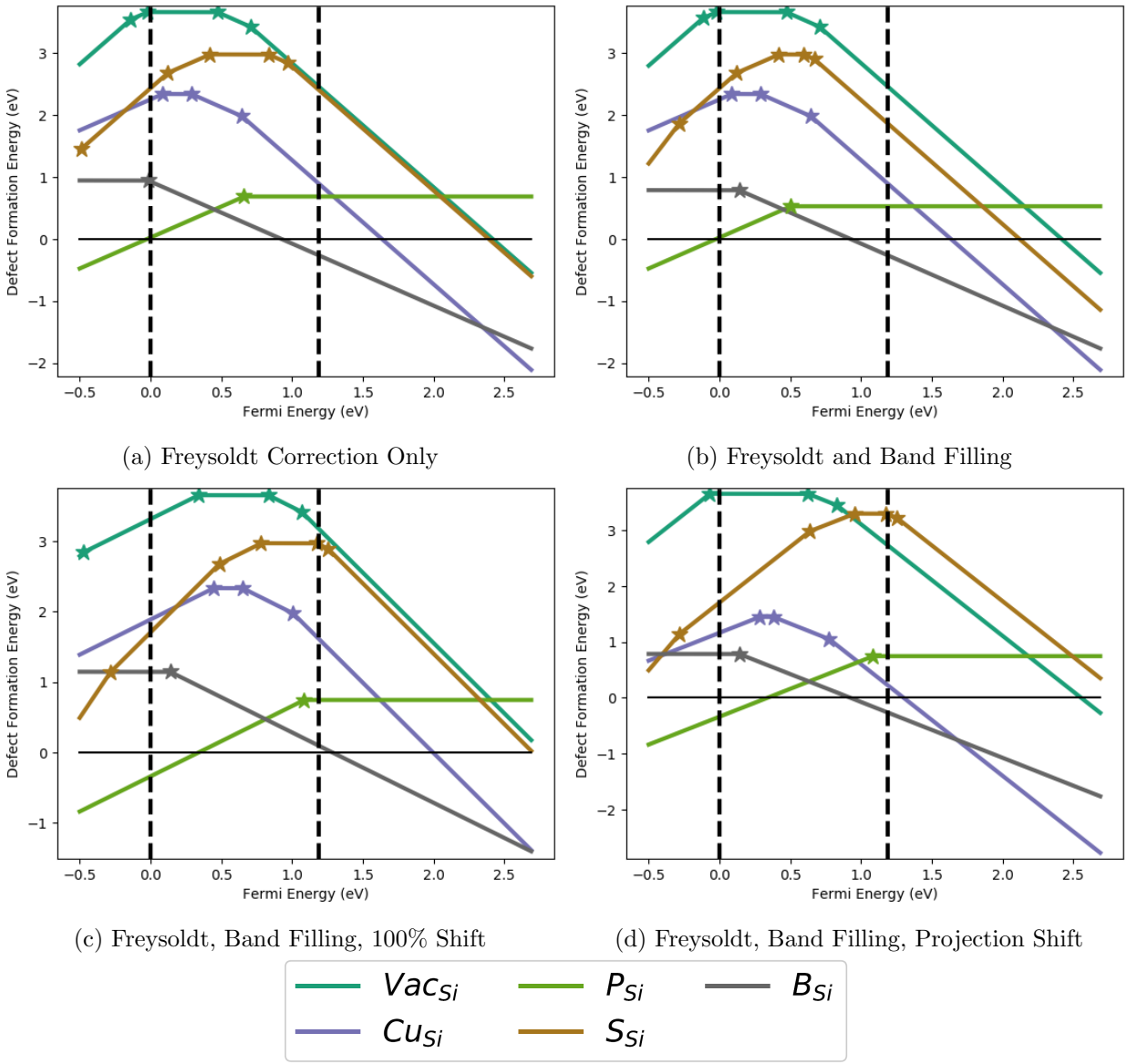
5 Results

5.1 Defect Transition Level Predictions

This section presents and discusses the predicted transition levels of the phosphorus, boron, copper, and sulfur substitutionals and single vacancy in silicon. Plots of the formation energies as a function of Fermi level E_F are shown for all correction schemes in Figure 2. Only the transition levels are discussed in detail below; analysis of correction method performance for formation energy prediction is left for future work. For reference, the GGA band gap calculated in this study is 0.61 eV. The HSE06 band gap is 1.19 eV and is denoted with black dotted lines on the formation energy plots.

