Automated Optimization of Pseudopotentials for Faster and More Accurate Plane Wave

Density Functional Theory Calculations in Materials

By

Casey Neil Brock

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Interdisciplinary Materials Science

May 31, 2019

Nashville, Tennessee

Approved:

D. Greg Walker, Ph.D.

Kalman Varga, Ph.D.

Kirk H. Bevan, Ph.D.

Cary L. Pint, Ph.D.

Peter T. Cummings, Ph.D.

To my family, for your love and support.

ACKNOWLEDGMENTS

I would first like to thank my advisor, Greg Walker, and my committee members, Kalman Varga, Kirk Bevan, Cary Pint, and Peter Cummings for guiding me during the pursuit of my Ph.D. Greg, I am grateful for your mentorship, your flexibility as I travelled the world in the name of science (mostly), and your patience during my careful convergence tests and slow manuscript revisions. I would also like to thank my Sandia mentor, Alan Wright, for his invaluable guidance during the research that became a large part of this dissertation, and Alan Tackett for answering my pseudopotential questions during many lunch outings. To my fellow Thermal Engineering Lab members, Matt, Drew, and Becca, thanks for listening to my practice talks.

I have had the opportunity to work with some very bright scientists. First, I'd like to thank Professor Varga for selecting me for the International Research Experience for Students at McGill University in Montréal, and Professor Bevan for his close mentorship while I was there. I'd also like to thank Nitin Muralidharan and Professor Pint for their collaboration studying battery materials and my second mentor at Sandia, Thomas Beechem, for showing me the ways of phonons and infrared optics.

I would like to thank my family for their unwavering support and encouragement, from awards ceremonies to band concerts to graduations. I have been truly fortunate to have so many people by my side from the beginning. I would especially like to thank my mom, who has sacrificed so much to make me the person I am today; my dad, for bragging about me to his audiences at shows; my sister Billie Jean for always believing in me; Grandmama, for all the trips to the library; Charlie, for fostering my passion for technology in middle school; Aunt Jackie and Uncle Chuck, who slipped me some money every month in college on the condition I put them in a nice retirement community one day (a smart investment on their part); Grandpa Bob, for making sure my cars always started; Jim and Laura, for teaching a young me how to approximate the area of a shape using rectangles (it would be

years before I learned to do it rigorously); and Uncle Bobby for teaching me to weld, which is not really relevant to this work but now I have a nice bed frame.

My friends have also been invaluable throughout the years, both for support and guidance, and also for debaucherous fun. There have been so many nights of loud music, smokey cocktails, video games, climbing, and island camping. And many mornings of freshly roasted coffee and breakfast feasts. These will no doubt remain some of my fondest memories.

And of course, many thanks to Courtney for standing by my side during the preparation of this dissertation, through the long days of writing and the many months apart gathering data in the desert, and for doing all the laundry since January. I suppose it's good grad school took so long, or I never would have met you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction and Background

Electronic structure calculations play a critical role in materials modeling by capturing phenomena at the quantum scale, and likely owe much of their success in materials research to speed improvements enabled by the pseudopotential approximation to density functional theory (DFT).[2,3] However, new pseudopotentials still need to be developed to meet the accuracy and speed demands of the materials design framework. The goal of our research is thus to improve on previous generations of pseudopotentials by improving the way pseudopotentials are evaluated, and to leverage this knowledge to enable automated searches for optimal pseudopotential parameters. The scientific impact of our work will be faster and more accurate pseudopotentials, enabling high-throughput calculations efficient enough to support the new materials by design paradigm and accurate enough to probe uncharted areas of the materials design space.

The pseudopotential is an effective atomic potential that speeds up DFT calculations by "freezing" core electrons so that only the valence electrons need to be calculated in the system. This frozen-core approximation is justified because the valence electrons are involved in bonding while the core electrons are relatively inert. Additionally, freezing the core electrons can smooth the valence wave functions, enabling further speedup of the calculations by reducing the number of plane waves needed to represent them in a plane wave basis set. Pseudopotentials were originally introduced in 1959[4] and several implementations have since been developed, including norm-conserving pseudopotentials, ultrasoft pseudopotentials, and the projector augmented wave method.[5,6,7] Since the 1990s efforts have focused less on new kinds of pseudopotentials and more on tuning the parameters of these pseudopotentials to make them faster and more accurate. This has resulted in several popular pseudopotential libraries.

A pseudopotential library is a collection of pseudopotentials for a large number of elements, sometimes arranged in a periodic table format for ease of use. For each element, a typical library has one or two pseudopotentials per exchange-correlation functional. The choices of exchange-correlation functional are usually limited to common parameterizations of the local density approximation (LDA) and generalized gradient approximation (GGA). The clearest advantage of libraries is the convenience of pseudopotentials ready for use. Another potential advantage is that the widespread use of standard sets of pseudopotentials allows the community to discover any flaws that exist, enabling improvement of libraries over time.[8] In recent years there has been a renewed interest in pseudopotential development and a surge in new libraries as metrics, testing protocols, and researcher needs evolve. Some recent libraries include the JTH library of PAWs packaged with the Abinit code,[9] the SG15 library of norm conserving pseudopotentials created using optimization techniques,[10] the EPAW library which was created especially for high pressure systems,[11] and the SSSP library which selects the best pseudopotentials from previously existing libraries according to their newly proposed testing protocol.[12]

The motivation of this work is not necessarily to create new libraries, although many of the discoveries could be used towards that end. Instead, the motivation is any case where new pseudopotentials need to be developed, including custom pseudopotentials. Below, we mention some drawbacks of pseudopotential libraries, and how these are addressed with custom pseudopotentials.

It is well known that, for any pseudopotential, a compromise between accuracy and efficiency must be chosen.[13] This places constraints on the design of libraries, where authors must make design decisions that may not be optimal for all users. For instance, in the interest of computational efficiency, authors may choose a large pseudopotential radius, causing sphere overlap in compressed materials or materials with defects. Additionally, a library must aim for transferability to all systems. This leads to cases where small radii and extra semi-core electrons,[14] unnecessary for many systems, introduce added expense to calcula-

tions for some researchers. In contrast, our work does not aim for complete transferability, instead opting to ensure pseudopotentials are optimal for a given application. This relaxes the above-mentioned constraints of developing libraries.

Another disadvantage of libraries is that they may have only have a few pseudopotentials per element, leaving the user in trouble if there are no pseudopotentials available that are appropriate for a given problem. This lack of pseudopotential options can lead to situations where off-the-shelf pseudopotentials are used without regard to their validity in the problem at hand. And for many calculations, the wrong pseudopotential can lead to seemingly valid but drastically inaccurate results.

As an alternative to choosing pseudopotentials from libraries, users can generate their own pseudopotentials using codes such as Atompaw.[15] This approach has the advantage of allowing the user to choose desired combinations of exchange and correlation functionals, core/valence partitions, and other pseudopotential parameters. However, tuning the parameters of these pseudopotentials can be a daunting task for users without years of experience and a detailed understanding of the nuances of pseudopotential development.

Pseudopotentials are not unique, and choosing appropriate parameters for them can be challenging. These parameters are typically adjusted by hand until calculations using the pseudopotential closely match results given by all-electron DFT calculations. The VASP documentation states that "[pseudopotential] generation was, and still is, a tricky, cumbersome, error-prone and time-consuming task" and "few groups can afford to generate a new [pseudopotential] for every problem at hand.[8] Gianozzi, the author of the popular Quantum Espresso DFT code, has published some notes online for hand-tuning pseudopotentials, and he describes the process as "closer to cooking than to science".[16] To date, the majority of the wisdom behind hand-tuning pseudopotentials is passed along through tips in user guides similar to these, and the process of pseudopotential generation relies on intuition and experience. To streamline the pseudopotential generation process, we are developing methods to generate custom pseudopotentials on demand for the user's specific

applications using an automated search for optimal pseudopotentials parameters. This automated optimization will eliminate the need for years of experience or hand-wavy intuition by automatically testing thousands of pseudopotentials.

Although pseudopotentials have been heavily used for decades, this approach of automated optimization is relatively new. In 2015, two separate groups applied a single-objective simplex method to automatically tune pseudopotentials.[10, 17] In our work, we show that pseudopotentials can be reliably tuned using a genetic algorithm to achieve gains in accuracy and speed. Since our original work on generating PAWs using evolutionary algorithms, Sarkar et. al. have applied a genetic algorithm to optimize several pseudopotentials for high pressure materials.[11] A unique advantage to our optimization technique is the generation of a Pareto front of pseudopotentials, which offers a user-selectable tradeoff between accuracy and efficiency. The Pareto front is a result of conducting a multi-objective optimization vs. using only a single objective. In our case, we optimize for both accuracy and speed, and an example of a Pareto front can be seen in Figure 3.3.

The push for reproducibility and validation in DFT has led to the proposal of new metrics and testing protocols for evaluating pseudopotentials. The most notable of the new metrics is the Δ-factor , originally introduced by Lejaeghere et. al. in Reference 18 and subsequently used to compare a wide array of published DFT codes.[19] Other metrics have been proposed by Prandini et. al.,[12] including a test suite for speed and accuracy. They claim the Δ-factor alone is not enough to determine the convergence properties (and thus speed) of a pseudopotential since different metrics can have different convergence properties for the same pseudopotential. Using our optimization tool, we have explored relationships between different pseudopotential quality metrics, deciding which ones are best suited for describing and optimizing pseudopotentials. The new metrics we have considered are an interatomic force-based accuracy metric motivated by situations where interatomic forces are important, and a metric for the atomic scattering property accuracy based on the arctangents of the logderivatives.

4

The previously mentioned arctangent metric addresses problems with the current standard for evaluation of scattering properties in pseudopotentials, the logderivatives. In this work, we show that the arctangent metric reliably represents disagreements in the logderivatives and captures ghost states, meaning it is valuable as a metric for evaluating the accuracy of a pseudopotential. Additionally, since the metric is very cheap to obtain compared to metrics calculated in the solid such as the Δ-factor, the arctangent metric lends itself as a screening metric during pseudopotential development or optimization. By screening out bad PAWs using the arctangent metric, more expensive and time consuming calculations of metrics such as the Δ-factor can be avoided for a large portion of pseudopotentials.

In this dissertation, I describe our optimization techniques using a genetic algorithm, the necessary formulations of the pseudopotential metrics, and the application of the optimizations to real-world problems. More specifically, Chapter 3 covers the application of a multi-objective genetic algorithm to the optimization of PAW pseudopotentials based on interatomic forces, arctangents, and efficiency metrics. Chapter 4 covers the arctangent metric in more detail, with illustrations of its efficacy. Finally, in Chapter 5, we apply our optimizations to the development of a low-valency gallium pseudopotential for defect calculations in GaN, and illustrate how the genetic algorithm improves the speed of optimizations.

Chapter 2

Theory

2.1   Pseudopotentials and Projector Augmented Wave Data Sets

To understand the pseudopotential parameters we adjust during optimizations, some pseudopotential theory is necessary. Several formulations of pseudopotentials have been developed, including norm conserving pseudopotentials,[5] ultrasoft pseudopotentials,[6] and projector augmented wave data sets (PAWs).[7] However, in this work we focus exclusively on PAWs. That being said, our optimization technique is not limited to PAWs, and is applicable to any pseudopotential type. An introduction to PAWs and pseudopotentials in general can be found in Reference 13, and details on the projector augmented wave method can be found in Reference 7. We will discuss only the relevant PAW parameters here.

Valence/core partition In a pseudopotential, the core electrons are frozen as if in an isolated atom, while the valence electrons are free to interact with the solid. The valence partition can also contain semi-core states are necessary if they affect the physics. These semi-core states would typically be considered core electrons, but are significantly involved in bonding for some materials. The semi-core electrons are no longer inert, for example, if their wavefunctions are extended in space (or if the materials under study are highly com-pressed) so that the orbitals are overlapping significantly with neighboring atoms. Including more electrons in the valence partition will typically increase accuracy and transferability of a pseudopotential, but will make calculations more expensive. Therefore, for compu-tational efficiency purposes, it is ideal to keep as many electrons in the core partition as possible.

Radii A core premise of the pseudopotential is that the wave functions and potential must be correct (same as all-electron) outside some sphere surrounding each atom.Inside

6

this sphere of radius r_c, the wave function and potential are smoothed out, or pseudized. This radius corresponds to r_{paw} in Reference 15. r_c is an important tunable parameter of pseudopotentials and is optimized in our work. In the PAW approach, several other radii are defined: r_{shape}, r_{vloc}, and r_{core} (see Reference 15). In the Atompaw code, these three radii are set equal to r_c by default, but can be specified independently if desired. Additionally, each partial wave has an associated radius $r_i \leq r_c$.

Projectors Projectors are used to map the pseudized wave functions to the all-electron partial waves within the PAW method. Each partial wave corresponding to a scattering state has a projector energy, E_i. In practice these projector energies are adjusted by pseudopotential developers to remove ghost states and ensure the logderivatives (and thus scattering properties) match all-electron values over some relevant energy range. By construction, the all-electron and pseudized logderivatives match at the projector energies, so additional projectors with strategically chosen projector energies can be used to enforce scattering accuracy to an arbitrarily high energy. Typically, however, the number of projectors is kept as low as possible to keep computational costs down, and because scattering property accuracy is not normally necessary for high energies. We expect that optimal values of these projector energies depend on the choice of the radii.

2.2   Evaluating Pseudopotentials

The ideal pseudopotential is accurate, efficient (fast), and transferable. These objectives compete with each other, making the pseudopotential generation process an art of compromise. For example, reducing the core radius of a pseudopotential can make the pseudopotential more accurate, but this increase in accuracy usually comes at the expense of computational efficiency; reducing this radius is likely to increase the size of the basis set required to converge calculations. Complete transferability is also never guaranteed because testing a pseudopotential in a variety of systems does not guarantee it will work in others. Better transferability can be gained by including more electrons in the valence or

reducing the core radius, but this again decreases the computational efficiency. There are various metrics for quantifying a pseudopotential's accuracy, efficiency, and transferability. The metrics relevant to this work are discussed in the remainder of this section.

2.2.1 Accuracy

The accuracy of a pseudopotential is gauged by comparison with all-electron calculations as opposed to experimental results. By comparing to all-electron calculations, one can evaluate the pseudopotential approximation only and not other approximations inherent to DFT. A fortunate advantage of this approach is that all-electron results can be obtained in situations where experimental data do not exist. A recent example of this is given in Reference 14, where they use zero-temperature lattice constant data for BCC, FCC, and simple cubic structures. Many of these structures are not stable in nature. Furthermore, calculations do not have to be converged to high accuracy since all that is needed is a comparison with all-electron results. For example, coarse k-point sampling can be used provided this sampling is the same in both the pseudopotential and all-electron calculations. The standard method for evaluating the accuracy of pseudopotentials against all-electron results has two parts, evaluation in the atom and evaluation in the solid. Both of these parts are discussed below, as well as methods for visually or automatically evaluating the quality.

2.2.1.1 Evaluating in the atom

A pseudopotential is evaluated in the atom by comparing pseudized (PS) and all-electron (AE) logderivatives for a single isolated atom. The logderivatives are related to scattering properties of the pseudopotential. Ideally, the scattering properties of the AE and PS potentials should be similar in order to correctly capture the physical interaction of the potential with electrons. When inspecting the log derivatives, any discontinuities should line up as much as possible. Any discontinuity present in the PS logderivatives that is not found in the AE logderivatives is a ghost state, and can result in unphysical behavior in calculations.

8

Typically the logderivatives are inspected visually. In some cases, the root mean square error of AE and PS log derivative curves over some energy range is used.[11] An alternative metric, introduced in this work, is based on the arctangent of the logderivatives. An advantage of the arctangents is that they are numerically smooth whereas the logderivatives contain singularities. This smoothness enables meaningful comparisons between the all-electron and pseudized arctangent curves. This advantage and other advantages of the arctangent metric are the subject of Chapter 4.

The calculation of the logderivatives is very fast (at least with the Atompaw code), requiring only seconds on a single processor. If the logderivatives are sampled very finely (on the order of 10,000 points), the evaluation of the logderivatives can take a few minutes. The arctangents of the logderivatives are then trivially fast to calculate.

2.2.1.2   Evaluating in the solid

Lattice constant/bulk modulus The lattice constant calculation is the classic first pass for evaluating a pseudopotential in the solid. In the simplest case, there is only one lattice parameter and no other structural degrees of freedom (e.g. BCC, FCC, simple cubic, rock salt, diamond, and zinc-blende structures). In this case, the energy vs. lattice constant relationship is simply $E(a)$, or in terms of volume, $E(V) = E(a^3)$. To calculate the equilibrium lattice constant, one must first calculate the total energy at several different lattice constants near some reasonable guess for the minimum. From this energy vs. lattice constant data, an equation of state can be fitted. Typical equations of state are the Murnaghan or Birch-Murnaghan equations of state. The bulk modulus also comes directly from the equation of state and can be used for comparison to all-electron calculations as well.

Δ-factor The Δ-factor was introduced by Lejaeghere et al. in 2014 and is becoming the standard metric for evaluating pseudopotential accuracy in the solid. The Δ-factor is similar to the equilibrium lattice constant calculation in that it requires a calculation of the equation of state across some lattice constant range for both the all-electron and pseudopotential

9

cases. However, instead of just comparing the error in lattice constant and bulk modulus derived from the equation of state, the energy error across the whole lattice constant range is considered. Thus, the Δ-factor is based on the integral of the difference between two equations of state.

$$\Delta = \sqrt{\frac{\int_{0.94V_0}^{1.06V_0}(E_1(V) - E_2(V))^2\,\mathrm{d}V}{0.12V_0}} \qquad (2.1)$$

Several alternative formulations exist,[19, 11] all variations on the same theme. Typically, a small range of volumes, $-0.94V_0$ to $1.06V_0$, near equilibrium is used. In contrast, Sarkar et al. have used a larger energy range to account for high pressure environments such as the Earth's core.

Forces An alternative to equilibrium crystal properties such as the lattice constant is the force-based metric, where interatomic forces are compared between all-electron and pseudopotential calculations in the same atomic structures. The force-based metric is motivated by applications where atoms are not in perfect crystals or molecular dynamics (where accurate forces are crucial).

In order to have non-zero interatomic forces to compare, atoms must be in non-equilibrium positions. There are multiple ways to set up these non-equilibrium structures. One could perturb the atoms from an equilibrium crystal structure. Alternatively, one could randomly place the atoms in the unit cell. Yet another option is to run molecular dynamics simulations at some temperature to get physically relevant atomic configurations. In any case, care must be taken to avoid configurations where atoms are unrealistically close. Ideally, the configurations adequately sample the coordinations with neighboring atoms, but an exhaustive sample is intractable. In practice, only a few configurations may be necessary.

The error in interatomic force for a single atom is defined as $r_f = |\boldsymbol{f}_{\mathrm{AE}} - \boldsymbol{f}_{\mathrm{PS}}|$. For the case of a two atom unit cell, the forces on the two atoms are equal and opposite, and we

can consider the force on only one of the atoms. Then the force based metric is defined as

$$s_f = \left(\frac{1}{N} \sum_{i=1}^{N} r_{f,i}^2 \right)^{1/2}, \tag{2.2}$$

where N is the number of configurations. This metric could easily be extended to account for more atoms in the unit cell.

2.2.2 Efficiency

In addition to accuracy, one wants to know the efficiency, or speed, of a pseudopotential. Pseudopotentials with which calculations are more computationally expensive are referred to as hard pseudopotentials. Soft pseudopotentials are more efficient. Usually the metric for efficiency is just the energy cutoff required for convergence of some physical parameter in the solid. For example, you could determine the required cutoff for a Si potential by converging the lattice constant within some tolerance. Different potentials will converge to that tolerance at different plane wave cutoffs. The idea is to tune parameters of pseudopotentials so that they require the lowest cutoff while still remaining accurate compared to all-electron results. In reality, the expense of calculations comes from the number of self-consistent iterations and bands as well. Therefore, we have used a new efficiency metric which is based not only on the energy cutoff, but also on self-consistent iterations and bands.

2.2.3 Transferability

It is usually desirable for a pseudopotential be accurate in various chemical environments. Early developers of pseudopotentials recognized that the scattering properties of the pseudopotential were critical for encouraging this transferability, hence the focus on logderivatives during pseudopotential evaluation. In addition to scattering property tests, pseudopotential developers typically run calculations in various solids to make sure the

pseudopotential gives correct results in multiple systems (see References 20, 14, 10 for example). Transferability is advantageous for high throughput materials studies because the pseudopotentials are expected to work in all materials. Transferability is also advantageous for reducing the barrier to entry for DFT users.

There is no perfect pseudopotential for every system, thus assuring the transferability of a pseudopotential inevitably leads to sacrifices in speed for many materials. For example, a small PAW radius that leads to the most accurate solution in the materials tested will likely be smaller than necessary for many systems, making the calculations unnecessarily expensive. Another example of transferability leading to sacrifices in speed is the inclusion of the 3d electrons in many studies of GaN. By removing these 3d electrons from the valence of gallium, the simulations of defects in GaN can be sped up by more than 40%, as will be discussed in Chapter 5. The gallium pseudopotential without 3d electrons in the valence is not transferable to many systems, but is perfectly acceptable for defect studies in GaN.

Chapter 3

Force-based optimization of PAW data sets

3.1   Motivation

Pseudopotential methods for density functional theory (DFT) calculations can dras-
tically reduce computational time by reducing calculations involving core electrons, and
have been widely used in computational materials science since their introduction in 1979.[5]
However, pseudopotentials that are both accurate in reproducing all-electron results and ef-
ficient in the computational work required are difficult to generate because these objectives
tend to compete with each other. In this chapter, we generate pseudopotentials based on
their accuracy in predicting atomic forces in non-equilibrium crystal configurations and
their computational work demands. Our force-based approach is different from previous
work, where pseudopotentials have been verified based on equilibrium properties such as
lattice constant and bulk modulus.[9,14]

Motivations for our force-based approach are applications where atoms are not in a
perfect-crystal environment, such as defects in lattices, material interfaces, or molecular
dynamics (MD). In the context of MD, Car-Parinello[21] or Ehrenfest dynamics[22] require
forces obtained from *ab initio* electronic structure calculations using DFT and the pseu-
dopotential approximation, therefore pseudopotentials should be designed to produce ac-
curate forces.

Pseudopotentials are typically tuned by adjusting several parameters until physical
quantities are calculated with sufficiently minimal errors as compared to all-electron re-
sults. Commonly these quantities are equilibrium lattice constant and bulk modulus,[14,9]
but other metrics based on perfect crystals are also used.[18] If a pseudopotential produces
low error for the physical quantity being calculated then the pseudopotential is said to be

good. The question that remains is will the pseudopotential optimized for equilibrium perfect crystals work for atoms that are not in their equilibrium positions? In other words, if we break symmetry by adding a defect or perturbing atoms in the lattice, how can we know that the pseudopotential that was derived for a perfect periodic crystal is still "good?" To answer this question, we have chosen to optimize based on forces in non-perfect crystals to ensure that subsequent force calculations are accurate.