Boron Substitutional The boron substitutional is a shallow single acceptor with an experimental transition level at 0.045 eV above the VBM [30]. It is commonly used as a dopant to make p-n junctions for computer chips and photovoltaics. Because boron is such a shallow acceptor, the electron density of the defect level will be delocalized beyond the 216-atom supercell used, which will decrease the accuracy of the supercell method DFT calculation. However, the boron substitutional is a good benchmark to observe the qualitative behavior of the correction methods, particularly band filling and band shifting. The Freysoldt correction only method (Figure 2a) predicts a slightly negative transition level (-0.02 eV). The band filling correction increases this transition level to 0.14 eV, which is unchanged by the band shifting corrections.

Phosphorus Substitutional The phosphorus substitutional is a shallow single donor with an experimental transition level at 0.045 eV below the CBM [30]. As with boron, it is often used in electronics applications. Because the extra electron in the phosphorus defect is delocalized, the supercell method is not expected to give accurate results. However, attempting the band shifting correction on a shallow level provides a simple benchmark to ensure that the method can shift a shallow donor with the conduction band edge. This is achieved by both the 100% shift correction



(a) Freysoldt Correction Only

(b) Freysoldt and Band Filling

(c) Freysoldt, Band Filling, 100% Shift

(d) Freysoldt, Band Filling, Projection Shift

Figure 2: Formation energy and transition level diagrams for the studied defects in Silicon, with different sets of corrections applied to each. The transition levels (indicated in the diagrams by star symbols) occur at the Fermi levels for which two charge states have equal formation energy. The 2x2x2 Monkhorst-Pack k-point mesh was used.

(Figure 2c) and the projection shift correction (Figure 2d). The transition level is 1.08 eV (0.11 eV below the CBM) for the 100% shift and for the projection shift.

Copper Substitutional Copper is a technologically important impurity in silicon because it increases leakage current in transistor devices and is common in semiconductor processing environments [31–33]. It exhibits both deep acceptor and deep donor character, with experimental transition levels at 0.207 (+/0) and 0.478 (0/-) eV above the VBM and 0.167 (-/--) eV below the CBM [34]. High-resolution Laplace-transform DLTS gave transition levels for +/0 and 0/- at 0.225 and 0.430 eV [35]. Using the HSE06 functional, Sharan, Gui, and Janotti predicted transition levels for the copper substitutional at 0.20 eV for +/0, 0.54 eV for 0/-, and 0.97 eV for -/-- [31].

The Freysoldt only and Freysoldt plus band filling methods give good predictions of the copper levels relative to the size of the band gap. The band shifting corrections give relatively good extrapolations of these levels to the true band gap (0.44 and 0.65 eV for 100% shift and 0.28 eV and 0.38 eV for the projection shift), with the 100% shift overestimating the transition levels and the projection shift underestimating the difference between the levels.

Sulfur Substitutional Sulfur is a deep donor with a level 0.32 eV below the CBM (+/0) and a level 0.51 eV above the VBM (++/+) [30]. The Freysoldt correction significantly underestimates the ++/+ level, placing it at 0.12 eV, which makes it appear like a shallow defect. In addition, the 0/- and -/-- transitions are incorrectly placed in the true band gap (though not the GGA band gap) at 0.83 and 0.97 eV. However, the ++/+ and +/0 levels are approximately the correct energy below the GGA conduction band minimum ($E_{g,GGA}$=0.612 [36]). Band filling corrections do not significantly affect the results.

With band shifting, agreement is found with experiment to about 0.1 eV (0.49 and 0.78 eV for 100% shift and 0.64 and 0.95 eV for projection shift), with slightly lower error for the 100% shift. The corrections both increase the 0/- and -/-- levels to 1.17/1.18 eV and 1.25 eV. The 0/- transition is 0.02 and 0.01 eV below the CBM for the 100% shift and projection shift, respectively, which is close enough to infer the instability of the -1 charge state.

Single Vacancy The single vacancy in silicon is known to be a donor with negative-U behavior which makes the +/0 transition level lower than the ++/+ transition level [37]. The +/0 transition is at 0.05 eV, and the ++/+ transition is at 0.13 eV [37]. LDA fails to capture this negative-U behavior [29, 38]. GGA was able to capture the negative-U behavior in one study [29]. The correction method used in that study chose ΔE_{VBM} to reproduce the ++/+ transition level, which is not a reliable strategy for predicting properties of previously unstudied defects. However, the choice of ΔE_{VBM} does not affect the difference between transition levels, so the negative-U behavior should be reproducible. The study did not use band filling or shifting corrections, so the behavior should be reproducible with only the Freysoldt correction. This is not the case for the current set of calculations, however (see Figure 2a), suggesting inadequate DFT parameters. Wright [29] used a 5x5x5 k-point mesh for the 215-atom supercell. This draws attention to the fact that parameters sufficient for total energy convergence in the bulk supercell are not necessarily sufficient for total energy convergence of the defect. To avoid errors like this, a high-throughput workflow would need to do convergence test calculations for the total energy of each defect and charge state, which is likely too computationally expensive for some applications. Interestingly, the projection shift correction method predicts negative-U behavior for the single vacancy, but the ++/+ and +/0 levels are incorrectly predicted to be negative.