We have used a multi-objective genetic algorithm (MOGA) to optimize pseudopotentials for accuracy and computational work by altering a set of parameters used to create pseudopotentials. The use of genetic algorithms for tuning pseudopotentials was introduced in Reference 23. In that work, though, the genetic algorithm was used to generate accurate, efficient, and transferable pseudopotentials based on equilibrium properties. We are not claiming in this work that pseudopotentials generated using the present approach are transferable. Instead we claim they are applicable to arbitrarily perturbed atomic configurations of the chosen material system. The accuracy of a pseudopotential is partially based on a comparison between forces calculated using the pseudopotential method and forces calculated using the all-electron full-potential linearized augmented-plane wave (FP-LAPW) method. The computational work of a pseudopotential is based on an estimation of the floating-point operations required to run the force calculations. After a specified number of pseudopotentials have been created, the non-dominated pseudopotentials are chosen for the Pareto set. A pseudopotential is non-dominated if it cannot be improved in one objective without worsening the other. The resulting Pareto set contains design points with varying compromises between accuracy and efficiency. Our method could make pseudopotentials more accessible to a broader range of scientists and engineers by automating the optimization of pseudopotentials. The optimizer handles the design, generation, and evaluation of pseudopotentials, resulting in not just one, but many optimized pseudopotentials with varying user selectable compromises between accuracy and efficiency.

In this chapter, we have completed three force-based optimizations: one for lithium and

14

fluorine in LiF, one for silicon and germanium in SiGe, and another for elemental molyb-denum. We chose these materials for two reasons. First, good pseudopotentials are already available to use as a reference when testing the pseudopotentials. Second, these materials are bonded differently, and we show that our method works well for a variety of chemi-cal environments. We have generated a Pareto set of optimized pseudopotentials for each material. To ensure that force-optimized pseudopotentials yield equilibrium properties, we also tested several of our optimized pseudopotentials for their ability to predict equilib-rium lattice constant and bulk modulus. Moreover, we test the optimized pseudopotentials in atomic configurations unknown to the optimizer. This step ensures that the optimized pseudopotentials are generally applicable to any atomic configuration and not just valid for the optimized configurations. Finally, results using our optimized pseudopotentials were compared with results given by pseudopotentials from the `ABINIT` library.

3.2   Methods

3.2.1   Pseudopotential optimization

A multi-objective genetic algorithm (MOGA), implemented using the DAKOTA soft-ware package,[24] was used for optimizing pseudopotentials. The two objectives for the optimization were accuracy and computational work, which MOGA tries to minimize (the accuracy objective should be minimized because it is based on force residuals). The genetic algorithm (GA), which is based on biological evolution, starts with a random population of design points, where each design point is a unique set of values for the design variables. The design variables are inputs to `atompaw` used to create pseudopotentials (one pseudopoten-tial for each element in the material), and the goal of MOGA is to find the best values for these design variables. For LiF and SiGe, which have two elements, each design point cre-ates a pair of pseudopotentials, so the elements are being optimized simultaneously. The design variables used for each material system are shown in Table 3.1. Each design point is

tested to determine its work and accuracy objectives, where lower objective values are desirable. Design points with lower objectives are given higher fitness values. Design points from each generation with low fitness values are discarded, and the remaining design points are subjected to genetic operations (mutation and crossover), which introduce diversity, to create the next generation of design points. We used MOGA's random shuffle crossover operator, which randomly selects design variable values from a specified number of parents to create the offspring. The specified number of parents was 10, the specified number of offspring was 20, and the crossover rate was 0.8. Mutation was carried out using uniform replacement, where every design variable for the design point is changed to a random value distributed uniformly in the allowed range of that design variable. The mutation rate was 0.3.

As MOGA progresses, the optimizer should find design points with higher fitness values. Because accuracy and work are competing objectives, there is no single best design point. Instead, all design points that are non-dominated are chosen for the Pareto set, which is the final product of MOGA (see Figure 3.3, for example). A non-dominated design point cannot be decreased in one objective without increasing the other objective. For this study, design points in the Pareto set that produce less error will require more work, and those that produce more error will require less work. In addition to searching for design points with higher fitness, the optimizer also attempts to fill in gaps in the Pareto front.

To determine the error given by the pseudopotential method, we compared atomic forces calculated using pseudopotentials to atomic forces calculated using the all-electron full-potential linearized augmented-plane wave (FP-LAPW) method. Both the pseudopotential method and the all-electron method share the approximations of DFT, so any differences in results between the two methods can be attributed to the pseudopotential used. The all-electron method was considered exact for the purposes of this work.

In addition to comparing forces, we also used the atomic wave functions to effectively screen pseudopotentials containing ghost states. Traditionally, ghost states are detected by

16
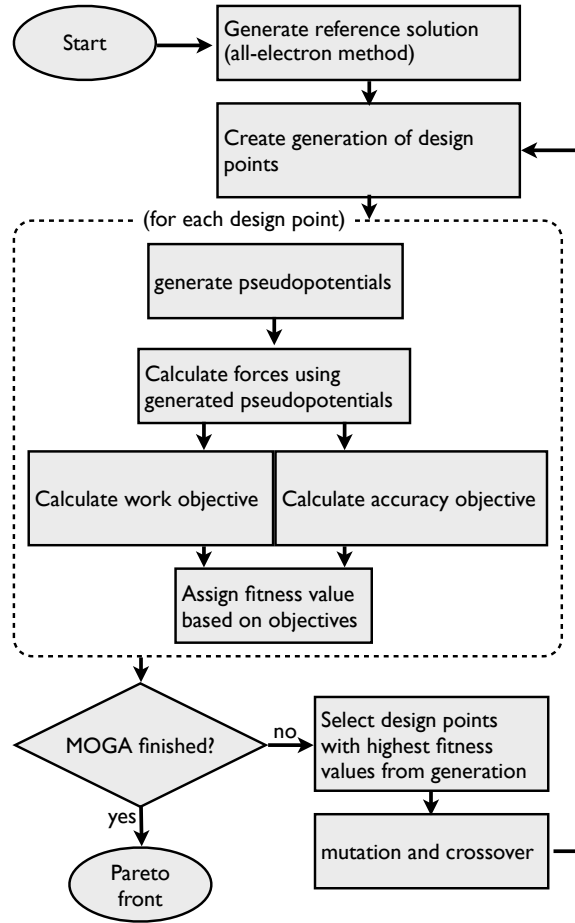
Figure 3.1: MOGA workflow. For each design point, a pair of pseudopotentials is created and tested simultaneously.

Material	Design Variable	Minimum	Maximum
LiF	RC_{Li}	0.2 Bohr	4 Bohr
	RC_F	0.2 Bohr	4 Bohr
	EP_F	0 Ry	15 Ry
SiGe	RC_{Si}	0.2 Bohr	4 Bohr
	EP_{Si}	0 Ry	15 Ry
	RC_{Ge}	0.2 Bohr	4 Bohr
	ED_{Ge}	0 Ry	15 Ry
Mo	RC_{Mo}	0.2 Bohr	4 Bohr
	EP_{Mo}	0 Ry	15 Ry
	ED_{Mo}	0 Ry	15 Ry

Table 3.1: Design variables and specified ranges used for MOGA. `RC` stands for cutoff radius. `EP` and `ED` are reference energies for P and D orbitals.

comparing the logarithmic derivatives of the pseudopotential and all-electron radial wave functions, defined as

$$\frac{1}{\Psi_{\text{El}}(r_c)} \frac{d\Psi_{\text{El}}}{dr}, \tag{3.1}$$

where Ψ_{El} is the wave function, r is the radial position, and r_c is the cutoff radius. Visual inspection of the logarithmic derivatives of the two wave functions allows pseudopotential designers to quickly assess the quality of a pseudopotential before it's put into a crystal. This otherwise manual operation can be codified for the purposes of optimization by looking at the root mean square (RMS) of the logarithmic derivative residuals. However, directly minimizing this RMS is a poor choice for automated evaluation due to discontinuities in the logarithmic derivatives. Any slight misalignment between the all-electron and pseudo wave functions leads to a large change in the residual sum. Even worse, as the misalignment decreases, the residual sum steadily increases until the wave functions significantly overlap, at which point the residual sum starts decreasing. As a result, trying to minimize this term using traditional gradient based techniques can lead to solutions where the all-election and pseudo wave functions have no overlap, resulting in a useless solution. Only if the initial pseudopotential parameters lead to a starting guess where the all-election and pseudo wave functions overlap sufficiently to be in the local minima or region of attraction will the optimization lead to a potentially useful solution.

Alternatively one can treat the logarithmic derivative as

$$\tan \Theta = \frac{1}{\Psi_{\text{El}}(r_c)} \frac{d\Psi_{\text{El}}}{dr} \tag{3.2}$$

and use RMS of the all-election and pseudo wave function Θ residuals. This approach doesn't suffer from having a region of repulsion surrounding a logarithmic discontinuity misalignment. In fact, as the original logarithmic derivatives align themselves, the Θ residuals decrease. Each traversal of a discontinuity coincides with a π phase shift which accumulates over the plot. Consequently, extra discontinuities, including those that are barely

visible due to the discrete energy sampling, have a dramatic effect on the resulting objective and hence the optimization. This makes using the RMS of the Θ residuals an excellent choice for evaluating the pseudopotential quality from an automated point of view.

The projector augmented wave (PAW) pseudopotential method[7] was implemented using the `socorro` code from Sandia National Laboratories.[25] The all-electron method was implemented with the `Elk` code.[26] The exchange and correlation functionals used the PBE version of the generalized gradient approximation (GGA).[27] For `Elk`, the exchange and correlation functionals were implemented using `LibXC`.[28] For `socorro` and `atompaw` the built in implementations of the exchange and correlation functionals were used. A $4 \times 4 \times 4$ Monkhorst-Pack k-point grid[29] was used for all calculations. This k-point mesh may not have been sufficient for convergence to physical values, but since both the all-electron and pseudopotential methods use the same k-point grid the results can be compared without convergence being required.[14]

Atoms in the crystals studied were perturbed from the equilibrium positions to produce net forces on the atoms. The equilibrium configuration is the natural crystal structure of each system; rock-salt for LiF, zinc-blende for SiGe, and BCC for Mo. The atomic forces were calculated for four different non-equilibrium random atomic configurations, which we call training configurations. After testing optimizations with two, four, and eight training sets, we concluded that four was sufficient for our systems because we did not observe significant changes in results of the optimizations when using more than four training configurations. Nevertheless, we concede that more training sets could be necessary for more complex systems. The generation of the random configurations used for the optimizations is discussed in more detail in Section 3.2.2.

We used the force residual to quantify force error between pseudopotential forces and all-electron forces. The force residual is the magnitude of the vector difference between the two forces on an atom:

$$r_f = \left| \mathbf{f}_\mathrm{p} - \mathbf{f}_\mathrm{AE} \right|, \tag{3.3}$$

where $\mathbf{f}_p$ is the force on the atom from the pseudopotential method and $\mathbf{f}_{AE}$ is the force on the atom from the all-electron method. A lower force residual represents a better match with the all-electron results and more accurate force. If the pseudopotential and all-electron methods give equal force vectors, the force residual is 0. For all sytems we defined the unit cell with two atoms, meaning the atomic forces will be equal and opposite. Therefore, we consider the forces on atom 2 only. The Θ residual is defined as

$$r_\Theta = \Theta_p - \Theta_{AE}, \tag{3.4}$$

where Θ_p is the arctangent of the logderivative for the pseudized wave function and Θ_{AE} is the arctangent of the logderivative for the all-electron wave function.

The accuracy objective for a design point is constructed as

$$S_A = s_f + W_\Theta s_\Theta, \tag{3.5}$$

where s_f is the sum of the force residuals over the atomic configurations (N),

$$s_f = \left(\frac{1}{N} \sum_{i=1}^{N} r_{f,i}^2 \right)^{1/2}. \tag{3.6}$$

And s_Θ is a sum of the arctan of logderivatives over each element and each state in each element (l),

$$s_\Theta = \frac{2}{\pi} \frac{1}{L} \sum_{\text{elements}} \sum_{l} \left(\frac{1}{M} \sum_{i=1}^{M} (r_{\Theta,i})^2 \right)^{1/2}, \tag{3.7}$$

where M is the number of sample points. The weight W_Θ in equation 3.5 is chosen so that both terms in S_A are similar orders of magnitude. The weights used for this chapter were 0.005 for LiF, 0.022 for SiGe, and 5.3 for Mo.

The work objective was an approximation of the number of floating point operations required to calculate the kinetic energy, potential, electron density, projected eigen-problem,

Hamiltonian, overlap operator, and residuals. The number of floating point operations is approximated as

$$S_W = m \left[2n_w + 10b^2 + b^3 + (b+1)\, n_d \log_2 n_d + b \left(3n_w + 4pn_w + 2n_w \log_2 n_w \right) \right] \quad (3.8)$$

where b is the number of bands, n_w is the size of the FFT grid for wave functions and potential, n_d is the size of the density FFT grid, p is the number of projectors, and m is the number of iterations. We chose to estimate computational work using this formula instead of wall time to avoid biases due to different computing environments.

Different pseudopotential pairs required different wave function cutoff (E_{cut}) values for the solution to converge. To determine the required wave function cutoff, we increased it from 10 Ry in steps of 10 Ry and considered the solution converged when

$$\left(\frac{E_{i-1} - E_i}{E_{i-1}} \right)^2 < 10^{-8}, \quad (3.9)$$

where E_i is the cell energy for the wave function cutoff being tested and E_{i-1} is the cell energy for the previous wave function cutoff tested. The density cutoff was set to $4E_{cut}$. Pseudopotential pairs that didn't converge by $E_{cut} = 70\,\text{Ry}$ were given artificially high work and accuracy objectives so the optimizer would recognize them as bad design points.

Pseudopotentials were generated using the `atompaw` code.[15] The core and valence configurations used are shown in Table 3.2. We chose to vary cutoff radii (`RC`) and projector energies (`EP` and `ED`) as design variables for these optimizations. Table 3.1 shows the `atompaw` input parameters that were used as design variables for MOGA. The PAW pseudopotentials generated by `atompaw` have four specified cutoff radii: r_{paw}, r_{shape}, r_{vloc}, and r_{core}.[15] For simplicity, we used a single value for all four radii. If `atompaw` could not create a pseudopotential given a set of parameters, an artificially high objective was returned to the optimizer to identify it as poor design point. An artificially high objective was also returned if pseudopotentials took longer than a specified wall time to converge or the

Element	core	valence
Li	none	1s, 2s, 2p
F	1s	2s, 2p
Si	1s, 2s, 2p	3s, 3p
Ge	1s, 2s, 2p	3s, 4s, 3p, 4p, 3d
Mo	1s, 2s, 3s, 2p, 3p, 3d	4s, 5s, 4p, 4d

Table 3.2: Specified valence configurations for generating pseudopotentials with `atompaw`.

pseudopotential resulted in non-convergence of a calculation.

3.2.2 Random atomic configurations

Atoms in the crystal structure must be moved from the equilibrium position to produce a net force on the atoms. The unit cells for LiF and SiGe have lattice vectors $(\frac{a}{2}, \frac{a}{2}, 0)$, $(0, \frac{a}{2}, \frac{a}{2})$, and $(\frac{a}{2}, 0, \frac{a}{2})$ where a is the corresponding lattice constant. The unit cell for Mo has lattice vectors $(a, 0, 0)$, $(0, a, 0)$, and $(0, 0, a)$. We generated atomic configurations for the optimization by randomly placing atom 2 in the unit cell while leaving atom 1 fixed at the unit cell origin.

To characterize the random atomic configurations, we have used the normalized nearest neighbor distance. This value is proportional to the distance between the closest two atoms in the crystal lattice for a given atomic configuration. It arises because the perturbation of an atom from its equilibrium position will move the atom closer to other atoms in the lattice (although in the zinc-blende structure, perturbations sometimes increase the nearest neighbor distance). The closest two atoms in the crystal are not necessarily in the same unit cell. For each configuration, the normalized nearest neighbor distance is defined as

$$d = \frac{h}{h_e} \tag{3.10}$$

where h is the (non-normalized) nearest neighbor distance for that configuration and h_e is the nearest neighbor distance for the equilibrium crystal structure. Normalizing to h_e aids

22

in analysis of atomic distances by letting $d = 1$ for equilibrium and $d < 1$ when atoms move closer. For LiF (rock-salt structure), $h_e = a/2 = 3.804\,\text{Bohr}$. For SiGe (zinc blende structure), $h_e = \sqrt{3}a/4 = 4.531\,\text{Bohr}$. For Mo (BCC), $h_e = \sqrt{3}a/2 = 5.157\,\text{Bohr}$. The lattice constants used for the force calculations only need to be approximate as long they are the same for both the the all-electron and pseudopotential calculations, ensuring a valid comparison between forces. The lattice constants used were $a = 7.6\,\text{Bohr}$ for LiF, $a = 10.5\,\text{Bohr}$ for SiGe, and $a = 6.0\,\text{Bohr}$ for Mo.

A random atomic configuration was not used if $d < 0.8$. This constraint prevents atoms from being too close and producing unrealistically large repulsive forces. Because many atomic positions violate this constraint, atom 2 is only allowed in a certain region in the unit cell. To illustrate this, we generated 100,000 random configurations and plotted the atomic locations in the unit cell (Figure 3.2). Atom 1 can be seen at the unit cell origin. The positions of atom 2 represent the region of allowed positions for atom 2. Table 3.3 contains the four random configurations generated for each crystal system that were used for the optimization.

3.2.3 Evaluation of optimized pseudopotentials

Ideally, pseudopotentials generated to calculate forces accurately would also predict equilibrium properties with similar accuracy. We calculated equilibrium lattice constant and bulk modulus using several design points from the Pareto fronts. The results using our optimized pseudopotentials were compared with results given by the all-electron method. To find lattice constant and bulk modulus, the energy of the unit cell was calculated for nine different lattice constants between 95% and 105% of the lattice constant initial guess. The energy vs. unit cell volume relationship was fit to the Murnaghan equation of state[30] and the lattice constant and bulk modulus were derived from this fit. A calculation of lattice constant was done with `Elk` using the lattice constants reported in Section 3.2.2 as the initial guess. The resulting lattice constants from those calculations were used as the

(a) LiF

(b) SiGe

(b) Mo
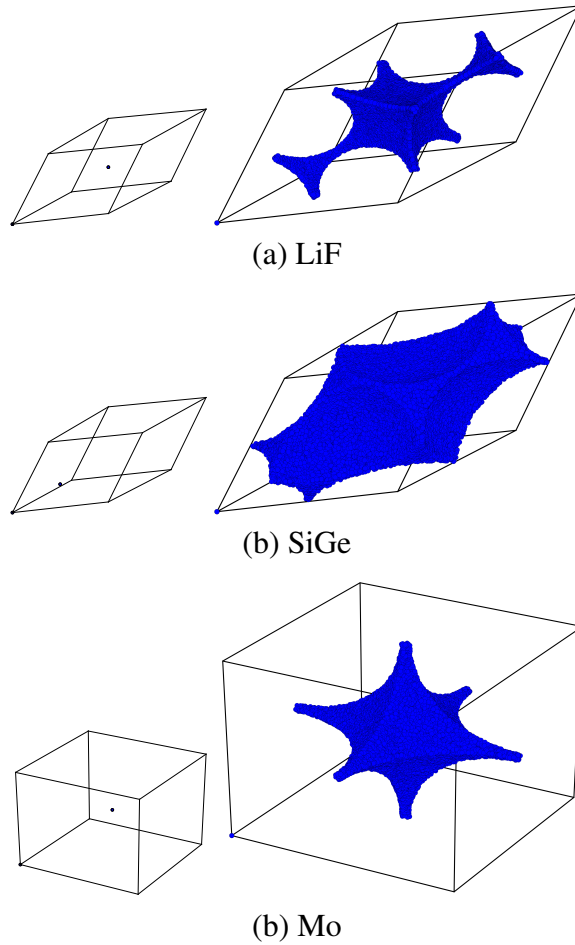
Figure 3.2: Left: Equilibrium positions of the two unit cell atoms for LiF in the rock-salt structure, SiGe in the zinc blende structure, and Mo in the BCC structure. Right: Illustrations of the region of possible positions of atom 2 for LiF and SiGe. Atom 1 is fixed at the origin. These images were created by generating 100000 random configurations using the method in Section 3.2.2.

Configuration	atom 2 coordinates	d
LiF 1	0.282, 0.187, 0.242	0.824
LiF 2	0.546, 0.314, 0.607	0.862
LiF 3	0.387, 0.496, 0.469	0.865
LiF 4	0.658, 0.478, 0.387	0.877
SiGe 1	0.740, 0.240, 0.448	0.825
SiGe 2	0.497, 0.507, 0.222	0.887
SiGe 3	0.591, 0.492, 0.329	0.958
SiGe 4	0.372, 0.506, 0.475	0.988
Mo 1	0.514, 0.050, 0.496	0.804
Mo 2	0.760, 0.431, 0.509	0.804
Mo 3	0.508, 0.142, 0.522	0.809
Mo 4	0.514, 0.431, 0.721	0.817

Table 3.3: Atomic configurations for LiF and SiGe used for this chapter. The training configurations were used for the optimization of pseudopotentials. The lattice vectors for LiF and SiGe are $(\frac{a}{2}, \frac{a}{2}, 0)$, $(0, \frac{a}{2}, \frac{a}{2})$, and $(\frac{a}{2}, 0, \frac{a}{2})$. The lattice vectors for Mo are $(a, 0, 0)$, $(0, a, 0)$, and $(0, 0, a)$. Atom 1 position is (0,0,0). The training configurations for each material are sorted by d.

guesses for the all-electron and pseudopotential lattice constant calculations reported in this chapter.

3.3 Results

3.3.1 Pareto fronts

For all crystal systems, MOGA created and tested a set number of design points—2500 for LiF and Mo and 7500 for SiGe. The initial generation size for each material was 300 design points. Subsequent generation sizes were determined by MOGA, and new generations were created until the specified number of design points were tested. At the end of the optimization, the non-dominated design points were identified for the Pareto front. The resulting objectives for each design point tested are shown in Figure 3.3 and points on the Pareto front are identified in blue. The best design points have low error and work objectives, so they are found towards the bottom and left in the plots. The Pareto front offers a diverse set of design points. In Figure 3.3, the design points vary from accurate but

expensive (top left) to inexpensive but not as accurate (bottom right) with a distribution of points in between.