The silicon vacancy also illustrates one of the limitations of the 100% shift correction. This method requires distinguishing valence and conduction band states from defect levels, and then shifts the valence and conduction states by a significant amount (in this case, -0.36 eV for the VBM shift and 0.22 eV for the CBM shift). Therefore, a small numerical error in the energy eigenvalues

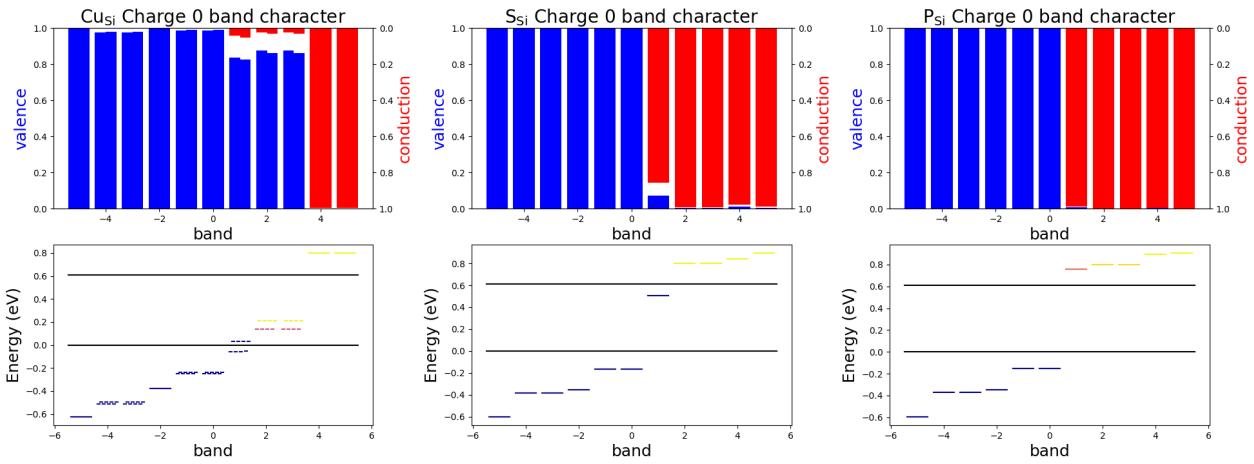


Figure 3: v_D and c_D for states near the band gap of several neutral defects. The upper portion of the plot shows proportion valence (blue) and proportion conduction (red). The lower part shows the energy levels at the k-points and spin states of each band. The color of the energy levels corresponds to the occupation, with the plasma color scale provided in Matplotlib [17] used to clearly illustrate partial occupancies. Dark blue levels have occupancy 1 and yellow levels have occupancy 0. The GGA band gap of 0.61 eV is denoted with the black lines.

or the Freysoldt potential alignment (Equation 2) can change the transition level significantly by changing whether a state is classified as a conduction, valence, or defect state. This appears to be the case with the silicon vacancy, where the Si ++/+ level is shifted down to -0.47 eV, 0.33 eV below its value with the Freysoldt correction only, because the +2 charge state contains an unoccupied level just below the VBM (Figure 4).

5.2 Understanding Energy Corrections Using Level Projections

Figure 3 shows the valence band character v_D and conduction band character c_D for levels near the band gap of the neutral copper, sulfur, and phosphorus substitutionals. Note that because the bulk basis set is finite and therefore incomplete, $v_D + c_D$ can be less than 1. The missing character, represented by white space between the red and blue bars in Figure 3, is due to the difference in the basis set between the bulk and defect. The defect contains different atomic wavefunctions at different coordinates than the bulk, and the set of plane-waves incorporated into the bands is not exactly the same in the two structures. For example, the bulk silicon basis contains no d states, so the white space in the Cu substitutional projection diagram might be due to d state character.

These diagrams are useful for reaching qualitative conclusions about the chemistry of the point defects. For example, in the neutral and +1 sulfur substitutional, the occupied defect level is derived mainly from the conduction band. This indicates that higher-level of theory correlation effects could raise the energy of this level and therefore raise the transition levels compared to the Freysoldt correction-only GGA prediction. Indeed, the basic Freysoldt correction scheme significantly underestimates the sulfur transition levels, and this underestimation is remedied by both band shifting corrections.