For comparison, we calculated the work and accuracy objectives of pseudopotentials available from a pseudopotential library. Lithium, fluorine, silicon, germanium, and molybdenum PAW pseudopotentials are available from the `ABINIT` website.[31] These pseudopotentials, which we used as reference pseudopotentials, are expected to be very accurate. Figure 3.3 shows the objectives of these reference pseudopotentials in relation to our Pareto front. We were able to generate pseudopotentials with comparable accuracy and efficiency compared to reference pseudopotentials from the literature, as is evident from the accuracy and work objectives of the reference pseudopotentials in Figure 3.3.

We chose three design points from each of our Pareto fronts to test further. These representative design points are labeled *accurate*, *balanced*, and *fast* based on their positions on the Pareto front. Figure 3.3 shows the locations on the Pareto fronts of these design points. For LiF and SiGe, the design point on the Pareto front with the best accuracy objective was chosen for *accurate*. For Mo, the second Pareto point was chosen for *accurate* instead of the first. Mo Pareto point 1's work objective is 160% higher than Pareto point 2's work objective while offering only a 3% improvement in accuracy objective, making it an unlikely choice for production runs. For all materials, the design point with the lowest work objective was chosen for *fast*. We chose a design point near the middles of our Pareto fronts for the *balanced* pseudopotentials to represent a balance between accuracy and work.

3.3.2   Genetic algorithm vs. random sampling and parametric sweep

Genetic algorithms (GA) are known to be advantageous for complex design spaces because they are less likely to be caught in a local minimum. However, we performed tests of the GA compared to brute force techniques to determine whether the GA (1) explored the entire design space, and (2) found the superior Pareto set. For the first test we randomly sampled the design space with 7500 points—the same number of total design points used

26
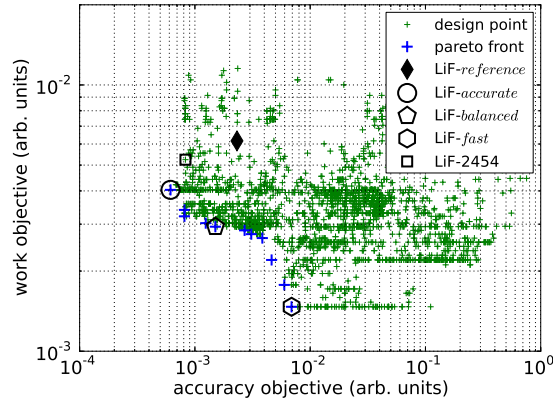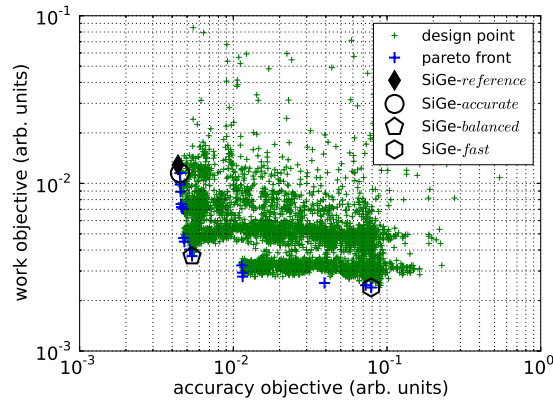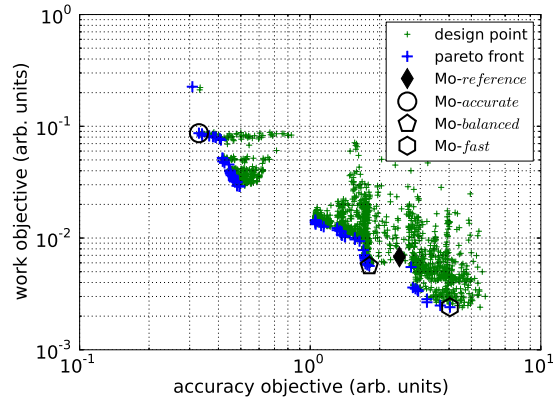
(a) LiF

(b) SiGe

(c) Mo

Figure 3.3: Design points and Pareto fronts for (a) LiF, (b) SiGe, and (c) Mo. More accurate design points have lower accuracy objectives and are towards the left in the plots. Faster design points have lower work objectives and are towards the bottom in the plots. We calculated the work and accuracy objectives of pseudopotentials from the literature; the objectives of these reference pseudopotentials are indicated on the plots. We chose three design points at different locations on the Pareto front to further evaluate. The chosen design points, *accurate*, *balanced*, and *fast*, are also indicated on the plots.

27
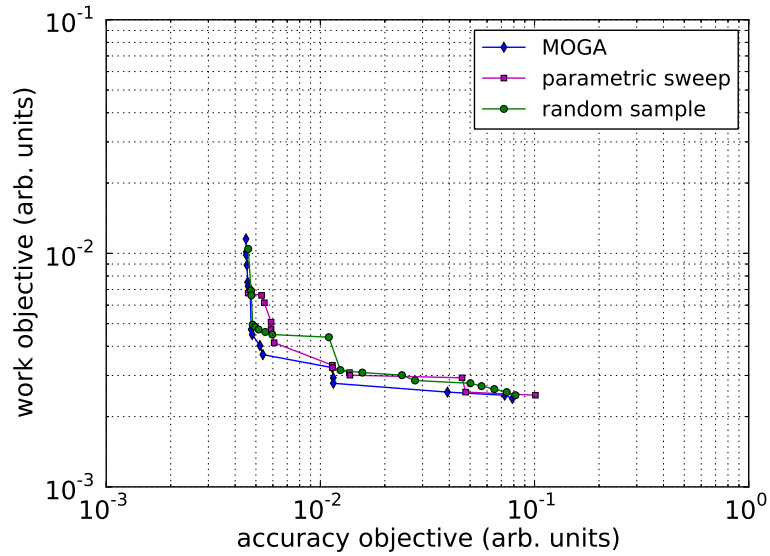
Figure 3.4: Pareto fronts for three different optimizations–one using the GA, one using a random sample of the design space, and one using a parametric sweep of the design space. The Pareto fronts are similar but the GA pushes the Pareto front farther towards lower objectives.

in the GA optimization. For the second, we gridded the 4D design space with 10 points in each of the the two cutoff radius dimensions and 9 points in each of the two projector energy dimensions, for a total of 8100 design points. We used 8100 design points instead of 7500 because we needed a product of 4 approximately equal integers to form the uniform grid. We avoided other optimization techniques because the majority of common optimization techniques can optimize only a single objective. Therefore, comparison to our study, which relies heavily on the quality of a Pareto optimal set with two competing objectives, is not meaningful.

The difference between an optimized Pareto front (using MOGA) and one created from random sampling or a parametric sweep were not as dramatic as we expected. Nevertheless, the GA does seem to push the Pareto front farther towards lower objectives as seen in Figure 3.4.

Interestingly, the GA approaches its final Pareto front in the first few generations, im-
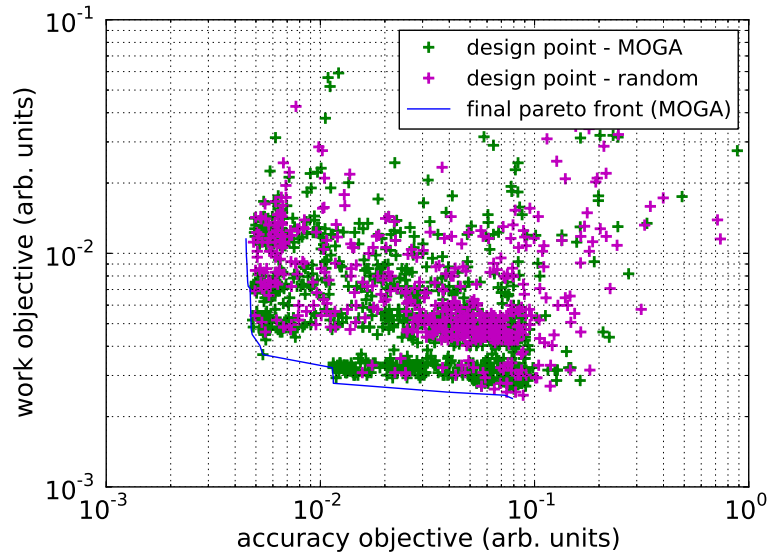
Figure 3.5: Design points after 10 generations (900 design points) of the GA alongside 900 randomly sampled design points. The final Pareto front from the GA after 7500 design points is shown as a solid line.

plying that the GA could be stopped much sooner and still give comparable results. Figure 3.5 shows the design points after 10 generations (900 design points) of the GA alongside 900 randomly sampled design points. The final Pareto front from the GA after 7500 design points is shown as a solid line. Clearly the design points are near optimal at this point in the optimization since they are near the final Pareto front. Notice the GA design points have approached the corner of the Pareto front, which is arguably the most important region.

3.3.3   Accuracy of forces in training configurations

The force magnitudes for the training configurations are plotted in Figure 3.6. Each plot shows the force magnitude on atom 2 in the unit cell calculated with our representative pseudopotentials and reference pseudopotentials for LiF, SiGe, and Mo. In general, the *reference*, *accurate*, and *fast* pseudopotentials all produce forces reasonably close to the all-electron solution while the *fast* pseudopotential forces are off by a larger amount. The *fast*

pseudopotentials for all materials are probably unusable due to high force errors. However, other points on the Pareto front offer fast solutions without the excessive accuracy penalty.

Trends in the results are noticeable in Figure 3.6a for configurations 2-4 and Figure 3.6c for configurations 2-4. For LiF, *reference* consistently under-predicts the forces. LiF-*accurate* is very near the all-electron solution with LiF-*balanced* and LiF-*fast* producing higher errors as expected. Similarly for SiGe, SiGe-*reference* consistently underpredicts the forces. SiGe-*accurate* and SiGe-*balanced* are close to the all-electron solution while SiGe-*fast* both overpredicts and underpredicts by large amounts. For Mo, all pseudopotentials tend to over-predict forces except for training configuration 1.

3.3.4 Accuracy of forces in random test configurations

We calculated forces for 100 random atomic configurations that were generated after the optimization was complete. These test configurations were generated with the same $d > 0.8$ constraint on positions that was used when generating the training configurations. Because these random configurations were not used for the optimization, they can be used to show how well different pseudopotentials predict forces in configurations unknown to the optimizer.

When generating the training configurations, we allowed d to be any value greater than 0.8. However, the configurations were generated randomly and because of the small sample size, d is in the range 0.824-0.877 for LiF, 0.825-0.988 for SiGe, and 0.804-0.817 for Mo. Table 3.3 lists d for all training configurations, and these optimization ranges are shaded in Figure 3.7. Because the test configurations represent a larger sample, the range of d for the test configurations is wider, and many configurations fall outside the range of d used in the optimization.

Figure 3.7 shows the percent error in the forces for each test configuration, calculated as

$$r_f^* = 100 \times \frac{r_f}{|\mathbf{f}_{\mathrm{AE}}|} \tag{3.11}$$
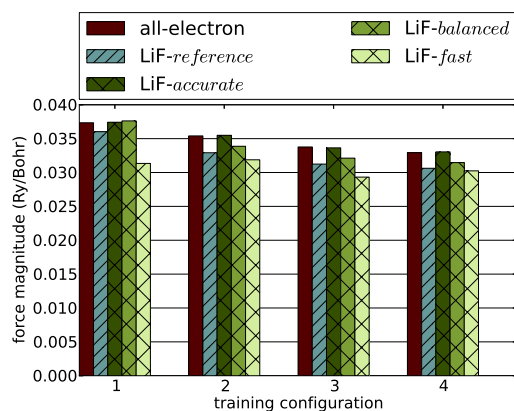
(a) LiF
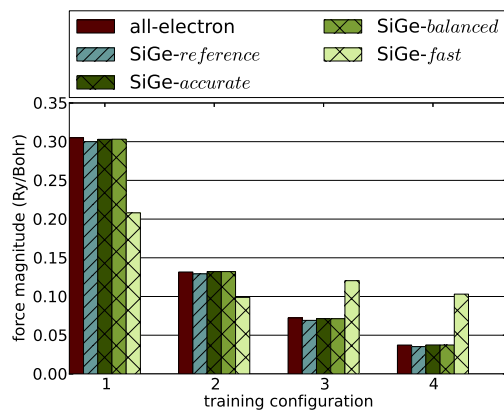
(b) SiGe
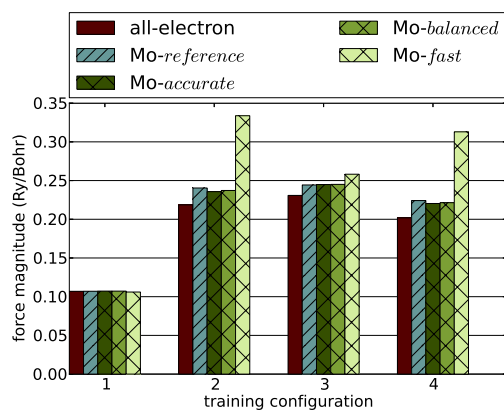
(c) Mo

Figure 3.6: Magnitude of force on atom 2 for (a) LiF, (b) SiGe, and (c) Mo in the training configurations as generated by the all-electron method, the reference pseudopotentials, and the representative pseudopotentials.
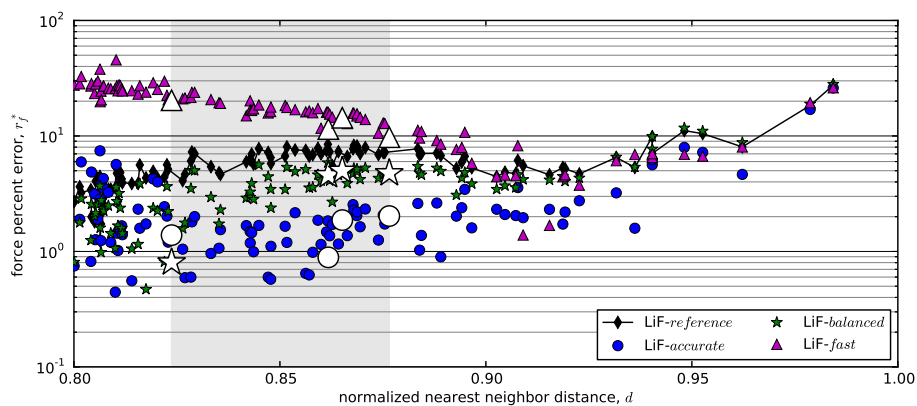
(a) LiF
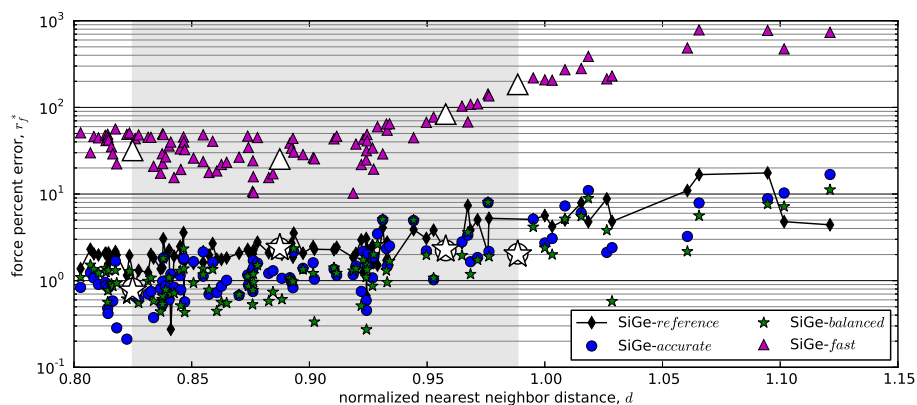
(b) SiGe
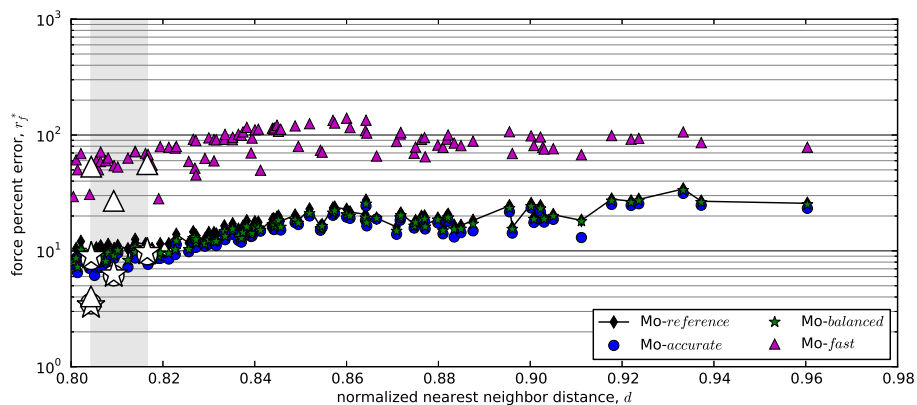
(c) Mo

Figure 3.7: Normalized force residuals for (a) LiF, (b) SiGe, and (c) Mo in their test configurations for the reference and representative pseudopotentials. The optimization region is indicated by the shaded area. Lower force residuals indicate more accurate force results. The hollow points represent the training configurations. The lines between points are for visualization and are not meant to imply a certain trend.

where $\mathbf{f}_{AE}$ is the force on the atom from the all-electron method. r_f is the force residual defined in section 3.2.1. Recall, lower force residuals means better accuracy, with $r_f^* = 0$ being the exact solution. These plots show how well our pseudopotentials calculate forces both inside and outside the optimization ranges. The *fast* pseudopotentials had by far the largest force residuals for every configuration. With the exception of the Mo training configuration 1 (leftmost training config in Fig. 3.7c), the residuals for the test configurations are consistent with the residuals for the training configurations. This implies that force residuals for a few training configurations can represent the residuals for a wide range of configurations. In other words, optimizing for a few training configurations will produce pseudopotentials with predictable accuracy in configurations unknown to the optimizer. For LiF, the *accurate* and *balanced* pseudopotentials consistently produced lower force residuals than the reference pseudopotential within the optimization region and for *d* greater than the optimization region. For the highest values of *d*, LiF-*balanced* produced slightly higher force residuals. For *d* lower than the optimization region, the residuals for both LiF-*accurate* and LiF-*balanced* increased for many configurations and were higher than the reference pseudopotential. The increase in residuals for these smaller *d* may result from increased sphere overlap that the optimizer did not anticipate. SiGe-*accurate* and SiGe-*balanced* produced lower residuals than the reference pseudopotentials across the entire *d* range, with only a few configurations having higher residuals. Compared to differences in force residuals for the other systems, Mo-*accurate* and Mo-*balanced* produced practically the same force residuals as the reference pseudopotentials, although Mo-*reference* did tend to have slightly higher residuals.

3.3.5 Accuracy of lattice constant and bulk modulus

The pseudopotentials optimized using forces were also evaluated for their ability to estimate equilibrium lattice constant and bulk modulus. The results were compared with the lattice constant and bulk modulus calculated using both the reference pseudopotentials

	design variables			objectives				
	RC_{Li}	RC_F	EP_F	S_A	S_W	a (%)	B (%)	E_{cut} (Ry)
ref	1.61	1.40	2.00	0.00232	0.00616	−0.22	+2.67	50
		1.20						
		1.40						
		1.20						
2454	2.07	1.37	2.13	0.00083	0.00524	−0.11	+5.00	50
1	2.06	2.12	4.05	0.00061	0.00404	+0.14	+4.66	30
2	1.62	2.16	0.55	0.00081	0.00339	+0.05	+2.80	30
3	1.62	2.16	1.06	0.00081	0.00321	+0.04	+2.97	30
4	1.62	2.12	4.64	0.00124	0.00303	+0.28	+2.70	30
5	1.88	2.16	5.37	0.00150	0.00294	+0.36	+1.47	30
6	1.88	2.27	1.75	0.00270	0.00284	−0.03	+2.71	30
7	1.88	2.03	4.26	0.00308	0.00275	−0.24	+2.27	30
8	1.89	2.27	0.62	0.00385	0.00266	−0.06	+2.07	30
9	1.82	3.08	9.68	0.00463	0.00220	+7.41	−36.89	30
10	3.95	3.73	3.49	0.00596	0.00177	+12.15	−21.75	20
11	3.88	3.39	14.37	0.00690	0.00147	+16.86	−64.14	20

Table 3.4: Design variables, accuracy objective, work objective, percent errors in lattice constant and bulk modulus, and converged E_{cut} for LiF reference pseudopotentials, LiF design point 2454, and all points on the Pareto front. LiF-*accurate*, LiF-*balanced*, and LiF-*fast* are shaded. The RCs specified for the reference pseudopotentials are r_{paw}, r_{shape}, r_{vloc}, and r_{core}.[1] If only one value is specified for RC, then all four cutoff radii are equal. The percent errors are relative to $a = 7.702$ Bohr and $B = 64.59$ GPa, which are derived from the all-electron calculation.

and the all-electron method. The percent errors in lattice constant and bulk modulus for LiF, SiGe, and Mo are reported in Tables 3.4, 3.5, and 3.6. With the exception of the few points near the bottom of each Pareto front, the optimized pseudopotentials produced lattice constants within 1% of the all-electron lattice constant and bulk modulus values within 5% of the all-electron bulk modulus.