In addition, the near-100% conduction character of the phosphorus substitutional level makes sense given how shallow the defect is. The strong valence character of the copper defect levels

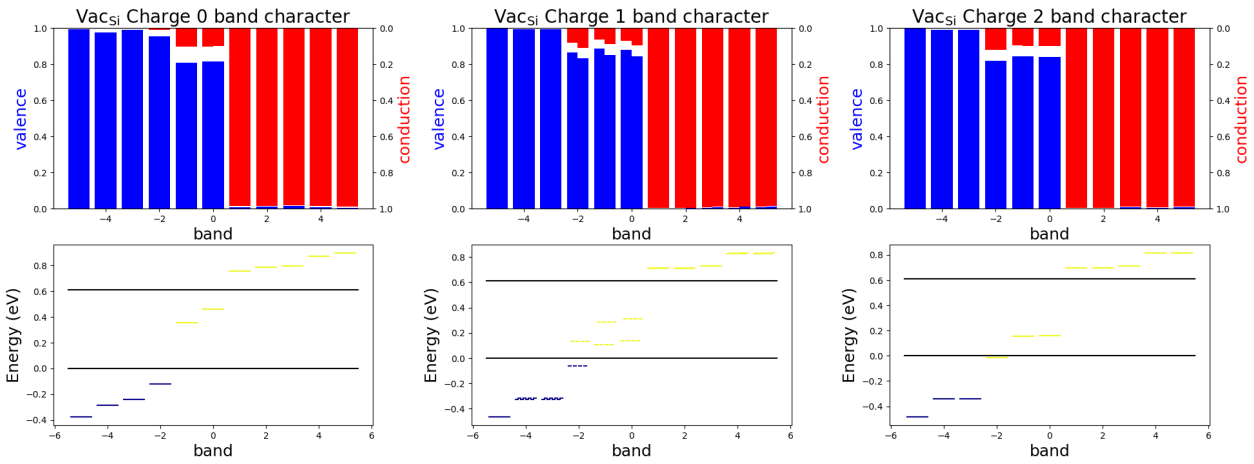


Figure 4: v_D and c_D for states near the band gap of several charge states of the silicon vacancy. The highest occupied band of the neutral vacancy has a larger v_D than the same band in the positive charge states, and this band is lower in energy than in the positive charge states.

confirm previous conclusions that these levels are not derived from the d shell, but rather ligand orbitals [31], since the d orbitals are not contained in the bulk silicon basis set.

Figure 4 shows the valence-conduction projection diagrams for the 0, +1, and +2 charge states of the silicon vacancy. When band -2 is unoccupied, it has $v_C = 0.82$. When it becomes occupied in the 0 charge state, is has $v_C = 0.95$. This suggests that the tetragonal distortion of the neutral vacancy allows this band to adopt more valence-type bonding character, which stabilizes it. This is consistent with the negative-U behavior in the silicon vacancy, which causes the +/0 level to be lower than the ++/+ level.

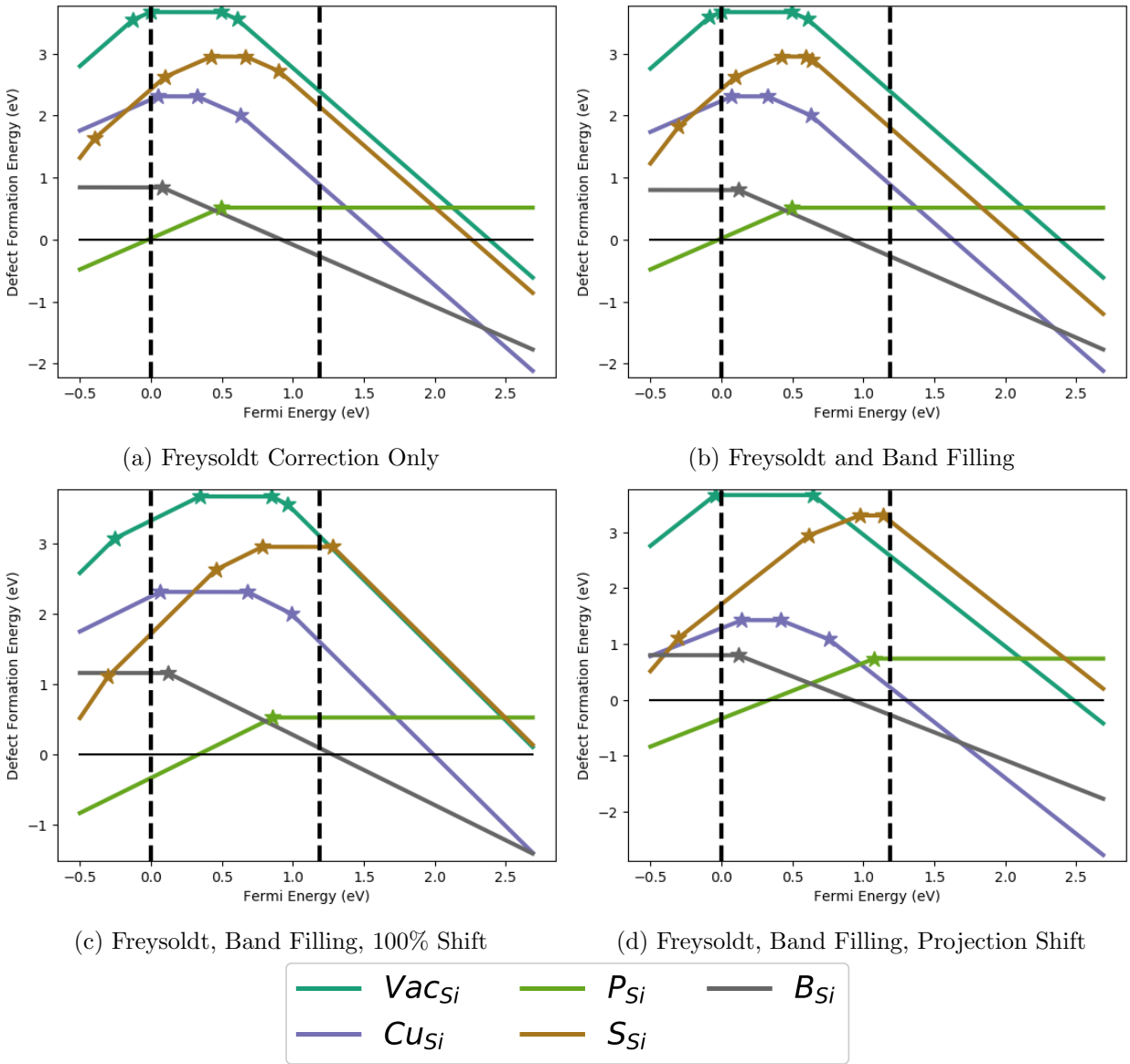
The energy level diagram for the +2 state illustrates why the ++/+ transition level is 0.47 eV below the VBM; the lowest unoccupied state is just slightly below the VBM, so that empty state gets treated as a hole and shifted with the VBM, which inaccurately increases the energy of the +2 state.