3.3.6 Design variables

Because LiF has only three design variables—RC_{Li}, RC_F, and EP_F—its design points can be plotted in a 3D parameter space. Figure 3.8 shows the work objective, represented by color, as a function of the design variables. This figure (as well as Figure 3.9) is meant to be

	design variables				objectives				
	RC_{Si}	EP_{Si}	RC_{Ge}	ED_{Ge}	S_A	S_W	a (%)	B (%)	E_{cut} (Ry)
ref	1.90	4.50	2.30	3.00	0.00436	0.01291	-0.08	$+0.58$	50
	1.60		2.00						
	1.75		1.80						
	1.20		1.90						
1	2.65	8.19	1.92	4.08	0.00450	0.01152	$+0.34$	-2.02	40
2	2.65	7.90	1.92	4.08	0.00451	0.01014	$+0.33$	-1.94	40
3	2.65	7.86	1.92	4.08	0.00451	0.00983	$+0.33$	-1.93	40
4	2.74	8.69	1.84	4.67	0.00454	0.00891	$+0.38$	-2.64	40
5	2.65	6.91	1.91	4.08	0.00458	0.00753	$+0.29$	-1.64	40
6	2.65	6.91	1.92	3.41	0.00460	0.00722	$+0.21$	-1.21	40
7	2.48	7.75	1.84	4.67	0.00470	0.00707	$+0.22$	-1.29	40
8	2.46	4.73	2.11	1.89	0.00475	0.00472	$+0.34$	-1.48	30
9	2.34	2.61	2.12	1.89	0.00480	0.00449	$+0.35$	-1.52	30
10	2.34	5.40	2.19	1.14	0.00522	0.00403	$+0.33$	-1.62	30
11	2.09	4.49	2.19	0.51	0.00539	0.00368	$+0.27$	-1.25	30
12	1.82	12.60	2.72	0.23	0.01137	0.00324	$+1.54$	-9.75	20
13	1.95	11.46	2.72	0.23	0.01149	0.00293	$+1.52$	-9.60	20
14	1.95	5.88	2.72	0.23	0.01149	0.00278	$+1.53$	-9.61	20
15	2.65	6.91	3.04	0.23	0.03920	0.00255	$+2.99$	-20.13	20
16	2.65	5.32	2.78	7.02	0.07254	0.00247	$+9.49$	-27.78	20
17	3.53	9.94	3.12	5.31	0.07879	0.00239	$+15.95$	-40.74	20

Table 3.5: Design variables, accuracy objective, work objective, percent errors in lattice constant and bulk modulus, and converged E_{cut} for SiGe reference pseudopotentials and all points on the Pareto front. SiGe-*accurate*, SiGe-*balanced*, and SiGe-*fast* are shaded. The RCs specified for the reference pseudopotentials are r_{paw}, r_{shape}, r_{vloc}, and r_{core}.[1] If only one value is specified for RC, then all four cutoff radii are equal. The percent errors are relative to $a = 10.59$ Bohr and $B = 72.48$ GPa, which are derived from the all-electron calculation.

	design variables			objectives				
	RC_{Mo}	EP_{Mo}	ED_{Mo}	S_A	S_W	a (%)	B (%)	E_{cut} (Ry)
ref	2.20	3.00	4.00	2.437	0.0068	+0.09	+2.88	30
1	1.35	4.89	11.14	0.308	0.2260	+0.08	+2.14	40
2	1.38	14.09	13.37	0.329	0.0868	+0.08	+2.29	40
4	1.38	12.48	13.37	0.340	0.0825	+0.08	+2.28	40
7	1.38	1.56	13.37	0.387	0.0803	+0.09	+2.20	40
10	1.38	0.17	13.86	0.410	0.0753	+0.09	+2.19	40
13	1.42	14.90	13.41	0.417	0.0486	+0.07	+2.43	40
16	1.44	13.80	13.37	0.447	0.0402	+0.07	+2.51	40
19	1.44	5.09	13.37	0.451	0.0385	+0.08	+2.42	40
22	1.45	14.20	13.37	0.462	0.0361	+0.07	+2.58	40
25	1.45	12.71	11.14	0.464	0.0351	+0.07	+2.55	40
28	1.45	0.20	14.51	0.465	0.0329	+0.08	+2.46	40
31	1.46	6.41	13.37	0.481	0.0315	+0.07	+2.55	40
34	1.46	0.26	7.58	0.494	0.0290	+0.08	+2.51	40
37	1.60	3.36	11.52	1.053	0.0136	+0.06	+3.04	40
40	1.67	6.41	11.14	1.144	0.0126	+0.06	+3.09	40
43	1.77	0.39	9.05	1.345	0.0116	+0.07	+2.96	40
46	1.78	0.53	4.96	1.372	0.0106	+0.08	+2.93	40
49	2.03	14.24	1.89	1.561	0.0101	−0.17	+4.86	40
52	2.03	9.70	3.70	1.701	0.0078	+0.10	+2.22	30
55	2.00	0.35	4.96	1.725	0.0068	+0.09	+2.96	30
58	2.00	0.53	4.06	1.734	0.0063	+0.09	+2.94	30
61	2.04	0.17	2.50	1.802	0.0056	+0.09	+2.98	30
64	2.96	0.35	1.16	2.857	0.0035	+0.73	+1.55	20
67	3.56	0.85	5.87	3.206	0.0028	+12.69	+24.25	20
70	3.82	0.17	14.51	4.036	0.0024	+14.16	−5.99	20

Table 3.6: Design variables, accuracy objective, work objective, percent errors in lattice constant and bulk modulus, and converged E_{cut} for the Mo reference pseudopotential and some points on the Pareto front. Mo's front contained 70 points because of the shape of the front, so all Pareto points are not shown in this table. Mo-*accurate*, Mo-*balanced*, and Mo-*fast* are shaded. The percent errors are relative to $a = 5.990$ Bohr and $B = 256.5$ GPa, which are derived from the all-electron calculation.
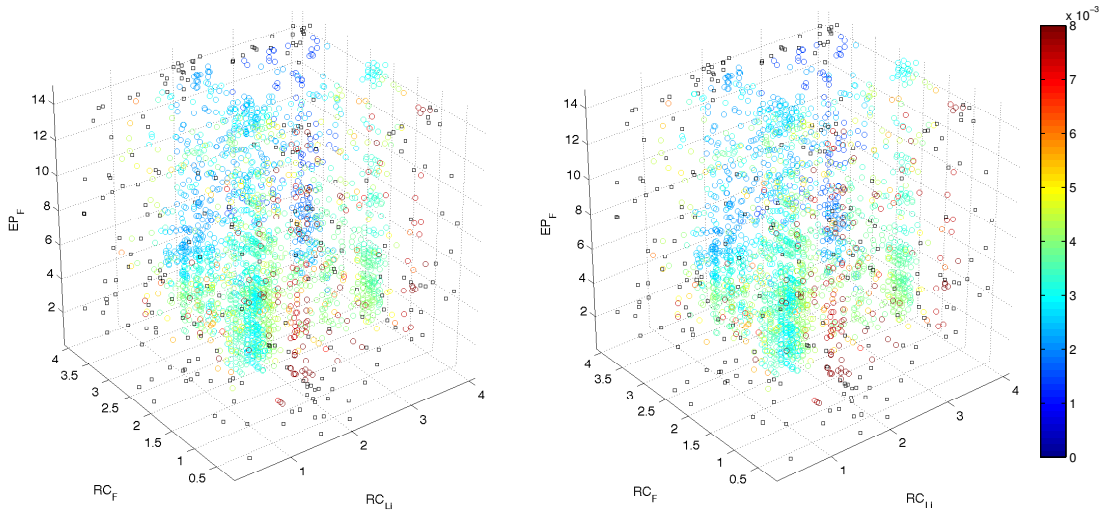
Figure 3.8: Every design point tested for LiF plotted by RC_{Li}, RC_F, and EP_F with color representing the work objective S_W (arbitrary units). This figure is a cross-view stereoscopic image that can be viewed in 3D. The black squares represent design points that returned artificially high objectives. The lack of significant color variation in the vertical direction indicates that EP_F has little effect on the work objective.

cross-viewed in 3D and helps illustrate the distribution of points sampled by the optimizer. Any dependence of the work objective on EP_F is visible as a variation of color in the vertical direction. In Figure 3.8, little variation of color appears in the vertical direction, meaning the work objective has only weak dependence on EP_F. Indeed, Tables 3.4, 3.5, and 3.6 show no obvious trends in projector energies for any of the systems. Similarly, Figure 3.9 shows only slight dependence on EP_F. Although the dependence is weak, in some regions of the design space, EP_F is used to refine the pseudopotential and is consequently kept as a design variable. Nevertheless, because of the weak dependence on projector energies, we focus the following discussion on cutoff radii.

We can compare the design variable values for our representative design points with the reference pseudopotentials from the literature (Tables 3.4, 3.5, and 3.6). The `atompaw` inputs for the reference pseudopotentials are different from our input files in several ways other than RCs and projector energies, so identical RCs and projector energies would not produce identical pseudopotentials. Therefore, the comparison of RCs and projector energies is
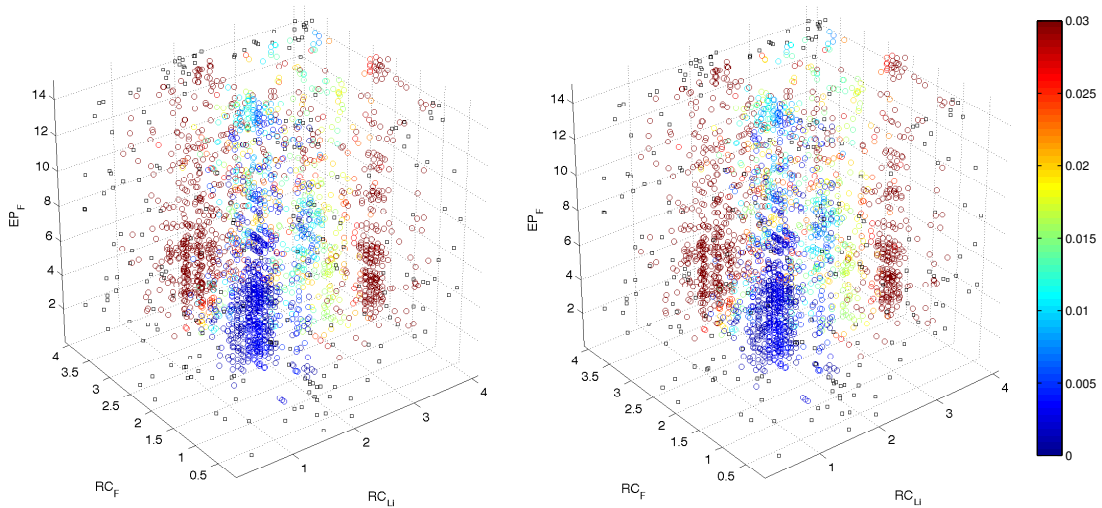
Figure 3.9: Every design point tested for LiF plotted by RC_{Li}, RC_F, and EP_F with color representing the accuracy objective S_A (arbitrary units). This figure is a cross-view stereoscopic image that can be viewed in 3D. The black squares represent design points that returned artificially high objectives. The lack of significant color variation in the vertical direction indicates that EP_F has little effect on the accuracy objective.

not a direct comparison. However, the RCs play an important role in pseudopotential properties and should be examined. Upon a cursory inspection, decreasing work corresponds to larger RCs, which is consistent with common understanding. For LiF, all of the RCs for our representative design points were greater than the reference pseudopotentials, which used $RC_{Li} = 1.6\,\text{Bohr}$ and $RC_F = 1.4\,\text{Bohr}$. Figure 3.8 shows that the region for $RC_{Li} = 1.6\,\text{Bohr}$ and $RC_F = 1.4\,\text{Bohr}$ is a region of high work objective, probably explaining why all of our representative design points have higher RCs, which would fall in regions of lower work objective.

For LiF, the lattice constant and bulk modulus results are noticeably worse after Pareto point 8, which coincides with a large jump in cutoff radii. It is well known that PAW sphere overlap can yield poor and unpredictable results. This motivates an examination of the cutoff radii for the more accurate points on the Pareto front. Pareto point 1 has large cutoff radii, much larger than LiF-*reference* and large enough to cause sphere overlap. In fact, most design points tested caused sphere overlap. Initially, the possibility of sphere

overlap did not concern us because we assumed cutoff radii that were large enough to cause overlap would produce bad force results and thus be ruled out by MOGA. Evidently, this was not the case since the overlap did not cause bad results for our Pareto points. This suggests we could improve our optimizations by constraining the cutoff radii or using more training configurations. Using more training configurations would increase the chances of overlap resulting in design points with smaller cutoff radii.

3.3.7   Larger unit cell

We tested our pseudopotentials in larger unit cells to understand the scalability of our results to larger systems. For LiF and SiGe, we used the respective rock-salt and zinc-blende conventional unit cells, which contain 8 atoms. For Mo, we used a $2 \times 2 \times 2$ conventional BCC unit cell, containing 16 atoms total. We perturbed the atoms from their equilibrium positions by a random amount in each direction. The maximum allowed perturbations were 1.9 Bohr for LiF, 2.1 Bohr for SiGe, and 1.2 Bohr for Mo. The configurations were constrained to $d > 0.85$ to produce configurations in the optimization ranges for LiF and SiGe. Random configurations were generated using the allowed perturbations until a configuration was found satisfying this constraint. We used the same $d > 0.85$ constraint for Mo even though $d = 0.85$ is not in optimization range. According to Fig. 3.7c, the Mo pseudopotentials perform similarly across the entire range of d so the choice of d for Mo should not matter. Having already shown that the *fast* pseudopotentials are too inaccurate to be useful, we only show results for *accurate* and *balanced* pseudopotentials in this section. Figure 3.10 shows r_f^* for each atom in the larger unit cells for LiF, SiGe, and Mo.

For LiF, our optimized pseudopotentials produced much higher force residuals than LiF-*reference* in the large unit cell, especially for the fluorine atoms. This result was unexpected since our pseudopotentials produced better residuals in the test configurations. We think this result is due to the PAW sphere overlap discussed in section 3.3.6, despite the interatomic spacings being in the same range as in Figure 3.7. For LiF, the large cut-
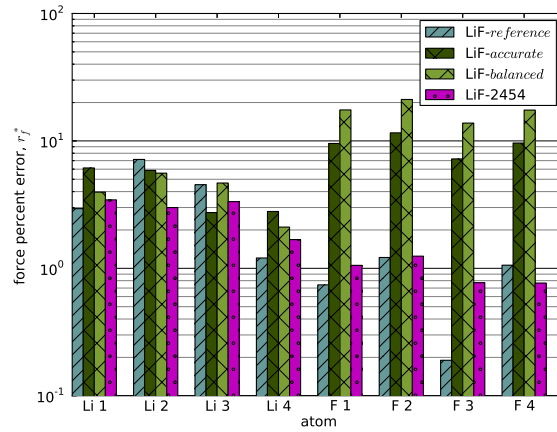
(a) LiF
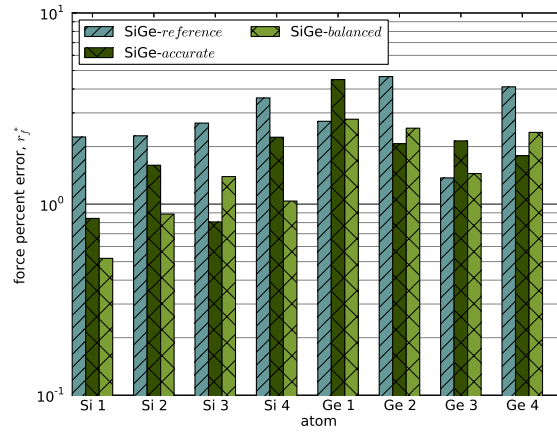
(b) SiGe
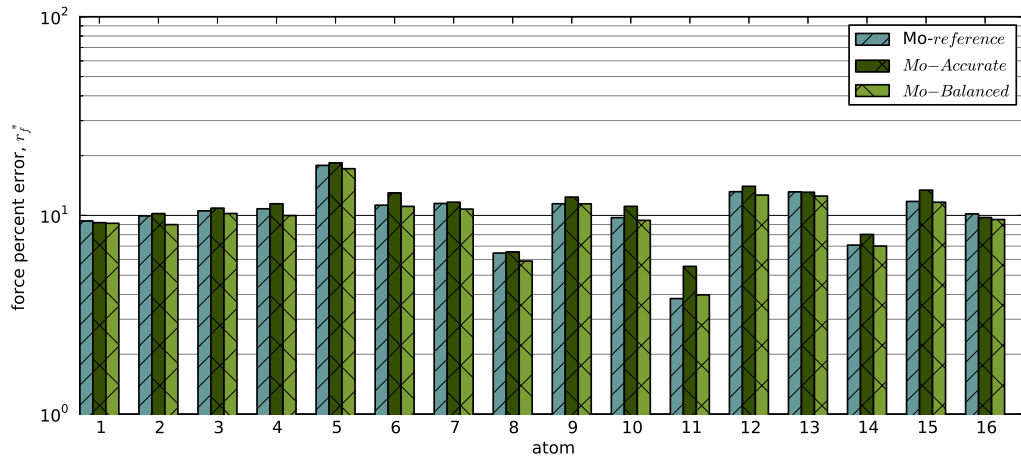
(c) Mo

Figure 3.10: Force percent error, r_f^*, for each atom in the large unit cell for (a) LiF, (b) SiGe, and (c) Mo as calculated by the reference pseudopotentials and various optimized pseudopotentials.

off radii did not hurt our force results for the training configurations, or the design points would have produced poor force results and been given worse accuracy objectives. Also, the sphere overlap was not an issue for the test configurations, as shown in Figure 3.7. Perhaps this disparity in results is because our random perturbations did not sample Li-Li interactions or F-F interactions, instead only varying distances and angles between neighboring atoms of Li and F. This implies improvements could be made in the method by sampling interactions for like atoms by using a larger unit cell.

We chose an additional LiF design point from the optimization to test, design point 2454. This point had smaller cutoff radii than any of the Pareto points and a higher work objective. It is indicated in Figure 3.3. If this point had a slightly lower accuracy objective, it would have been on the Pareto front. Figure 3.10a shows that this design point produces similar residuals to the the reference pseudopotential. We expect that if the Li-Li and F-F interactions had been sampled more, smaller cutoff radii would have been necessary to produce accurate forces and the optimizer would have selected points with smaller cutoff radii.

For SiGe, the lack of Si-Si sampling or Ge-Ge sampling did not seem to degrade the force results when using the larger unit cell. The force residuals for our optimized pseudopotentials were mostly lower than reference pseudopotentials, as expected from the results in Figure 3.7. Perhaps it was acceptable not to sample Si-Si and Ge-Ge interactions in this case because silicon and germanium are chemically similar elements, making an Si-Si or Ge-Ge interaction similar to a Si-Ge interaction. The low residuals produced by SiGe-*accurate* and SiGe-*reference* in the low *d* range in Figure 3.7 also imply that SiGe is not sensitive to the sphere overlap in this case when calculating forces. SiGe-*balanced* had low residuals despite being farther down the Pareto front. This is due to the shape of the Pareto front (Figure 3.3b). The left section of the Pareto front is nearly vertical, meaning all of the points on the left edge of the front have very similar accuracies. This feature of the Pareto front is advantageous because one can decrease the work significantly while maintaining

high accuracy, as shown by the low force residuals produced by SiGe-*balanced*.

For Mo, the trends in force residuals for the larger unit cell are comparable to trends in force residuals for the test configurations, with residuals being similar between the different pseudopotentials. Our hypothesis—that the lack of same-element sampling in LiF caused poor results for larger LiF unit cell—is supported by the Mo findings because the residuals for this plot are the same as in Figure 3.7.

3.3.8   SiGe in cesium chloride structure

For our pseudopotentials to be useful, they must have some degree of local transferability, meaning they should work in materials related to the optimized material even if they aren't meant to be universally transferable to all materials. To examine the local transferability of our optimized pseudopotentials, we tested our SiGe pseudopotentials in the cesium chloride structure. The CsCl structure is chemically different from the optimized structure because it uses a cubic primitive cell instead of a rhombohedron, meaning the atoms are coordinated differently. The unit cell for the CsCl structure has lattice vectors (a, 0, 0), (0, a, 0), and (0, 0, a) where $a = 6.14$ Bohr for CsCl structured SiGe. A new set of test configurations was generated for the CsCl structure with the constraint $d > 0.8$ where $h_e = \sqrt{3}a/2$. The procedure for testing the pseudopotentials in the CsCl structure is the same as in section 3.3.4. We tested the pseudopotentials in 100 random test configurations by comparing forces calculated with pseudopotentials to forces calculated with all-electron DFT.

As expected, SiGe-*accurate* and SiGe-*balanced* produced reasonable force values for the 100 test configurations (Figure 3.11). The optimization range in this figure is the same as in Figure 3.7(b), but plotted using absolute nearest-neighbor distances instead of normalized nearest neighbor distances. For larger h, both SiGe-*accurate* and SiGe-*balanced* produced higher errors than SiGe-*reference*, but still within reasonable values. These high h configurations had values of h outside the optimization range, probably explaining the
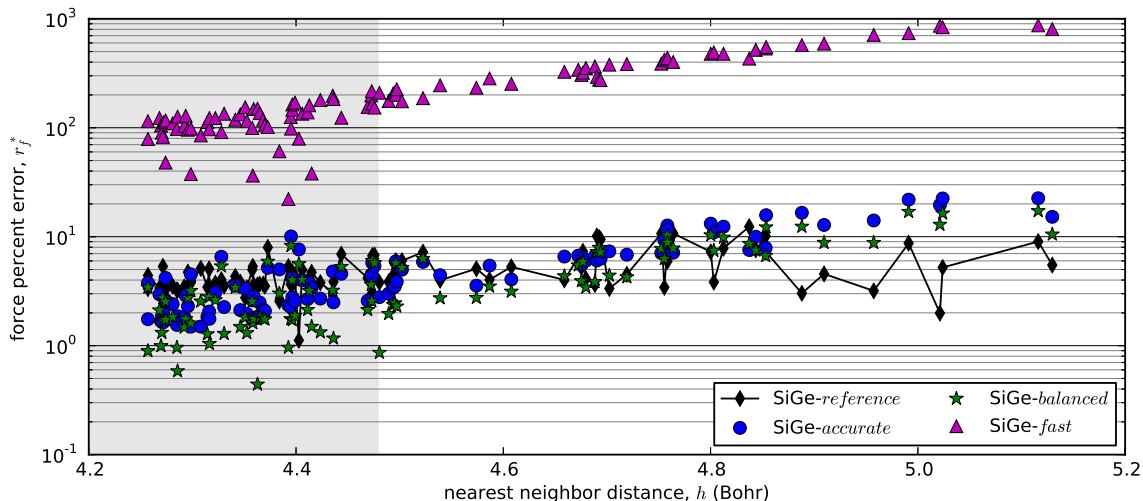
Figure 3.11: Normalized force residuals in CsCl structured SiGe test configurations for the reference and representative pseudopotentials. The shaded area represents the optimization region. Lower force residuals indicate more accurate force results. The lines between points are for visualization and are not meant to imply a certain trend.

increase in force errors in that region. The higher force residuals outside the optimization region suggest that these pseudopotentials could be further improved with a better sample of configuration space using larger unit cells or various unit cells. In other words, an intelligent selection of configurations could improve the optimization compared to a random selection of configurations.

3.3.9 Computational work

For LiF, all of the Pareto points have lower work objectives than the reference pseudopotentials, meaning the Pareto points should all be faster. The slowest LiF Pareto point had a work objective 34% lower than LiF-*reference*, and the fastest had a work objective 76% lower. LiF design point 2454, which was discussed in Section 3.3.7, has a work objective 15% lower than LiF-*reference*. The results for LiF design point 2454 alone justify using optimization to develop pseudopotentials to achieve speedup with arguably comparable accuracy. This point is further supported by results for SiGe, where all Pareto

points again had a lower work objective than SiGe-*reference*. SiGe-*accurate* is 11% lower, SiGe-*balanced* is 71% lower, and SiGe-*fast* is 81% lower. However, these results are not universal.

In contrast to LiF and SiGe, most Mo Pareto points had higher work objectives than Mo-*reference*, with 54 Pareto points being higher, one being equal, and the remaining 15 being lower. Mo-*accurate* had a work objective 33 times higher than Mo-*reference*, while Mo-*balanced* was 17% lower and Mo-*fast* was 64% lower. To understand why we didn't get the same favorable results for Mo as we did for LiF and SiGe, we consider the crystal environments.

For Mo, the optimizer created a pseudopotential, Mo-*balanced*, that is practically identical to Mo-*reference* in terms of accuracy and work. Even though both pseudopotentials produced very similar results, they are not equivalent pseudopotentials. In fact, the design variables in Table 3.6 indicate that these are very different pseudopotentials. Nevertheless, the performance between these two pseudopotentials for the test configurations, large unit cell, and lattice constant/bulk modulus is similar. We believe this similarity in the performance of the pseudopotentials is a result of optimizing Mo in the same target structure (BCC) as Mo-*reference*.[9] Considering both pseudopotentials were optimized for the same crystal structure, it is not surprising that the optimized pseudopotential is comparable to the reference pseudopotential. The optimization for the same crystal structure may explain why the results from the Mo pseudopotential optimization do not show any significant gains in speed.

In contrast, for Si and Ge the target structure of the optimization was different than those of the reference Si and Ge pseudopotentials, which were separately optimized as elemental crystals. We were able to optimize Si and Ge together to produce speed gains that optimizing Si and Ge individually may not have produced. This result also supports our idea that generating pseudopotentials on demand for specific systems can allow speedup without sacrificing accuracy. Perhaps large scale simulations can be sped up dramatically

while maintaining a high level of accuracy simply by tuning the pseudopotentials for the environment.

Chapter 4

Atan metric

4.1   Motivation

The pseudopotential and projector augmented wave (PAW)[7] methods,[*] which are core frameworks of plane wave density functional theory (DFT) codes such as VASP,[32,33,34,35] Abinit,[36] and Quantum Espresso,[37] enable accurate and tractable calculations in periodic and semi-periodic systems. The choice of pseudopotentials for a study can affect the accuracy and speed of calculations, so a key step in using these methods is choosing or generating pseudopotentials that are appropriate for the problem at hand. Widespread use of plane wave DFT in both physics and materials science means the quality of available pseudopotentials affects the accuracy and throughput of a large variety of research.

Especially important for the accuracy and transferability of pseudopotentials are their scattering properties.[5,13] However, the current approach for evaluating scattering properties, which involves visual inspection of logderivative plots,[1,16] is subjective and requires time consuming human inspection. Therefore, this approach does not allow for generation of new pseudopotentials using automated optimizations, which are the subject of several recent efforts in pseudopotential development.[10,23,38,11] Additionally, the current push for standardized pseudopotential verification[19,12] is still lacking a scattering property metric. In this work, we suggest a new metric, S_A, that addresses shortcomings of visual inspection of the logderivatives, and provides a low-barrier but meaningful metric for standardized evaluation and reporting of the quality of pseudopotential scattering properties.

The new metric S_A is formulated by taking the arctangent of the logderivatives. This

---

[*]Although pseudopotentials and PAW data sets (PAWs) are distinct solutions to modeling the core electrons and atomic charges in DFT, we can discuss both under the same framework. As such, we will refer to both as pseudopotentials except when specified otherwise.

extension of the canonical approach to evaluating PAWs using logderivatives alone provides valuable information to the user or optimizer analyzing the PAW. We show that this metric is capable of robustly predicting ghost states without resorting to density of states or equation of state (EOS) calculations in the solid, and accurately represents the amount of disagreement between all-electron and pseudized solutions. Furthermore, we show that S_A predicts accuracy in the solid in many cases, meaning it is valuable as a screening metric during optimizations.

In the past, the process of generating pseudopotentials was somewhat of an art requiring extensive hand tuning[1,16] by long-time experts in the field. More recently, the need for custom pseudopotentials, as well as the desire for new libraries, has sparked interest in automated optimization of pseudopotentials.[10,23,38,11] Both hand tuning and automated optimization can benefit from improved pseudopotential quality metrics. For hand tuning, meaningful and reliable metrics become a tool for more efficiently evaluating pseudopotentials as they are created, allowing faster throughput. From there, optimizations are a natural next step, where loosening the burden of tedious visual inspection of the logderivatives allows the search of much larger parameter spaces in a more rigorous way. For automated optimizations, reliable and meaningful metrics are critical for guiding the optimization.

Before discussing the details of the proposed metric S_A, we summarize the standard metrics currently in use by the community. These metrics fall into two categories: atom metrics (S_A will fit in this category) and solid metrics. Atom metrics involve comparisons between calculations of the isolated atom using an all-electron solution and a pseudized solution. The standard atom metrics are related to scattering properties of the atom. The quality of the scattering properties are typically evaluated visually using a plot of the logderivatives, where the agreement between pseudized (PS) and all-electron (AE) logderivatives indicates quality. Small discrepancies at higher energies can be tolerated. However, singularities in the PS wave function that do not exist in the AE wave function are known as ghost states and result in non-physical solutions. Because of the small system size (one

atom) and because much of data necessary for evaluation of the pseudopotential atom metrics is a by-product of pseudopotential generation, atom metrics are computationally cheap to obtain.

Solid metrics, in contrast, involve calculations in elemental solids, compounds, or both. These metrics are typically derived from energy vs. volume calculations. From these data, equations of state (EOS), equilibrium lattice parameters, and bulk moduli can be obtained and then compared to the same parameters from all-electron calculations. More recently, the Δ-factor,[18,19] which represents the error in the EOS with a single number, has gained popularity. Regardless which solid metric is preferred, all are expensive compared to atom metrics due to setup time, larger system sizes, multiple electronic structure calculations, and compulsory convergence studies. Furthermore, codes for all-electron solutions of periodic systems (using the full-potential linearized augmented plane wave (FP-LAPW) method, for example) can have their own uncertainties that can be on the order of errors associated with pseudopotentials.[19] This uncertainty, combined with inherent differences between all-electron and pseudopotential methods, means that a small bias exists any time pseudopotential results are compared to all-electron calculations in the solid. This bias, however, does not exist with atom metrics. Nevertheless, the advantage of solid metrics is that one obtains evidence that the pseudopotential gives reasonable results in a calculation of a material (as opposed to an isolated atom). Moreover, one can infer the transferability of the pseudopotential between different systems by sampling various materials structures.

In many cases, researchers have relied on expensive solid metrics to detect ghost states. A recent (2018) example is given in Reference 12, where they analyze the accuracy of band energies given by using the pseudopotentials in various solids. In this work, which does not generate new pseudopotentials but instead analyzes a handful of previously existing pseudopotential libraries, the expense of the band structure calculations is not a limiting factor. Additionally, the band structure calculations were already necessary for estimation of the required plane wave energy cutoff for different pseudopotentials. In contrast, we are

trying to facilitate automated optimization efforts, where thousands, or even orders of magnitude more pseudopotentials need to be evaluated. For this many pseudopotentials, band structure calculations and other solid metrics become prohibitively expensive for detecting ghost states. Evidently, an atom metric that could reliably identify a pseudopotential that is accurate in the solid would greatly reduce the barrier to generating new pseudopotentials.

The arctangent approach, which yields an atom metric, to evaluating pseudopotentials is advantageous over logderivatives because the arctangent method produces numerically smooth results, making numerical comparisons more meaningful. Additionally, it can identify ghost states that would otherwise go undetected. Finally, the quality of a pseudopotential can be reduced to and reliably indicated by a single number, which is useful for automated evaluations.

The arctangent of the logderivatives has been used in previous studies for visual comparison of scattering properties. In his chapter on pseudopotentials, Don Hamann describes the advantage of plotting the arctangent vs. the logderivatives because the arctangent does not have singularities.[39] Hamann has also used this metric in Reference 40, citing the advantage of the smoother curves for visual comparison. Mattsson has used the arctangent curves for visual inspection during customization of VASP Li and Mo PAWs for high energy density physics applications.[41,42] While studying xenon at high pressures, Root et al. found ghost states in a VASP Xe PAW, which VASP then improved.[43] The scattering properties of the old and new Xe PAWs are compared visually with arctangent plots.[44]

We note that the computation of the arctangent must be phase unwrapped to use the metric for automated pseudopotential evaluations. The phase unwrapping prevents discontinuities in the arctangent. However, phase unwrapping is not always the norm. A non-phase-unwrapped version of the arctangent metric has been used by Kiffner et al.[45] for visual evaluation of pseudopotentials. A non-phase-unwrapped version is also included in the Atompaw EXPLORE capability. In this case, the utility of the arctangent as an automated metric is questionable.

4.2 The arctangent metric

The logderivatives of the wave functions of a pseudopotential defined for each angular momentum channel l as

$$L_l(E) = \left. \frac{\Psi'_l(E,r)}{\Psi_l(E,r)} \right|_{r=r_c}, \tag{4.1}$$

where $\Psi(E,r)$ is the electron wavefunction at a given energy, E, and r_c is the cutoff radius of the PAW. These values are related to the atomic scattering properties and are important indicators of accuracy and transferability of pseudopotentials. Logderivatives are calculated for both the all-electron (AE) and pseudized (PS) wave functions, and the two are compared visually. The logderivative vs. energy curves often contain one or more singularities, which are related to nodes in the wave functions. The locations of these singularities should match as closely as possible between AE and PS logderivatives. However, numerically representing the agreement between these singularities in logderivative curves is challenging because the infinities do not behave well numerically. In especially bad cases, singularities in the PS log derivatives exist that are not present at all in the AE log derivatives. These spurious singularities are known as ghost states, and their existence can cause significant errors in the solid calculation. Therefore, pseudopotentials with ghost states are unusable. The arctangent is better than the logderivatives at indicating ghost states for a variety of reasons that will be illustrated below.

We first consider the logderivative metric as an evaluation tool to illustrate its shortcomings. Because we are comparing two curves that we expect to be the same—the all electron (AE) and pseudo (PS) logderivatives—it makes sense to subtract one from the other to get a residual error. Because both of these curves are evaluated on a discrete sample of points, the difference gives an error value at each sample point. Then the RMS of these errors represents the disagreement between the curves as a single number:

$$S_{L,l} = \sqrt{\frac{1}{N} \sum_{E=E_1}^{E_2} \left(L_{l,\mathrm{PS}}(E) - L_{l,\mathrm{AE}}(E) \right)^2}, \tag{4.2}$$

where L_{AE} is the all-electron logderivative, L_{PS} is the pseudo logderivative, and N is the number of sample points. The fundamental problem with this metric is that the disagreement between the curves is not faithfully reflected in the metric, meaning worse agreement does not always equal a larger value, and a larger value does not always mean worse agreement. This feature is illustrated in the following toy examples.

Because the logderivative metric can be extremely sensitive to the alignment of logderivative singularities with the discrete sample points. We consider a mock logderivative of the form

$$L_{\text{mock}}(E) = \frac{1}{E-a}. \tag{4.3}$$

The $1/E$ form is similar to the shape of a singularity in an actual logderivative. The mock AE logderivative is represented by equation 4.3 with a displacement of $a = 0$, and is shown by the solid line in Figure 4.1a. The mock PS logderivative is also represented by equation 4.3 but with the displacement a as a free parameter. By sweeping a, various disagreements between the AE and PS logderivatives are represented and shown in Figure. 4.1 as the dashed lines, where larger a represents greater disagreement. To quantify this disagreement (pretending we don't know that the mock AE and PS logderivatives are the same shape), we attempt to apply the logderivative metric defined in equation 4.2 with $L_{l,\text{AE}}(E) = L_{\text{mock}}(E)\big|_{a=0}$ and $L_{l,\text{PS}}(E) = L_{\text{mock}}(E)\big|_{a}$. The value of the metric is now dependent on the displacement a. What we expect from a meaningful metric is that it increases monotonically as the magnitude of the disagreement a increases. However, what we see instead is the relationship shown in Figure 4.1b, where the metric varies erratically. This behavior is due to the sensitivity of the logderivative metric to the alignment of energy sample points with the singularity and is clearly undesirable.

However, another issue with the logderivative metric that is being masked by this singularity alignment issue can can be illustrated by smoothing the mock logderivative at the

singularity. Consider an alternative mock log derivative of the form

$$L_{\text{smooth}}(E) = \frac{1}{E-a}\left(1 - e^{-100(E-a)^2}\right),$$ (4.4)

where the singularity at $E-a$ is softened by the factor of $1 - e^{-100(E-a)^2}$. As before, the mock AE logderivative is represented by $a = 0$, and the mock PS logderivatives by varying a. These AE and PS mock logderivatives are shown in Figure 4.2a. Again, we apply the metric in equation 4.2 as an attempt to quantify the disagreement, with $L_{l,\text{AE}}(E) = L_{\text{smooth}}(E)\big|_{a=0}$ and $L_{l,\text{PS}}(E) = L_{\text{smooth}}(E)\big|_{a}$. For a meaningful metric, we would expect a monotonic increase of the metric as the magnitude of the disagreement a increases. Instead, we get the relationship shown in Figure 4.2b, where the metric does not increase monotonically with displacement. This effect is due to the finite energy range for the evaluation of the metric (-5 to 5 in this case) and could be eliminated by extending the energy range. However, for real logderivatives the appropriate energy range is decided beforehand and is not adjusted to account for the location of singularities. In fact, a typical logderivative could have multiple singularities, and extending the evaluation range to account for a nearby singularity would likely uncover additional singularities.

Because of the issues discussed above, a metric based on logderivatives is fundamentally inappropriate for describing the disagreement between AE and PS scattering properties. However, we can address both problems using instead the arctangent of the logderivatives,

$$A_l(E) = \tan^{-1} \frac{\Psi'_l(E,r)}{\Psi_l(E,r)}\bigg|_{r=r_c},$$ (4.5)

where l is the angular momentum channel. We define the arctangent metric for a single angular momentum channel as

$$S_{a,l} = \sqrt{\frac{1}{N}\sum_{E=E_1}^{E_2}\left(A_{l,\text{PS}}(E) - A_{l,\text{AE}}(E)\right)^2},$$ (4.6)

52

where AE and PS have the same meaning as before.

To demonstrate the efficacy of the metric, we take the arctangent of the mock logderivative

$$A_{\text{mock}} = \tan^{-1} L_{\text{mock}}(E) = \tan^{-1} \frac{1}{E-a} \tag{4.7}$$

Here, we do not have to worry about smoothing at $E - a$, because taking the arctangent removes the $1/E$ singularity. We let A_{mock} at $a = 0$ represent the AE arctangent curve, shown by the solid line in Figure 4.3a. Again, the parameter a is used to shift the singularity feature along the x-axis, representing various amounts of disagreement with the AE mock arctangent. A_{mock} for various values of a are shown in Figure 4.3a as the dashed lines. Now we can evaluate the arctangent metric for various displacements a using equation 4.6 with $A_{l,\text{AE}}(E) = A_{\text{mock}}(E)\big|_{a=0}$ and $A_{l,\text{PS}}(E) = A_{\text{mock}}(E)\big|_{a}$. Note that we enforced the arctangent function to return values between $-\pi$ and 0 to ensure continuity of the arctangent curve (for general arctangent curves, we ensure continuity using our phase unwrapping algorithm discussed in Section 4.4.1). As illustrated in Figure 4.3b, larger displacements of the feature mean larger values of the arctangent metric, even when the feature is displaced outside the energy sampling range. Also, there is no erratic variation of the arctangent metric with small changes in displacement as was observed in Figure 4.1b. Therefore, S_A better represents the disagreement between AE and PS scattering properties.

We can further demonstrate how the arctangent handles features in the logderivatives by using the $l = 1$ logderivative and arctangent from a silicon pseudopotential created during this work. In Figure 4.4, we provide the logderivative and arctangent for two different energy sample densities. The ghost state near -2 Ry is visually apparent in all four representations. However, we notice that the magnitude of the chirp (numerical representation of the singularity) is sensitive to the sample density. Yet the arctangent appears to be insensitive to the sample density. We can see the variation in both S_L and S_A metrics in Figure 4.5 when varying the number of energy sample points used to calculate these metrics. S_A remains relatively constant (Figure 4.5b), while S_L is very sensitive to the sampling (Figure 4.5a). The
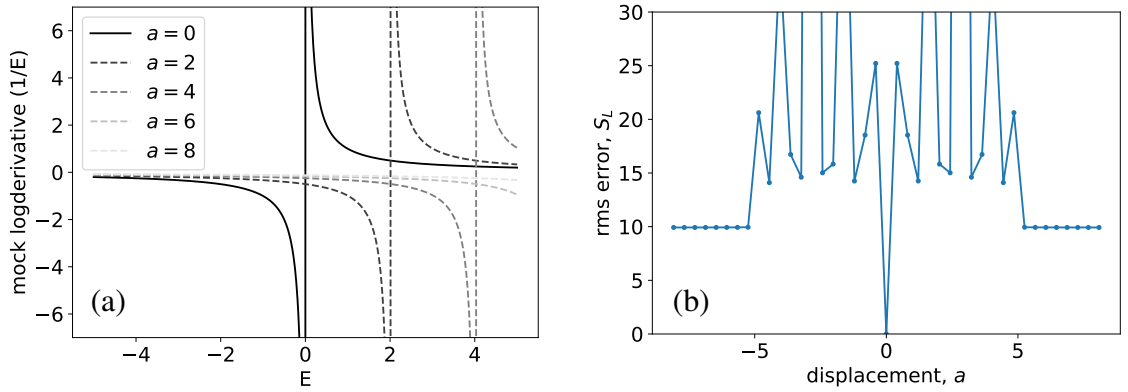
Figure 4.1: (a) Logderivative metric S_L shown with a few representative displacements and (b) the RMS error of the displaced logderivative relative to $a = 0$.
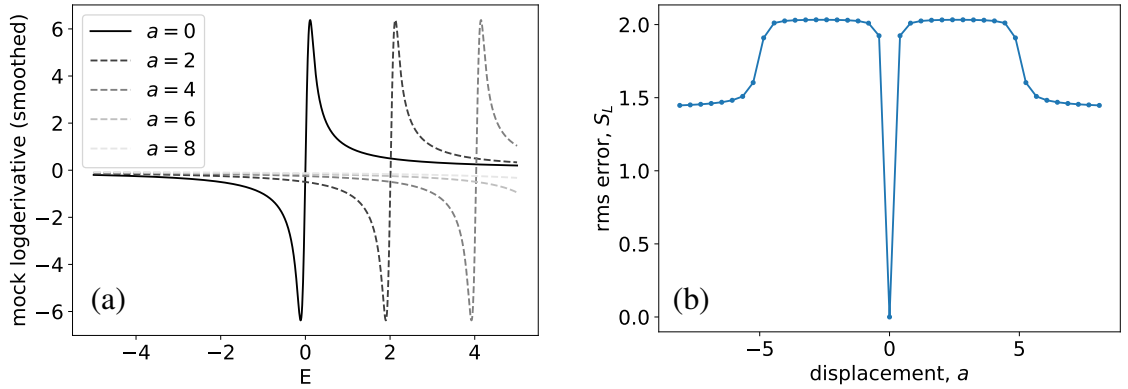
Figure 4.2: (a) Smoothed logderivative metric S_L shown with a few representative displacements and (b) the RMS error of the displaced smooth logderivative relative to $a = 0$.
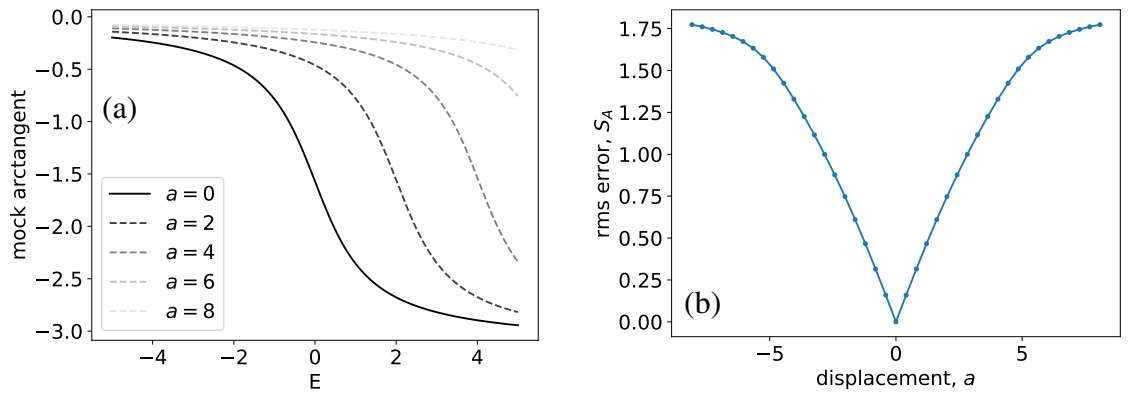
Figure 4.3: (a) Arctangent metric S_A shown with a few representative displacements and (b) the RMS error of the displaced arctangent relative to $a = 0$.

54

sensitivity of S_L to the sampling is undesirable, since the metric should ideally be independent of this sampling. Of course, even for S_A, the sampling must be fine enough to resolve the features corresponding to singularities in the logderivatives. But above this threshold, which is apparently lower than the sampling density shown in Figure 4.5, S_A remains stable as desired. The global increasing trend in the logderivative metric with increasing sampling is due to the singularity at the ghost state energy (the metric would likely be infinite for continuous sampling), and the erratic variations are related to how the singularity at the ghost state lines up with the sample points.

The arctangent has the property that it shifts by $\pm\pi$ whenever there is a singularity in the logderivative, which allows the approach to find flawed PAWs. We have already seen this shift behavior in Figure 4.4b and d. Figure 4.3a also shows this characteristic shift, but the transition is much slower due to the width of the singularity. Regardless of the rate of shift, this behavior leads to disagreements in the arctangents if singularities in the logderivatives are slightly misaligned. In Figure 4.6, we show the additional benefit that these phase shifts can illuminate ghost states that do not register noticeably on the logderivatives. Figure 4.6 shows a logderivative curve from another Si PAW selected from the design space. Here, the agreement seems perfect, and no singularities are visible. However, taking the arctangent, you see a shift of $-\pi$ near $E = -0.5$, indicating a singularity in the logderivatives. Indeed, if you zoom in near $E = -0.5$ on the logderivatives you can see a small blip in the plot. This blip will not show up as a large error when evaluating the logderivative metric S_L but will result in a large value of S_A. Even a user inspecting visually will likely miss this ghost state in the logderivatives, whereas it readily appears in the arctangent.