5.3 Γ-centered K-point Mesh DFT Calculations

Because the 2x2x2 k-point mesh used in this study does not sample the band edges of silicon, a second set of calculations was performed in which the 2x2x2 mesh was Γ-centered. However, this change does not result in significantly more accurate transition level and formation energy predictions. See Figure 5 for formation energy plots for the band shifting corrections. This suggests that the errors in transition levels in this study are not due simply to the k-point mesh generation method. The density of the k-point mesh may be too low for some systems (such as the vacancy), and the limitations of the GGA functional also limit the accuracy of the transition level predictions.

5.4 Quantitative Comparison of 100% Shift and Projection Shift

Tables 2 and 3 compare the energies of experimentally observed transition levels in the studied defects to the theoretical predictions in this work. The comparison is performed for the shifted Monkhorst-Pack k-point mesh calculations (Table 2) and the Γ-centered k-point mesh calculations (Table 3) to investigate the effects of the corrections for different choices of k-point sampling. To



(a) Freysoldt Correction Only

(b) Freysoldt and Band Filling

(c) Freysoldt, Band Filling, 100% Shift

(d) Freysoldt, Band Filling, Projection Shift

Figure 5: Formation energy and transition level diagrams for the studied defects in Silicon. As opposed to Figure 2, the 2x2x2 Γ-centered k-point mesh was used.

Table 2: Errors of predicted transition levels in eV for the DFT calculations performed with MP k-point meshes. Format: "level in eV (error as percentage of band gap)."

Transition Level	Expt.	Freysoldt Only	Freysoldt/B.F.	100% Shift	Projection Shift
Vac ++/+	0.13 [37]	-0.14 (-34%)	-0.12 (-31%)	-0.47 (-51%)	-0.04 (-15%)
Vac +/0	0.05 [37]	-0.02 (-8%)	-0.02 (-7%)	0.34 (24%)	-0.09 (-12%)
P +/0	1.07 [30]	0.66 (12%)	0.50 (-14%)	1.09 (-5%)	1.09 (-5%)
B 0/-	0.04 [30]	-0.02 (-7%)	0.14 (19%)	0.14 (8%)	0.14 (8%)
Cu +/0	0.23 [35]	0.09 (-6%)	0.09 (-6%)	0.44 (17%)	0.28 (4%)
Cu 0/-	0.43 [35]	0.29 (9%)	0.29 (9%)	0.65 (16%)	0.38 (-7%)
Cu -/--	0.95 [34]	0.65 (20%)	0.65 (20%)	1.01 (-1%)	0.77 (-20%)
S ++/+	0.51 [30]	0.12 (-26%)	0.12 (-26%)	0.49 (-5%)	0.64 (8%)
S +/0	0.80 [30]	0.42 (-2%)	0.42 (-2%)	0.78 (-6%)	0.95 (9%)
MAE		14%	15%	15%	9%
MAE (excl. vac)		12%	14%	8%	9%

Table 3: Errors of predicted transition levels in eV for the DFT calculations performed with Γ-centered k-point meshes. Format: "level in eV (error as percentage of band gap)."