4.3   Efficient Development of Pseudopotentials and PAWs

In the previous section, we demonstrated that the arctangent metric S_A reliably represents disagreements between AE and PS scattering properties because it is numerically smooth and increases monotonically as features in the logderivatives become more mis-
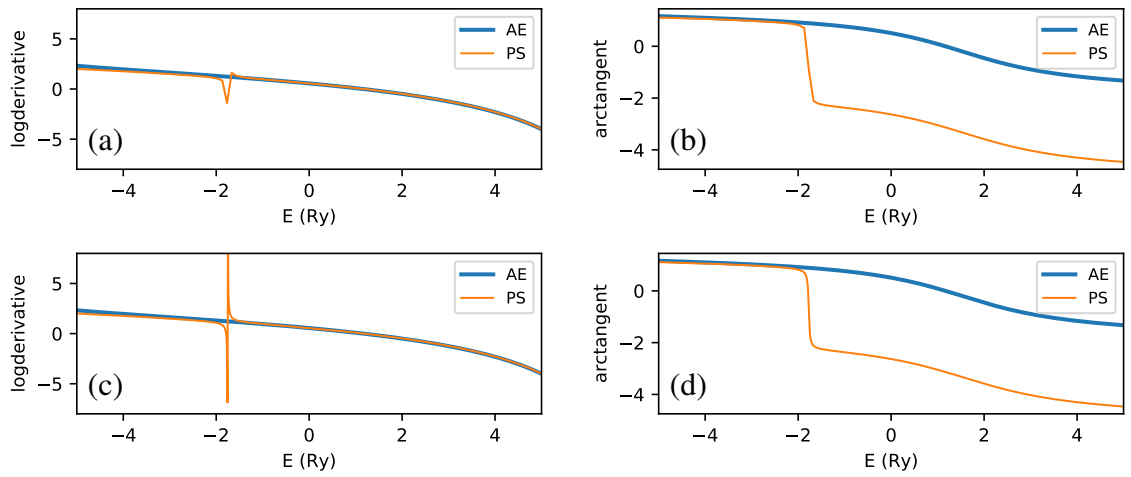
Figure 4.4: The logderivative and arctangent curves are shown for sampling of 100 (a-b) and 1000 (c-d) points to illustrate the difference in resolution at the ghost state near -2 Ry. This silicon PAW has a narrow ghost state near $E = -2$ in the $l = 1$ angular momentum channel, which is visible in (a) and (b). The arctangent metric is relatively unaffected by the sampling (b to d, also refer to Figure 4.5(b) (b)), while the logderivative metric varies erratically (a to c, also refer to Figure 4.5(a)).
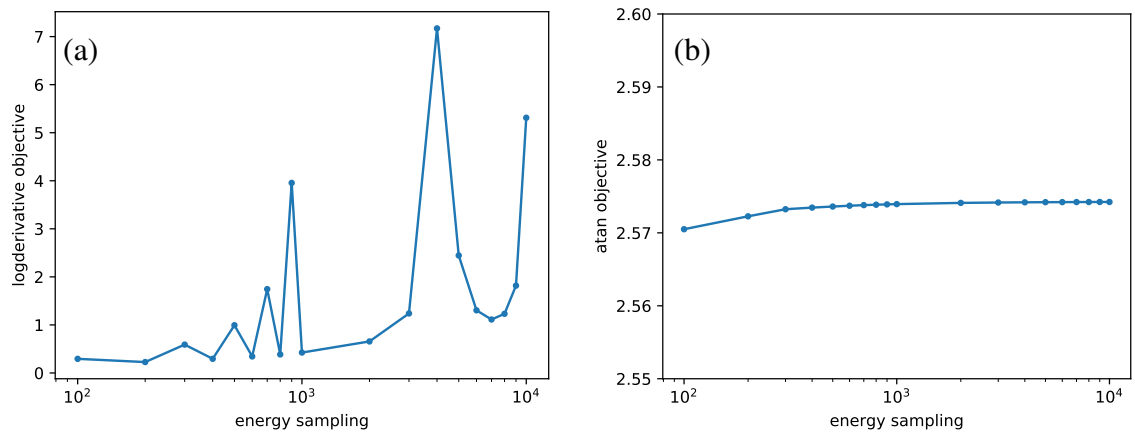
Figure 4.5: Relationship between the logderivative and arctangent metrics and the sampling of the logderivative energy range.
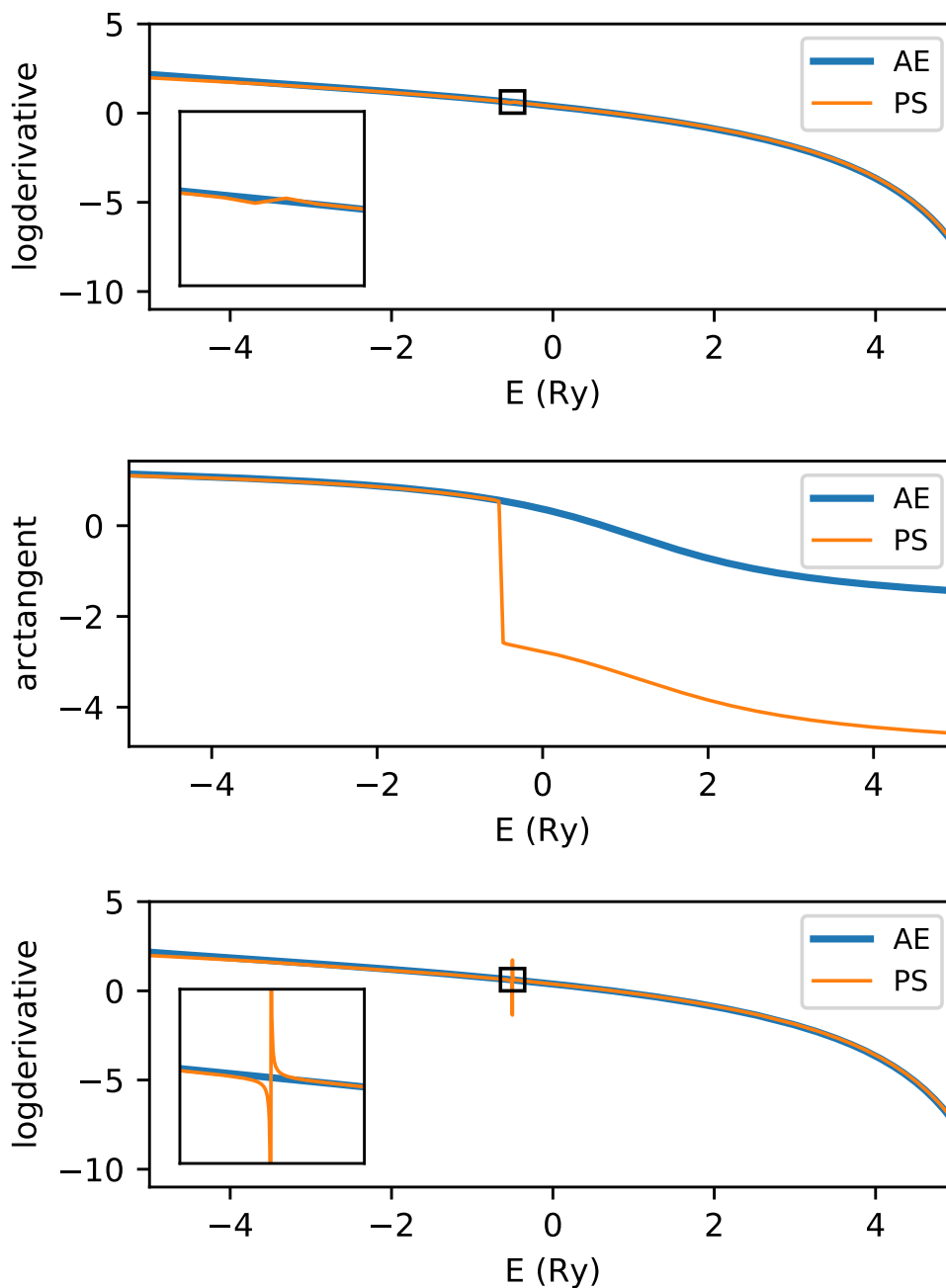
Figure 4.6: Example of a hidden ghost state in a silicon PAW. In the logderivatives (top) no ghost state is visible. However, a π shift is shown in the arctangent curve near -0.5 Ry (middle), indicating a ghost state. Indeed, if you zoom in to this region of the logderivatives (top, inset), you can see a small blip indicating a ghost state. The number of sample points is 200. Not only would this ghost state not be captured by a logderivative based metric, it is also so narrow that it is practically undetectable by visible inspection. This demonstrates the power of the arctangent metric for detecting ghost states, both automatically, and via visual inspection. The logderivative curve for a finer energy sampling of 10,000 is also shown (bottom) to prove that the ghost state truly exists.

aligned. Additionally we have shown how S_A reliably indicates ghost states due to its $\pm\pi$ shifting property, even at coarser energy sampling. Already we see the value of the metric, as it cheaply captures scattering agreement with a single number. In this section, we will describe how we can utilize the simplicity and effectiveness of this metric to speed up the development of pseudopotentials.

4.3.1   The Δ-factor

The Δ-factor has been posited as the de facto standard for pseudopotential and PAW accuracy. After all, it compares an all-electron solution to a pseudized solution using physical quantities (lattice constant and bulk modulus derived from an equation of state) in its measure. The prevailing thought is that if the pseudized solution produces the same physical results as the all-electron solution, then the pseudized part must be correct. The assumption from the user who is armed with the knowledge of a low Δ-factor then is that the designer of the pseudopotential has ensured the scattering properties are sufficiently accurate.

Consequently, we will evaluate the PAWs in the context of the Δ-factor, whose calculation details are provided in Section 4.4.2. The Δ-factor is a solid metric, meaning an equation of state must be calculated using the PAW in a solid. It is related to the bulk modulus and equilibrium lattice parameter calculations in that they all are derived from the equation of state.

4.3.2   Parameter sweeps and testing the Δ-factor

Of the metrics we have discussed to indicate the quality of a pseudopotential (S_A, S_L, and Δ), none was initially assumed to be better than the others, although we've demonstrated some advantages of S_A over S_L already. In fact, we can't identify any single metric as the ultimate and universal standard. Instead, we can only examine relationships between these metrics and look for trends and make judgments about their suitability for identifying the highest quality pseudopotentials. To explore the relationships between the various

metrics used to evaluate the quality of pseudopotentials, we generated a large collection of PAW pseudopotentials for Na, Mo, and Si using parameter sweeps. We chose these three elements, which we tested in three different materials, to sample different types of bonding: Mo in BCC Mo for metallic bonding; Li in rock-salt LiF for ionic bonding; and Si in zinc-blende SiC for covalent bonding. The details of the parameter sweep are included in Section 4.4.3. By spanning a relatively large design space, we created many PAWs with a range of quality from exceptional to unusable.

For each PAW we calculated the Δ-factor, a solid metric, as well as the atom metrics S_A and S_L which are calculated for each angular momentum channel of the pseudopotential and combined for the total metric. The total logderivative metric is then

$$S_L = \sum_l S_{L,l}, \tag{4.8}$$

and the total arctangent metric is

$$S_A = \sum_l S_{A,l}. \tag{4.9}$$

We proceed by first plotting the arctangent metric S_A against the Δ-factor (Figure 4.7) for every pseudopotential in each material we tested. Small Δ-factor values indicate that the physical quantities produced by the pseudized solution match those of the all-electron solution; small S_A values indicate the lack of ghost states and close agreement between the AE and PS scattering properties. The two metrics measure seemingly two different things, yet the relationship between the Δ-factor and S_A is remarkable.

First consider the Na case (top two images in Figure 4.7). In general the two metrics corroborate the quality of the PAW, with the possible exception of a handful of outliers. Upon closer examination (see the zoomed plot on the right), however, we find that many PAWs have small Δ values but not so small arctangent values. This means that pseudopotentials that produce good match with physical quantities as indicated by the small Δ values may actually contain disagreements in the scattering properties at some energies. More in-

terestingly, a small arctangent automatically guarantees a small Δ-factor, as long as we can remove outliers, meaning the arctangent alone predicts accuracy in the solid in this case. For Mo, a similar correlation between the Δ-factor and S_A can be observed in Figure 4.7. For Si, the trend is not so clear, but in the region of reasonable values for Δ and S_A (see zoomed plot), a correlation is visible. Si and Mo points with good arctangents but bad Δ values will be discussed in the next section.

Because the logderivative has been the canonical approach for finding pseudopotentials with ghost states, we show the relationship between the logderivative metric S_L and Δ-factor in Figure 4.8. Unlike the arctangent, the logderivative metric does not predict accuracy in the solid. Even in the zoomed regions near the optimal points, no meaningful relationship is seen. Moreover, any apparent trends are different between the three materials suggesting that the there is nothing universal about these relationships.

Some of the pseudopotentials we generated caused Socorro to crash during calculations of the total energy. These crashes, the result of the failure of a linear algebra routine, likely indicate that these pseudopotentials are of poor quality. These pseudopotentials are indicated at the bottom of the Mo and Si plots in Figures. 4.7 and 4.8 on the line labeled X1. Evidently, a good S_A does not always indicate a quality pseudopotential, since some of these ill-conditioned points had optimal values of S_A yet still caused crashes in Socorro. Additionally, with some pseudopotentials the total energy is not converged within 10^{-8} Ry after 100 self consistent cycles (10^{-7} Ry after 200 cycles for Mo) for one or more of the electronic structure calculations during calculation of the Δ-factor. This alone is not an indicator of a bad PAW, but does imply the PAW is not optimal for speed. These pseudopotentials are also indicated in Figures. 4.7 and 4.8, labeled X2.

4.3.3   Reducing computational expense of optimizations

We can use the fact that S_A identifies pseudopotentials with high accuracy in the solid to expedite the process of developing new pseudopotentials. Expensive Δ-factor calculations

or other tests in the solid are necessary, but really only need to be performed for pseudopotentials with good scattering properties. Therefore, it would be valuable to filter out points with high S_A during optimizations before calculating solid metrics. For example filtering out Na PAWs from our parameter sweep with $S_A > 0.1$ reduces the number of pseudopotentials by 75%. The S_A cutoff is somewhat arbitrary, and acceptable values of S_A may depend on the application of the pseudopotential. Nevertheless, $S_A < 0.1$ seems to be a suitable minimum requirement for quality PAWs based on the zoomed plots in Figure 4.7. Applying this same threshold for Mo and Si means that we have reduced the number of solid calculations that need to performed by 78% and 57% respectively. However, for Mo and Si, low S_A does not guarantee a good value of the Δ-factor like it does for Na. In other words, the scattering properties alone may not be sufficient to determine if a PAW is accurate in the solid. Although the results for Mo and Si are, in general, similar in character to those of Na, the metal and the covalent material exhibit a few more outliers. That is, we find PAWs with small arctangent values but large Δ-factor values. This means that arctangent alone is not a universal indicator of high-quality PAWs.

The apparent discrepancy can partially be explained by PAW projectors that are too large. To compensate for this shortcoming, we introduce a new metric for PAWs that we call the projector magnitude ratio. For a single projector,

$$d_i = \sqrt{\frac{\langle \tilde{p}_i | \tilde{p}_i \rangle}{\langle \tilde{\psi}_i | \tilde{\psi}_i \rangle}}. \tag{4.10}$$

Here $\tilde{p}_i$ is the projector function and $\tilde{\psi}_i$ the PS partial wave for quantum numbers $n_i l_i$. The projector metric is then defined as the RMS of the projector magnitude ratios. That is,

$$S_P = \sqrt{\frac{1}{N_i} \sum_i d_i}. \tag{4.11}$$

The projector magnitude ratio represents the size of the projectors compared to the size of the wave functions. An ideal projector has a similar magnitude to the corresponding PS

wave function, and thus has a value near unity. What should be avoided are projectors much larger than their corresponding PS wave functions,[7,1] which would have larger values such as 100 (see Figure 4.9). Like the logderivative and the arctangent, the projector magnitude is an atom metric used to indicate the quality of a PAW. In Figure 4.7 each point represents a different PAW, and the color corresponds to projector magnitude ratio. Examining the colors of the points, we find that the points with small arctangent values but large Δ values do, in fact, have large projectors. These data suggest that the arctangent combined with the projector magnitude ratio could be used as a universal method for indicating quality PAWs, further meaning that an atomic metric can indeed indicate accuracy in the solid.

Returning to the idea of filtering out pseudopotentials based on atom metrics before running Δ-factor calculations, we now can now filter out points with high projector magnitudes (> 60) as well. Filtering by both projector magnitude and arctangent leads to a reduction in points of another 15-40% on top of those filtered out by S_A alone. Filtering by both $S_A < 0.1$ and $S_P < 60$ we see a total reduction of pseudopotentials of about 50-75% for all materials tested.

4.4   Methods

4.4.1   Phase unwrapping of the arctangent

Phase unwrapping of $A_l(E)$ is necessary to ensure suitability of the the arctangent as a metric. Without this phase unwrapping, there are discontinuities in the arctangent curve where it should be continuous. We can enforce continuity of the arctangent curve by shifting $A_l(E)$ by $\pm\pi$ whenever the curve crosses $\mp\pi/2$.

We have defined the arctangent in Equation 4.5, d. The phase unwrapped arctangent is defined as

$$\hat{A}_i = A_i + s_i \tag{4.12}$$

where i is the index for the energy sampling, A_i is the arctangent as defined in Equation 4.5
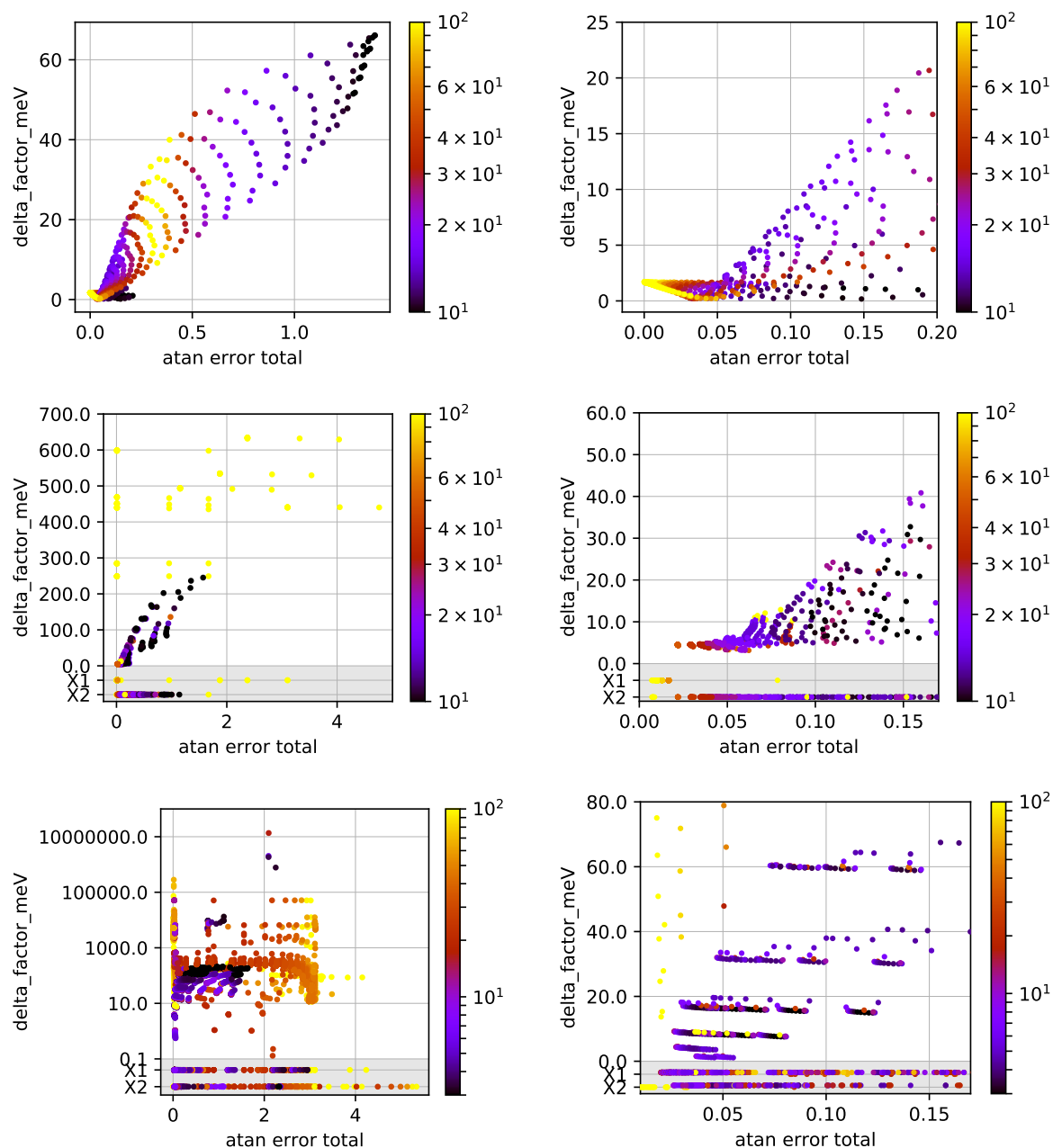
Figure 4.7: Δ-factor vs arctangent metrics for for Na (top), Mo (middle), and Si (bottom). The full objective space is shown in the left plots and a zoomed view is shown in the right plots. For some pseudopotentials, an error is thrown during the Δ-factor calculations. We have labelled these at the bottom of the plot with an X1. Additionally, with some pseudopotentials the total energy is not converged within 10^{-8} Ry after 100 self consistent cycles (10^{-7} Ry after 200 cycles for Mo) for one or more of the electronic structure calculations during calculation of the Δ-factor. These are labeled with an X2 at the bottom of the plot. The color is the projector metric.
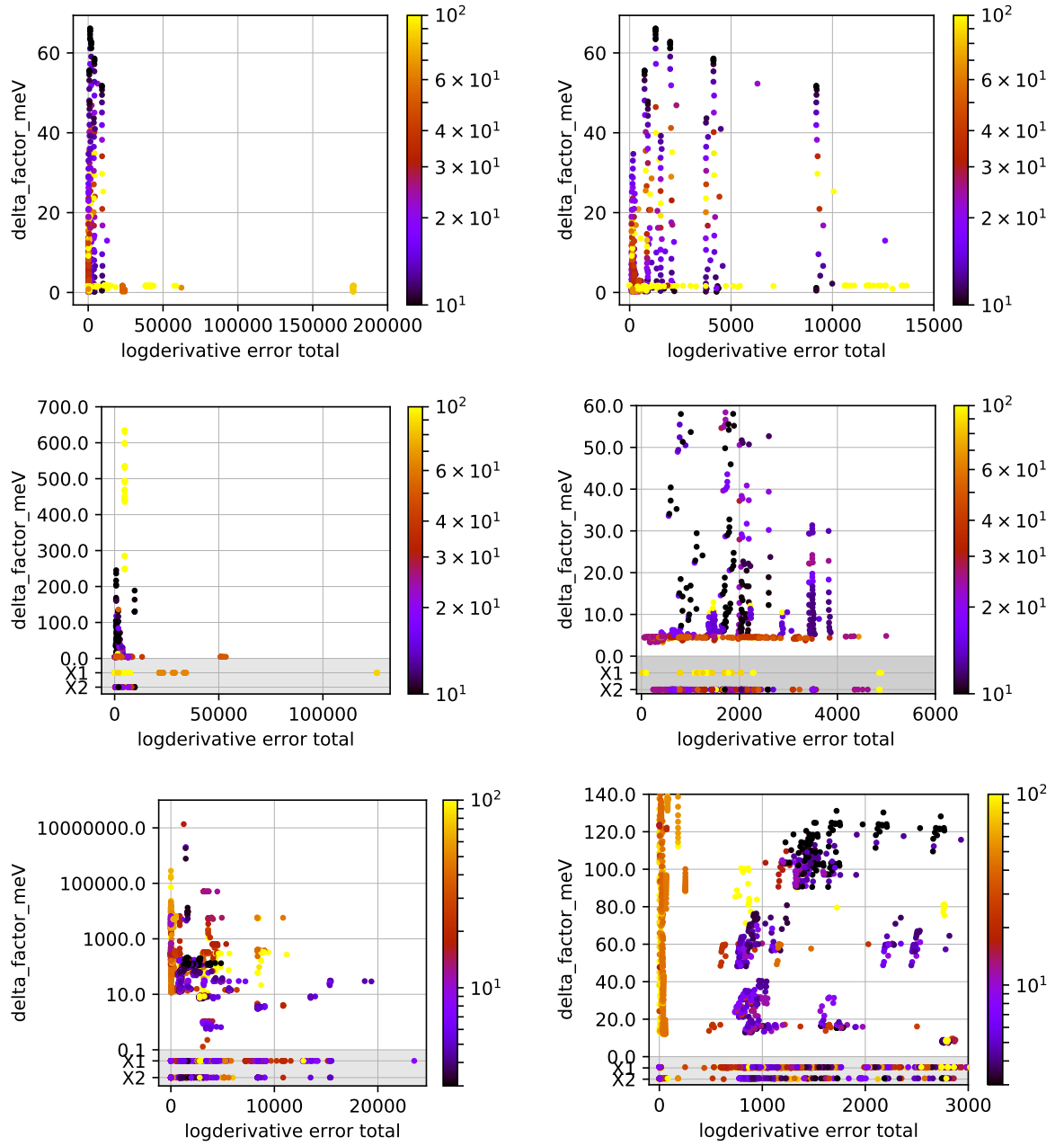
Figure 4.8: Δ-factor vs. logderivative metrics for Na (top), Mo (middle), and Si (bottom). The full objective space is shown in the left plots and a zoomed view is shown in the right plots. For some pseudopotentials, an error is thrown during the Δ-factor calculations. We have labelled these at the bottom of the plot with an X1. Additionally, with some pseudopotentials the total energy is not converged within 10^{-8} Ry after 100 self consistent cycles (10^{-7} Ry after 200 cycles for Mo) for one or more of the electronic structure calculations during calculation of the Δ-factor. These are labeled with an X2 at the bottom of the plot. The color is the projector metric. One outlier with $S_L > 3 \times 10^6$ is not shown in the top left figure.
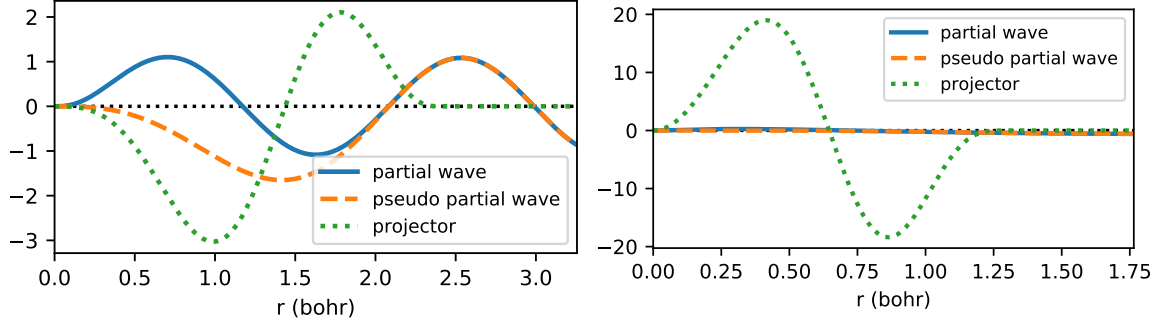
Figure 4.9: Examples of projectors with appropriate magnitude (left, $d = 1.8$) and large magnitude (right, $d = 53$) compared to the PS wave function.

```
for i in range 1 to N
    if A(E_i)·A(E_{i-1}) < 0 and Ψ(E_i, r_c) < 0,
        s_i = s_{i-1} − π·sign(Ψ'(E_i, r_c))
    else if (Ψ(E_{i-1}) − Ψ(E_{i-2}))·(Ψ(E_i) − Ψ(E_{i-1})) < 0
            and |(Ψ(E_{i-1}) − Ψ(E_{i-2})) − (Ψ(E_i) − Ψ(E_{i-1}))| > 0.01
            and s_{i-1} = s_{i-2} = s_{i-3},
        s_i = s_{i-1} + π·sign(Ψ(E_{i-1}) − Ψ(E_{i-2}))
    else,
        s_i = s_{i-1}
done
```

Figure 4.10: Algorithm for determining shifts for phase unwrapping the arctangents.

using a four quadrant arctangent function, s is a shift that depends on the energy sample point, and we have dropped the angular momentum subscript l for clarity. The algorithm for determining the shift s_i is pseudocoded in Figure 4.10. $\hat{A}$ is the arctangent used to calculate metrics throughout this paper, not A.

The logderivatives and thus $A_l(E)$ are evaluated on a set of discrete points, and sometimes the ghost states are narrow. Therefore the energy sampling must be fine enough to sufficiently resolve the singularities. In previous work, we found a sampling of 600 points/Ry to be sufficient. For this work we use a sampling of 1000 points/Ry. This sampling is much finer than Atompaw's current default value of about 10 points/Ry. The logderivatives and the arctangents are both evaluated on the energy range -5 to 5 Ry.

4.4.2 Calculating the Δ-factor

The Δ-factor, originally proposed in Reference,[18] is defined for this work as

$$\Delta = \sqrt{\frac{\int_{0.94V_0}^{1.06V_0}(E_1(V)-E_2(V))^2dV}{0.12V_0}}, \tag{4.13}$$

where a comparison of the energy-volume curve between an all-electron solution (E_1, considered exact) and a pseudopotential DFT solution (E_2) is made. In our case, the two solutions are derived from the full-potential linearized augmented plane wave (FP-LAPW) code Elk[26] version 4.3.0.6 and the plane wave DFT code Socorro,[25] respectively. V_0 is the average of the equilibrium volumes given by two solutions. To find the energies and equilibrium volumes we calculated the total energy at 5 volumes in the range $(0.95a_{\text{guess}})^3$ and $(1.05a_{\text{guess}})^3$, with guess lattice parameters of $a_{\text{guess}} = 10.82$ Bohr for rock-salt NaCl, $a_{\text{guess}} = 5.98$ Bohr for BCC Mo, and $a_{\text{guess}} = 8.29$ Bohr for zinc-blende SiC. These energy-volume data are then fit with a Murnaghan equation of state.[30] Equation 4.13 is evaluated numerically using the trapezoidal rule to get the Δ-factor. The energy calculations for the Δ-factor were performed using the GGA-PBE exchange correlation functional in both codes. We used a 7x7x7 Monkhorst Pack k-point grid for NaCl and SiC and an 11x11x11 Monkhorst Pack k-point grid for Mo. It is not necessary that the k-point grid be fine enough to converge the physical properties, as we are only interested in a comparison between the pseudopotential calculation and an all-electron calculation, and we use the same k-point grid for both.[14] For the calculations in Socorro, a relatively high plane wave cutoff of 120 Ry was used to ensure convergence.

There are multiple versions of the Δ-factor which are all in the same spirit but vary slightly in implementation. Differences include normalization of the metric, which solution (if not both) is used for V_0, which equation of state is fitted, and which structures and compounds are used. In many cases, the final Δ-factor is the average over many compounds. In our case, we consider only one compound per element as a representative compound.

4.4.3 Parameter Sweeps

All pseudopotentials were generated by sweeping select input parameters normally used in specifying a PAW pseudopotential. The inputs were established from templates based on published pseudopotentials, and the parameters swept uniformly within specified ranges. The AtomPAW code version 4.1.0.3 was used to generate the pseudopotentials.[15] See Table 4.1 for the parameters varied, the ranges of the design space, and the parameter values in the original published pseudopotentials. The parameter r_c is the radius of the augmentation spheres. For the original Si pseudopotential, the last three values for r_c indicate r_{shape}, r_{vloc}, and r_{core} as described in Reference 15. Note that, for each pseudopotential in this work and for the original Na and Mo pseudopotentials from the literature, r_{shape}, r_{vloc}, and r_{core} were set equal to r_c. The energies E_l are matching energies for projectors, where l is the angular momentum channel.

The Na and Mo input file templates were based on the published PAWs from the Atompaw web page $<$2013 library. The radius r_c and the projector energies were varied during the parameter sweep. For Si, the template was from Atompaw's $>$2013 library. In this case, several modifications to the template were necessary, including the pseudization scheme because Socorro does not support l-dependent shape functions. For all pseudopotentials, we increased the logderivative sampling to 10,000 from the default value of 200. This increase in sampling does not affect the resulting pseudopotential, only the logderivative and arctangent data. Such a large sampling was used to improve resolution of the features in the logderivatives and arctangents.

element	parameter	range	step	original
Na	r_c	0.8-2.4	0.05	1.8
	E_p	0-10	0.5	2
Mo	r_c	1.2-2.6	0.1	2.2
	E_p	0-10	1	4
	E_d	0-10	1	6
Si	r_c	1.0-2.4	0.1	1.7, 1.5, 1.7, 1.7
	E_s	0-20	1	14
	E_p	0-20	1	14

Table 4.1: The parameters varied in the parameter sweeps. The r_c are PAW radii and the E_i are energies for projectors for angular momentum channel i. All radii are in Bohr; energies in Rydberg. The total number of PAWs tested were 693 for Na, 1815 for Mo, and 6615 for Si.

Chapter 5

Application of Improved Pseudopotential Optimization Techniques to a Low-Valency Gallium PAW for Faster DFT Calculations in GaN

5.1 Motivation

The study of point defects in gallium nitride (GaN) will benefit from accurate and fast ab-initio molecular dynamics studies using plane wave DFT with the PAW method. For studies of defects, large supercells are required because of long-range Coulombic interactions of the defects and long-range strain fields. Either the supercell must be large enough to reduce these effects, or multiple supercell sizes must be used to extrapolate cell-size-dependent results to infinite cell size. The large cell sizes and necessity of cell size convergence studies make quality defect calculations extremely computationally expensive. Furthermore, if one wants to perform MD studies of defects, even more expense is added since the electronic structure must be calculated at each time step. Therefore, efficient PAWs are a necessity for plane wave DFT studies of point defects in GaN. The primary objective of this chapter is thus to optimize a custom Ga PAW that will give accurate results for defect studies in GaN while significantly reducing the computational expense compared to existing Ga PAWs. We do not optimize an N PAW and instead use a previously published 5-electron PAW because we expect the greatest improvement in speed to be gained by optimizing the Ga PAW with fewer electrons in the PAW valence partition.

The goal is not to find the perfect PAW for Ga that is transferable to all materials. Instead, we customize a Ga PAW for significant performance gains in studies of defects in GaN. We take this approach because, to ensure transferability, compromises must be made (e.g. more valence electrons or smaller radii) that can drastically reduce efficiency. Therefore, because our Ga pseudopotential is not required to be accurate for a wide chemical

space, we can customize the pseudopotential to be as fast as possible while still capturing the physics necessary for the problem at hand.

The main change to our Ga PAW compared to the previously published PAWs in the Atompaw or Abinit libraries is the removal of the 3d electrons from the valence partition, meaning the Ga PAW has 3 electrons in the valence instead of 13.[*] Removing the 3d electrons from the valence partition improves the computational efficiency of calculations by reducing the number of electrons that need to be calculated. The number of electrons reduces from 18 to 8 per formula unit since the N PAW has 5 electrons in the valence partition, enabling a potential reduction in computation time of about a factor of two. Previous research has suggested that the gallium 3d electrons are important in studies of GaN since they affect band gap and structural property results. This effect is attributed to the overlap and hybridization of the gallium 3d electrons with the nitrogen 2s electrons.[46,47,48] However, other research has shown that this interaction is an artifact of the approximation to the exchange-correlation functional, which falsely raises the energy of gallium 3d electrons so that it overlaps with the nitrogen 2s; in reality the energy of the gallium 3d band is about 3 eV lower than the nitrogen 2s band.[49] In this chapter, we show that the accuracy of our quantities of interest, namely defect transition levels, is not adversely affected in calculations in GaN by removing the 3d electrons from the valence compared to calculations using a 13-electron Ga PAW. To demonstrate this accuracy in GaN using our optimized Ga potential, we calculate lattice parameters, band gaps, and defect levels in w- and zb-GaN and compare to results given by a published 13-electron Ga PAW.

Given the valence/core partition, the next step in generating a PAW is selecting the radii and other parameters. For this we use automated approaches where thousands of PAWs can be efficiently generated and then tested using quality objectives. We have demonstrated the effectiveness of these techniques for pseudopotential development in the previous chapters, and they are also discussed in the literature.[11,23] In this chapter, we first use a parameter

---

[*]VASP and GPAW have 3-electron PAWs, but they are not useable with Socorro, and we are unsure of their quality.

sweep to uniformly search the space, varying important parameters for PAW quality. As a demonstration, we also run a genetic algorithm in the same search space and show that it more efficiently finds the same or better PAWs compared to a parameter sweep.

Following standard practice for pseudopotential testing, we test each pseudopotential both in an isolated atom and in calculations in the solid. The accuracy of each calculation is determined by comparing to all-electron DFT calculations, which are considered exact for the purposes of this work. In the isolated atom, we use the arctangent metric, which we have already shown in Chapter 4 to be the best known scattering property metric. We will further demonstrate in this chapter the advantages of this metric for prediction of accuracy in the solid. For evaluating PAW accuracy in the solid, we use a standard lattice parameter calculation. Some studies also optimize PAWs or pseudopotentials for speed[23, 10] using metrics such as estimated FLOPS or the energy cutoff required for convergence within a certain tolerance. In this work, we do not optimize the Ga PAW for speed since we assume that that the required cutoff for our calculations is determined by the N PAW and not the Ga PAW. We assume this because the N atom is more compact than Ga (it has smaller PAW radius) and thus should require more plane waves in the basis set. In any case, the reduction of the number of gallium valence electrons provides a significant speed advantage already.

5.2   Methods

5.2.1   Optimizations

A single PAW can be described as

$$d = \{r_c, E_p, E_{d1}, E_{d2}\} \tag{5.1}$$

where r_c, E_p, E_{d1}, and E_{d2} are parameters used to specify the PAW. The other parameters for the PAW are not varied during the optimization. The choice of which parameters to vary, and other settings for the PAW are discussed in Section 5.2.1.1. Each PAW, d, is

71

evaluated by calculating two accuracy metrics, S_a and S_s, which both need to be minimized by the optimization. S_a is a scattering property metric for the isolated atom and S_s is a metric based on lattice constant calculations in the solid. Both objectives are discussed in Section 5.2.1.4. In this chapter we present results from both a parameter sweep, described in Section 5.2.1.2, and a multi-objective genetic algorithm, described in Section 5.2.1.3.

5.2.1.1 PAW design

For the optimizations, we vary several parameters important for specifying the PAW, which are inputs to the Atompaw code. For reference, the template for the Ga PAW input file is shown in Figure 5.1. The parameters varied are indicated in curly braces, and the other parameters are not varied. The PAW input parameters relevant to this chapter are described here in further detail.

We treat only the 4s and 4p electrons as valence, and the remaining electrons (including the 3d electrons) are frozen in the core. We set all radii in the PAW equal to r_c. This includes r_{paw}, r_{shape}, r_{vloc}, r_{core}, and the matching radius for each basis function. Setting the radii equal is necessary in order to keep the number of design variables small so that the number of points in the parameter sweep is not prohibitively large. Setting these radii equal is common in published PAWs, and we do not expect it to hinder our search for high-quality PAWs. We found it necessary to include 1 additional s, 1 additional p, and 2 additional d projectors to get both accurate scattering properties and lattice constants in GaN. The matching energies for these projectors are E_s, E_p, E_{d1}, and E_{d2}. In order to keep the number of design variables at 4, we fixed E_s at 5 Ry while varying E_p, E_{d1}, and E_{d2}. We found that the PAW results are fairly insensitive to the choice of E_s. The PAWs were generated with the GGA-PBE functional, and all DFT calculations in this chapter using Elk and Socorro use this functional.

```
Ga 31
XC_GGA_X_PBE+XC_GGA_C_PBE logderivrange -5 5 6001
4 4 3 0 0 0
4 1 1
0 0 0
c
c
c
v
c
c
v
c
2
{Ga_RC}
y
5
n
y
{Ga_EP}
n
y
{Ga_ED1}
y
{Ga_ED2}
n
CUSTOM RRKJ
3 0
{Ga_RC}
{Ga_RC}
{Ga_RC}
{Ga_RC}
{Ga_RC}
{Ga_RC}
SOCORROOUT
END
```

Figure 5.1: Input file template for Atompaw Ga. The variable parameters are in curly braces. Notice that all radii are set equal.

5.2.1.2 Parameter Sweep

We use 11 values for each parameter varied in order to adequately sample the space, meaning the time required to perform the parameter sweep scales as 11^N where N is the number of design variables. With that in mind, we chose to use four design variables so that there are a total of 14,641 design points in the parameter sweep. See Table 5.1 for the values of each parameter during the sweep. For each set of parameters, we calculate S_a and S_s to evaluate that PAW.

parameter	min	max	N
r_c (Bohr)	2	3	11
E_p (Ry)	0	10	11
E_{d1} (Ry)	0	10	11
E_{d2} (Ry)	0.5	10.5	11

Table 5.1: Parameter sweep variable ranges. The energy ranges and step sizes for E_{d1} and E_{d2} were chosen so that two different d projectors would never have the same energy.

5.2.1.3 Genetic Algorithm

The DAKOTA software[24] is used for the multi-objective genetic algorithm. The settings for the genetic algorithm are shown in Table 5.2. For the multi-objective genetic algorithm, we vary the same parameters in the same range as the parameter sweep. The objectives S_a and S_s are also the same as for the parameter sweep.

5.2.1.4 Objectives

The objective S_a is designed to evaluate a PAW for its scattering properties in the isolated atom. Typically, the PS logderivatives are visually compared to the AE logderivatives to ensure that the curves match and that discontinuities line up. What we need instead for our optimization is a way to evaluate the logderivatives automatically, representing the agreement of the curves with a single metric. The integral of the difference between the

population size	100
max function evaluations	2500
initialization type	unique random
crossover type	shuffle random
num offspring	2
num parents	2
crossover rate	0.8
mutation type	replace uniform
mutation rate	0.1
fitness type	domination count
replacement type	below limit = 6
shrinkage fraction	0.9
num generations	40

Table 5.2: Genetic algorithm settings for DAKOTA. The variable ranges are the same as for the parameter sweep shown in Table 5.1.

logderivative curves,

$$S_L = \sqrt{\frac{1}{N} \sum_{E=E_1}^{E_2} (L_{\text{PS}}(E) - L_{\text{AE}}(E))^2} \tag{5.2}$$

where

$$L(E) = \left. \frac{\psi'(r)}{\psi(r)} \right|_{r=r_c} \tag{5.3}$$

is an obvious first choice, and has indeed been used by other researchers.[11] However, we use the arctangent of the logderivatives instead for two reasons: (1) the arctangent of the logderivatives is numerically smooth without discontinuities, making the error integral more meaningful, and (2) the arctangent curve can identify ghost states that are visually imperceptible in the logderivatives and completely missed by integrating the logderivatives. We discussed the advantages of the arctangent metric in more detail in Chapter 4. We will further demonstrate in Section 5.3 that the arctangent metric more reliably predicts accuracy in the solid. The arctangent metric is defined as the root mean square of the error in the atan curve,

$$S_a = \sqrt{\frac{1}{N} \sum_{E=E_1}^{E_2} (A_{\text{PS}}(E) - A_{\text{ES}}(E))^2} \tag{5.4}$$

where the atan curve is defined as

$$A(E) = \tan^{-1} \left. \frac{\psi'(E,r)}{\psi(E,r)} \right|_{r=r_c}. \tag{5.5}$$

We use a range of -5 to 5 Ry to integrate the logderivatives and atan.

There is an additional consideration for the arctangent objective that arises because we are pseudizing the 3d electrons in Ga. There is now a discontinuity in the AE logderivatives around -1.5 Ry that corresponds to the 3d bound state. We do not want to represent this discontinuity in our PS logderivatives. Therefore, we consider logderivatives and atan only for $E > 0$ for the d channel.

As described in Section 4.4.1, the arctangent curve must be phase shifted at appropriate points to ensure its continuity. The phase unwrapping algorithm we use relies on a very fine sampling of the logderivatives to resolve all discontinuities in the arctangent curve. We determined that a sampling of 6000 points was sufficient over the range of 10 Ry to resolve all discontinuities. With coarser sampling, the phase unwrapping algorithm will miss some discontinuities. To ensure 6000 points was sufficient, we calculated the arctangent metric for each PAW with a sampling of 100000 instead of 6000 and got the same values within a small tolerance.

The objective S_s is designed to evaluate a PAW for its accuracy in the solid. To analyze the accuracy of each PAW in the solid, we calculated the zb-GaN equilibrium lattice parameter using each PAW and compared the values to the all-electron value. The general methods used for calculating equilibrium lattice parameters are described later. Our preliminary tests have shown that the zb-GaN lattice parameters correlate to the w-GaN lattice parameters, meaning no information is gained by calculating both. Because the zb-GaN lattice parameter is faster and simpler to calculate, we use the zb-GaN lattice parameter for the test in the solid during our optimizations.