Transition Level	Expt.	Freysoldt Only	Freysoldt/B.F.	100% Shift	Projection Shift
Vac ++/+	0.13 [37]	-0.13 (-32%)	-0.09 (-26%)	-0.25 (-33%)	-0.01 (-13%)
Vac +/0	0.05 [37]	-0.00 (-5%)	-0.00 (-5%)	0.34 (24%)	-0.08 (-11%)
P +/0	1.07 [30]	0.50 (-15%)	0.50 (-15%)	0.86 (-24%)	1.08 (-6%)
B 0/-	0.04 [30]	0.08 (9%)	0.12 (15%)	0.12 (6%)	0.12 (6%)
Cu +/0	0.23 [35]	0.05 (-12%)	0.07 (-8%)	0.06 (-15%)	0.14 (-8%)
Cu 0/-	0.43 [35]	0.32 (14%)	0.32 (14%)	0.68 (19%)	0.42 (-4%)
Cu -/--	0.95 [34]	0.63 (18%)	0.63 (18%)	0.99 (-2%)	0.76 (-21%)
S ++/+	0.51 [30]	0.10 (-29%)	0.10 (-29%)	0.46 (-7%)	0.62 (6%)
S +/0	0.80 [30]	0.43 (-1%)	0.43 (-1%)	0.78 (-6%)	0.98 (11%)
MAE		15%	15%	15%	9%
MAE (excl. vac)		14%	14%	11%	9%

compare the theoretical transition levels to the experimental levels, each transition level is divided by the band gap for the respective level of theory. This means non-band-shifted method predictions get divided by the GGA band gap of 0.612 eV, band-shifted predictions by the HSE06 band gap of 1.192 eV, and experimental predictions by the experimental band gap of 1.12 eV. Errors are then reported as the difference between the percentage of the band gap of the theoretical and experimental levels. Sze and Ng's compilation of impurity levels is used for the boron, phosphorus, and sulfur substitutionals [30]. The copper data is taken from [35] and [34] because the levels are assigned and the data is more recent. The vacancy data is taken from [37]. Below is a list of notable performance differences between the methods.

- On average, only the projection shift method outperforms the standard Freysoldt correction when the vacancy is included, but the 100% shift is also an improvement when the vacancy is ignored.
- The Freysoldt correction errors are not systematically positive or negative. Performing a simple VBM shift of the Fermi level increases all transition levels by the same amount (assuming

the VBM shift is negative), so the VBM shift alone is inadequate to correct the errors of the Freysoldt correction.

- The projection shift underestimates the copper -/-- level compared to the 100% shift.
- The 100% shift gives poor transition level predictions for the vacancy because the lowest unoccupied orbital is incorrectly assigned as a hole.
- The projection shift predicts negative-U behavior for the silicon vacancy ++/+ and +/0 levels (though the energies are unphysically negative).
- The 100% shift overestimates the copper +/0 and 0/- levels
- The phosphorus donor level is slightly below the conduction band in the Γ-centered calculation, so it is not shifted with the conduction band by the 100% shift method. This results in a large underestimation of the transition level, which is remedied by the projection shift method.

6 Discussion

6.1 Effectiveness of Band Filling and Band Shifting Corrections

It should be noted that the set of defects studied here is quite small, and only one system is studied. The purpose of this work is to identify possible strengths and drawbacks of the different correction methods and present utilities to perform these corrections in a standardized way. The data collected is certainly not adequate to draw strong general conclusions about any correction method, and no attempt is made to do so.

For the set of defects studied here, the band shifting correction methods provide modest decreases in the average error of the transition levels. The 100% shift method improves most transition levels but produces large errors for the vacancy because an unoccupied defect level in the +2 state gets labeled as a hole in the valence band and shifted in energy. The projection shift gives a net decrease in transition level errors compared to all other methods, but it gives higher errors than other methods for some levels. The 100% shift method has the advantage of simplicity and low computational cost, but it also requires identifying valence band and conduction band states in order to shift energies. As demonstrated by the ++/+ transition in the vacancy and the Γ-centered calculation of the phosphorus substitutional, this can cause significant errors. The projection shift method is more computationally expensive but can shift levels in the band gap as well as valence and conduction states.

The projection of defect levels onto conduction and valence bands does not provide complete information about how the energy of the defect level will shift when attempting to extrapolate to a more accurate band structure from DFT. One reason for this is that the chemical environment of an electron in a given state is significantly different in the defect than in the bulk, so the correction to the XC term could be very different. One way to account for the local chemical environment around the defect is to partly base the shift in energy of defect levels on the projection onto local atomic wavefunctions around the defect. For example, the *s*-type defect states around the copper substitutional have mainly valence character (80-90%), which results in the perturbative correction presented here shifting the energy of the levels down, which makes the correction less accurate for the -/-- level. However, one might predict that additional repulsion by the *d* states in copper makes the XC energy less negative than expected from the XC energy of the bulk. If the XC energy correction included a contribution from the XC correction for atomic states of copper, the accuracy of the correction might be improved.

6.2 Other Applications of Pawpyseed

Pawpyseed gives a user easy access to the all electron (AE) wavefunctions from the PAW method, which opens up the possibility that it can be used as a general tool for analyzing PAW DFT wavefunctions. The following are some potential applications of the formalism developed in this paper.

Wavefunction Visualization Pawpyseed currently contains utilities which can be used to visualize the AE wavefunctions of a VASP calculation. The user can select either the charge density of a given state or the wavefunction itself (with one output file each for the real and imaginary parts). The user can also choose real-space grid dimensions for printing charge density files, which enables printing AECCAR-style files. All volumetric data file output from pawpyseed is formatted like VASP volumetric data for easy visualization in tools like VESTA [39].

Population Analysis Mulliken population analysis [40] is a method used in quantum chemistry to assign partial charges to atoms for analysis purposes. This method, as well as more modern methods derived from it [41], require projecting wavefunctions of a structure onto localized atomic wavefunctions. Such tools are common in localized basis set codes but not in plane-wave basis set codes. The tools in pawpyseed can be extended to perform population analysis on plane-wave DFT output.

Wavefunction Type Conversion Some codes only support certain types of wavefunction formats. For example, some GW codes, such as BerkeleyGW, require as input the wavefunctions from norm-conserving (NC) pseudopotential calculations [42]. The framework in pawpyseed can be used to implement a mapping between PAW wavefunctions and norm-conserving (NC) or ultrasoft (US) pseudopotential wavefunctions.

Orbital Localization Certain orbital localization procedures, such as the SCDM and SCDM-k methods [43, 44], are still in need of efficient open-source implementations. This could be conveniently provided by pawpyseed by using the existing computational framework, as it already supports reading, real-space projection, and visualization of PAW wavefunctions.

Optical Property Calculations When a solid or molecule is electronically excited by a photon, the ionic structure can undergo a relaxation in addition to the electronic structure. Pawpyseed's utilities can be extended to evaluate the electric field operator matrix elements between wavefunctions in the ground state and excited state structures.

General Nonlinear Operators The algorithm developed in this paper is sufficient for the evaluation of general nonlocal operators in the PAW formalism, so applications which require the evaluation of nonlocal operators can use the framework developed in pawpyseed.

7 Summary and Conclusions

An open-source code has been presented which can calculate the overlap between PAW wavefunctions from different structures. This utility has been used to implement a perturbation theory-based energy correction for charged point defect DFT calculations. There are multiple ways to derive a band shifting correction from the general perturbation theory approach, and the utilities in pawpyseed enable the evaluation of overlap operators necessary for these approaches, which provides other researchers with a fast and easy way to develop new corrections. The parallelization and other optimizations in pawpyseed execute the computationally expensive overlap operator evaluations quickly. The framework developed in pawpyseed is capable of evaluating general nonlinear operators,

and development plans include expanding pawpyseed to support utilities not available in other PAW codes and post-processing packages.

Pawpyseed, PyCDT, pymatgen, and VASP were used to perform a benchmarking study of several correction schemes for charged defect calculations, including the new projection shift method presented here. The results suggested that the projection shift method improves the average error of the transition levels for the studied defects. The 100% shift method improves most transition levels compared to the Freysoldt-only correction but sometimes produces large errors.

Future development plans for pawpyseed include improving the precision of overlap operator evaluations, particularly the O_R and O_S terms, which currently limit the precision of the overlap operator evaluation to 10^{-2}. Future work on the perturbative band shifting method will focus on accounting for XC effects from changes to the local chemical environment introduced by defects.

8 Acknowledgments

I would like to thank my advisor, Mark Asta, and my PhD student advisor, Danny Broberg, for the opportunity to perform this research and for guidance and support for this project. I also thank the UC Berkeley College of Chemistry for an undergraduate research stipend for the summer of 2017. This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer).

References

(1) Freysoldt, C.; Grabowski, B.; Hickel, T.; Neugebauer, J.; Kresse, G.; Janotti, A.; Van De Walle, C. G. *Reviews of Modern Physics* **2014**, *86*, 253–305.

(2) Yin, W. W.-j.; Yang, J. J.-h.; Kang, J.; Yan, Y.; Wei, S.-h. *J. Mater. Chem. A* **2015**, *3*, Advance.

(3) Lany, S.; Zunger, A. *Phys. Rev. B* **2008**, *78*, 235104.

(4) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, DOI: 10.1103/PhysRev.140.A1133.

(5) Komsa, H. P.; Rantala, T.; Pasquarello, A. In *Physica B: Condensed Matter*, 2012; Vol. 407, pp 3063–3067.

(6) Meggiolaro, D.; De Angelis, F. *ACS Energy Letters* **2018**, DOI: 10.1021/acsenergylett.8b01212.

(7) Deák, P.; Duy Ho, Q.; Seemann, F.; Aradi, B.; Lorke, M.; Frauenheim, T. *Physical Review B* **2017**, *95*, DOI: 10.1103/PhysRevB.95.075208.

(8) Broberg, D.; Medasani, B.; Zimmermann, N. E.; Yu, G.; Canning, A.; Haranczyk, M.; Asta, M.; Hautier, G. *Computer Physics Communications* **2018**, DOI: 10.1016/j.cpc.2018.01.004.

(9) Freysoldt, C.; Neugebauer, J.; Van De Walle, C. G. *Physical Review Letters* **2009**, DOI: 10.1103/PhysRevLett.102.016402.

(10) Kumagai, Y.; Oba, F. *Physical Review B - Condensed Matter and Materials Physics* **2014**, DOI: 10.1103/PhysRevB.89.195205.

(11) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. *Computational Materials Science* **2013**, DOI: 10.1016/j.commatsci.2012.10.028.

(12) Bogusl/awski, P.; Briggs, E. L.; Bernholc, J. *Phys. Rev. B* **1995**, *51*, 17255–17258.

(13) Blochl, P. E. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(14) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(15) Feenstra, R. M.; Widom, M. WaveTrans: Real-space wavefunctions from VASP WAVECAR file.

(16) Jones, E.; Oliphant, T.; Peterson, P.; Others SciPy: Open source scientific tools for Python., 2007.

(17) Hunter, J. D. *Computing in Science and Engineering* **2007**, DOI: 10.1109/MCSE.2007.55.

(18) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558–561.

(19) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49*, 14251–14269.

(20) Kresse, G.; Furthmller, J. *Computational Materials Science* **1996**, *6*, 15 –50.

(21) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(22) Talman, J. *Computer Physics Communications* **2009**, *180*, 332 –338.

(23) Meurer, A. et al. *PeerJ Computer Science* **2017**, DOI: 10.7717/peerj-cs.103.

(24) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(25) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(26) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *Journal of Chemical Physics* **2003**, DOI: 10.1063/1.1564060.

(27) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. *Journal of Chemical Physics* **2006**, DOI: 10.1063/1.2404663.

(28) Monkhorst, H. J.; Pack, J. D. *Physical Review B* **1976**, DOI: 10.1103/PhysRevB.13.5188.

(29) Wright, A. F. *Phys. Rev. B* **2006**, *74*, 165116.

(30) Sze, S. M.; Ng, K. K., *Physics of Semiconductor Devices*, 2007, pp 164, 682.

(31) Sharan, A.; Gui, Z.; Janotti, A. *Physical Review Applied* **2017**, DOI: 10.1103/PhysRevApplied.8.024023.

(32) Graff, K. *Springer Series in Materials Science* **1995**, *24*, 143.

(33) Böhm, R; Klose, H *physica status solidi (a)*, *9*, K165–K168.

(34) Knack, S; Weber, J; Lemke, H *Physica B: Condensed Matter* **1999**, *273-274*, 387 –390.

(35) Yarykin, N.; Weber, J. *Phys. Rev. B* **2013**, *88*, 085205.

(36) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. a. *APL Materials* **2013**, *1*, 011002.

(37) Watkins, G. D.; Troxell, J. R. *Physical Review Letters* **1980**, DOI: 10.1103/PhysRevLett.44.593.

(38) Schultz, P. A. *Physical Review Letters* **2006**, DOI: 10.1103/PhysRevLett.96.246401.

(39) Momma, K.; Izumi, F. *Journal of Applied Crystallography* **2011**, DOI: 10.1107/S0021889811038970.

(40) Mulliken, R. S. *The Journal of Chemical Physics* **1955**, *23*, 1833–1840.

(41) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *The Journal of Chemical Physics* **1985**, DOI: 10.1063/1.449486.

(42) Deslippe, J.; Samsonidze, G.; Strubbe, D. A.; Jain, M.; Cohen, M. L.; Louie, S. G. *Computer Physics Communications* **2012**, *183*, 1269–1289.

(43) Damle, A.; Lin, L.; Ying, L. *Journal of Chemical Theory and Computation* **2015**, *11*, 1463–1469.

(44) Damle, A.; Lin, L.; Ying, L. *Journal of Computational Physics* **2017**, *334*, 1–15.