The all-electron full potential linearized augmented plane wave (FP-LAPW) value as

76

calculated using the Elk code was considered to be the correct answer for the purposes of the lattice constant calculations. We compare to FP-LAPW results instead of experiment because we only want to test the effect of the pseudopotential on the results, and not the approximations inherent to DFT. We use a $3 \times 3 \times 3$ k-point sampling which may not completely converge the solution, but convergence is not necessary for comparison between two DFT calculations using the same sampling.[14] This low k-point sampling speeds up the optimization to make finding optimized PAWs tractable.

For the lattice constant calculation using Elk, three different convergence settings were used: normal, high quality (highq=.true.), and very high quality (vhighq=.true.). The lattice constant using very high quality settings was selected for the reference value. The difference between the very high quality and high quality settings was 0.0023 Bohr, so we considered this value converged.

For the calculations with Socorro using the PAWs, the calculations were performed at an energy cutoff of 80 Ry, which we determined was sufficient to converge the lattice constant values for all PAWs.

The objective for the accuracy of the PAW in the solid is then

$$S_s = |a_{\mathrm{PAW}} - a_{\mathrm{elk}}| \tag{5.6}$$

where a_{PAW} is the zb-GaN lattice constant given by the PAW being evaluated and a_{elk} is the reference lattice constant given by the FP-LAPW method.

The lattice constant calculations were performed exactly the same way for the all-electron and PAW cases. For zinc-blende GaN, we sampled 5 values of the lattice constant between -5% and +5% of the starting guess of 8.6 Bohr. This is a reasonable guess value because the all electron lattice constant was found to be 8.6012 Bohr. The Murnaghan equation of state[30] was fitted to the energy vs. unit cell volume data, and the minimum of the EOS curve was chosen as the equilibrium lattice constant.

When analyzing the optimization results, we calculated the wurtzite lattice parameters for some PAWs. For the wurtzite structure, both $a = b$ and c are varied. At each value of a and c, the internal coordinates are relaxed to get the u value for the wurtzite structure. This gives us an energy surface, $E(a, c, u)$, where the minimum gives the equilibrium lattice parameters. To find the minimum, we fit the energy data to the bivariate polynomial

$$
\begin{aligned}
E(a,c) = C_0 + C_1 a + C_2 c + C_3 a^2 + C_4 ac + C_5 c^2 \\
+ C_6 a^3 + C_7 a^2 c + C_8 ac^2 + C_9 c^3.
\end{aligned}
\tag{5.7}
$$

from which the local minimum can be found. We chose a starting value and varied $\pm 2\%$. The starting guesses were $a = 5.98$ Bohr and $c = 9.62$ Bohr. This is more expensive than a zinc-blende lattice parameter calculation because (1) the energy is evaluated on a 2d grid instead of a 1d grid, (2) there are 4 atoms instead of 2 in the primitive cell, and (3) a relaxation of the ionic positions is necessary at every grid point.

5.2.2   Defect Calculations

We calculated defect transition levels of three native point defects in w-GaN, a gallium vacancy, and the gallium interstitial at two different sites. The structures of each defect are shown in Figure 5.9. The defect calculations were performed in a $3 \times 3 \times 2$ w-GaN supercell containing 72 atoms, not including the defect. We compared band gaps and defect levels between the 13-electron and 3-electron PAWs to gauge the effect of removing the 3d electrons from the valence.

The defect transition level for charge states $q - 1$ and q is defined as

$$
E_L[q-1, q] = E[q-1] - E[q] - E_{\text{shift}}
\tag{5.8}
$$

where E_{shift} is conventionally the valence band edge energy. For this work, the shift will be defined on a case by case basis when comparing defect levels between different PAWs

(see Section 5.3.2). The energy, $E[q]$, at charge state q is calculated using the PAW being tested. The energy cutoffs used were 45 Ry for the 3-electron PAW and 50 Ry for the 13-electron PAW. All defect level calculations used a $2 \times 2 \times 2$ k-point sampling. A uniform compensating background charge is used to keep the total charge of the cell neutral. The ionic positions are fully relaxed for each charge state with a force tolerance of 10^{-4} Ry/Bohr. For all defect calculations, the lattice parameters are set to the lattice parameters of the bulk cell as calculated by the PAW being tested so that the results are self-consistent. Therefore, the lattice constants are slightly different for each PAW. Using the 3-electron Ga PAW, the calculated lattice parameters are $a = 6.092$ Bohr and $c = 9.891$ Bohr. Using the 13-electron PAW the calculated lattice parameters are $a = 6.095$ Bohr and $c = 9.929$ Bohr. The ionic positions are fully relaxed for each charge state.

For valid defect charge states, electrons added or removed are localized on the defect site. Otherwise, electrons are delocalized. For an infinite size supercell, we could identify valid charge states as those with defect transition levels within the bulk band gap. However, the bounds for valid charge states for a finite-sized supercell are not equal to the band edges in the bulk cell.[50] The effective bounds are calculated as

$$E_{\text{b,upper}} = \frac{E[-1] - E[0]}{Q} - E_{\text{shift}} \qquad (5.9)$$

$$E_{\text{b,lower}} = \frac{E[0] - E[-1]}{Q} - E_{\text{shift}} \qquad (5.10)$$

where $Q = 4/72$, the fractional charge of a 4 atom cell, equivalent to a 72 atom cell and E[q] is the energy for the non-defective structure calculated with a unit cell charge of q. The value for Q is fractional because the unit cell is small (4 atoms) but we want to simulate adding a single electron to a larger cell (72 atoms). E_{shift} is the same as in Equation 5.8.

5.2.3 Timing Calculations

To test the speedup of the 3-electron Ga PAW vs the 13-electron Ga PAW, we ran structure relaxations for w-GaN with a Ga vacancy. We chose a structure relaxation as a timing test because the underlying calculations are the same as MD, which is the desired application of the new Ga PAW. In both MD and structure relaxations, the interatomic forces are calculated and the atomic positions are updated at each step. To compare timing on a per-iteration basis, we fixed the number of iterations to 20 for the self consistent calculations of the electronic structure, and the number of relaxation steps to 11 for a total of 220 total iterations. All calculations were performed in the same computing environment with the same number of compute nodes, and the same nitrogen PAW was used in all cases. We ran the timing tests in two different cell sizes to evaluate the scaling of the speedup.

5.3   Results and Analysis

5.3.1   Optimization results

We are searching for a PAW with good qualities as indicated by the objectives S_a and S_s. The relevant portion of the objective space for the parameter sweep is shown in Figure 5.2. Each point represents one Ga PAW tested. Points towards the left of the plot have better scattering properties as indicated by a low value of S_a. Points towards the bottom of the plot have smaller errors in the equilibrium zb-GaN lattice parameter compared to all-electron results, as indicated by a low value of S_s. Therefore, an optimal PAW would appear at the bottom left of the plot. Optimal PAWs with both accurate S_a and S_s are evident in the objective space, meaning we have optimized a PAW with high accuracy according to our objectives.

We chose an optimal design point ($S_a = 0.027$, $S_s = 0.0037$) to test further. First, we manually inspect the logderivatives and arctangent curves of the optimal PAW, which are shown in Figure 5.3. From this figure it is clear that the PS logderivatives and arctan-

gent curves match the respective AE curves. Recall that the discontinuity in the d channel corresponds to the 3d bound state, and we do not want to capture that behavior in the PS logderivatives. The results in this figure show that accurate scattering properties are achievable within our PAW search space.
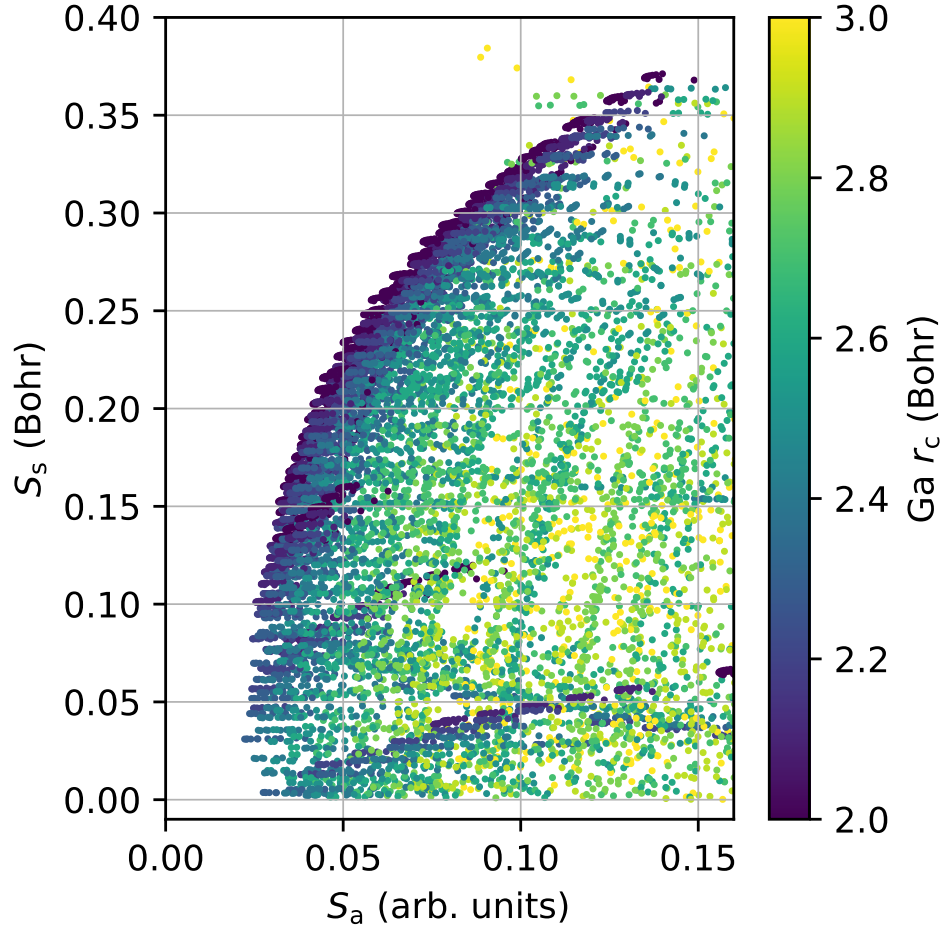
Figure 5.2: Objective space for parameter sweep. Points towards the bottom left of the plot are more optimal. Notice that a low arctangent objective (S_a) predicts a low lattice constant error (S_s) in the solid.

An interesting finding is that our objective S_a for the scattering properties also predicts accurate lattice parameter in the solid, as evident in Figure 5.2. This result supports our findings in Chapter 4, where the arctangent metric predicted accuracy in the solid for silicon, sodium, and molybdenum. To compare the arctangent metric to the logderivative metric, we also calculated S_L as defined in Equation 5.2. The objective space using S_L is
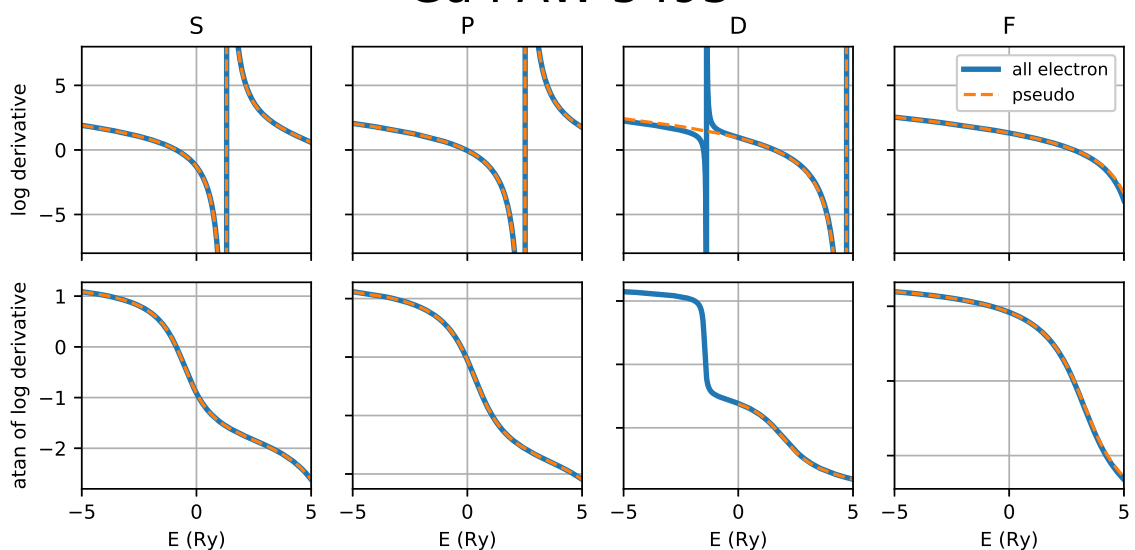
Figure 5.3: The first row shows the PS and AE logderivatives for each angular momentum channel. The second row shows the arctangents of the log derivatives for each angular momentum channel. Notice the discontinuity in the logderivatives in the D channel which corresponds to the 3d bound state. With the exception of this discontinuity (which we ignore due to pseudizing the 3d electrons), the agreement between log derivatives is excellent in the full energy range, and the atan curves capture this agreement.
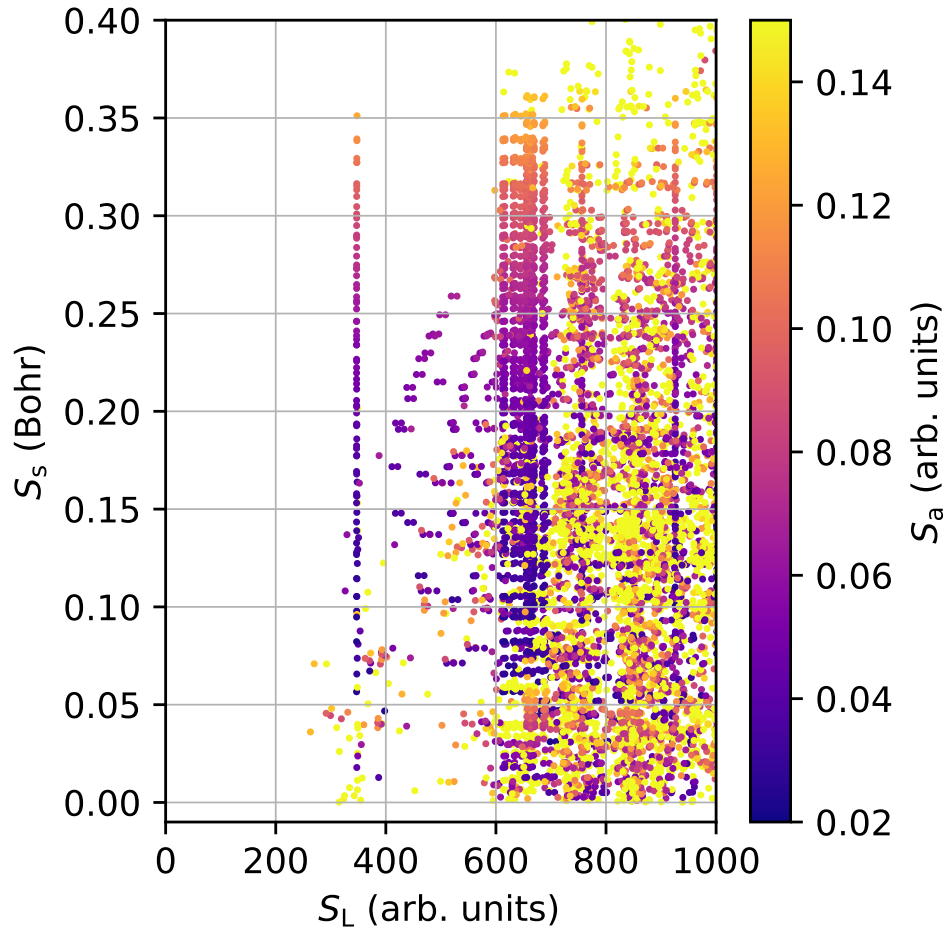
Figure 5.4: Objective space for the parameter sweep using S_L instead of S_a. Points towards the bottom left of the plot are more optimal. Unlike the arctangent objective, a low logderivative objective does not predict accurate lattice constants. Compare to Figure 5.2.

shown in Figure 5.4. It is evident that the logderivative does not predict lattice constant accuracy as the arctangent does.

We have shown that an accurate PAW exists in this search space. As a final proof of concept for PAW optimization, we show that a genetic algorithm can find optimal PAWs with fewer iterations than the parameter sweep. Thus using a multi-objective genetic algorithm (MOGA) will enable searches of larger parameter spaces in the future which would not be tractable with parameter sweeps. Figure 5.5 shows the objective space for an optimization of the GaN PAW using a genetic algorithm. In this case, optimal points are discovered as in the parameter sweep, but the number of required design points is much lower. This is
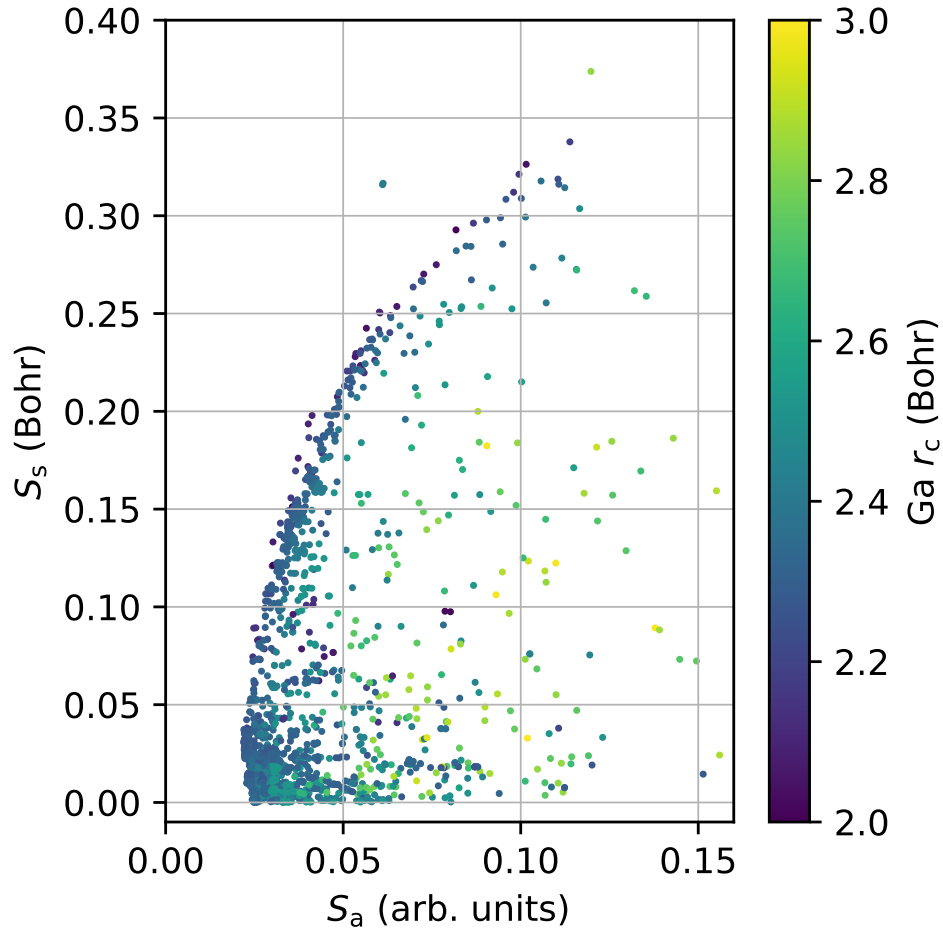
Figure 5.5: Objective space for the optimization using a genetic algorithm. Compared to the parameter sweep (Figure 5.2), the genetic algorithm found optimal PAWs with a much sparser search space.

evident visually by noting the non-optimal region of the objective space is much sparser using a genetic algorithm vs. a parameter sweep (compare Figure 5.5 with Figure 5.2). A total of 2500 PAWs were tested in the MOGA test, but only 1052 were necessary to find a point that dominated PAW 5493 from the parameter sweep. This means the number of evaluations required has been reduced by an order of magnitude compared to the parameter sweep.

5.3.2 Defect calculations using optimized PAW

Given an optimized PAW according to our optimization objectives, we want to validate this PAW in defect calculations. These tests will confirm that the Ga PAW optimized to give the correct lattice constant in zb-GaN is transferable to the defect calculations in w-GaN we are interested in, and that freezing the 3d electrons does not significantly affect our defect calculation results.

First, the calculated w-GaN Kohn-Sham band gap is increased by 0.36 eV using the 3-electron PAW as compared to the 13-electron PAW (see Figure 5.6, for example). This is possibly due the fictitious gallium 3d and nitrogen 2s interaction present when treating the 3d electrons as valence electrons, which was mentioned in Section 5.1. By freezing the 3d electrons we can avoid this spurious interaction, thus correcting the band gap by some amount. Although it's true that improvement of the band gap is somewhat of a fluke, that does not change the fact that a fictitious interaction has been avoided by freezing the 3d electrons. In both cases, including or omitting the 3d electrons from the valence, an approximation is being made. In the former case, the spurious repulsion of the VBE upward is included, and in the latter, the other interactions of the 3d electrons during bonding are neglected.

The calculated defect transition levels are shown in Figures 5.6, 5.8, and 5.7 for a gallium vacancy, and two different gallium interstitial sites. The structures for the defects are shown in Figure 5.9. It is unknown to what extent the difference in band gap is due to the valence band shifting downward vs. the conduction band shifting upward. Therefore, it does not make sense to align the levels by the valence band edges or the conduction band edges. We have therefore shifted the energies for the optimized 3-electron PAW downward by 1/2 the difference in band gaps to enable better comparison of the defect levels. This shift is somewhat arbitrary, but our conclusions are insensitive to the choice of shift. Qualitatively, the defect transition levels are similar for all defects calculated. For every defect, both PAWs predict the same number of defect transition levels within the

bounds. One possible exception is the +5/+6 level for the tetrahedral site. However, since this level for both PAWs is very near the bound, these can still be considered to agree. Furthermore, both PAWs predict the negative-U behavior for the tetrahedral site, indicated by the +1/+2 level being lower than the +2/+3 level, which means this defect can trap two holes with the second hole bound more strongly than the first.[51] This behavior has also been reported in Reference 52. Quantitatively, all defect levels agree within 0.25 eV.

5.3.3 Speedup achieved with optimized PAW

Given that we have found optimal PAWs using our search, we want to know if they afford any speedup compared to a 13-electron PAW. We expect they will, due to fewer electrons being included in the valence. However, the 3-electron PAW requires more additional projectors, which adds computing time to calculations. We ran timing calculations as described in Section 5.2.3 for two different supercell sizes, 72 and 300 atoms. In each supercell, one Ga atom is removed to create the Ga vacancy. We see a speedup using the 3-electron PAW for both supercell sizes. The speedup for both cell sizes is 30-40%, and is greater for the larger cell.
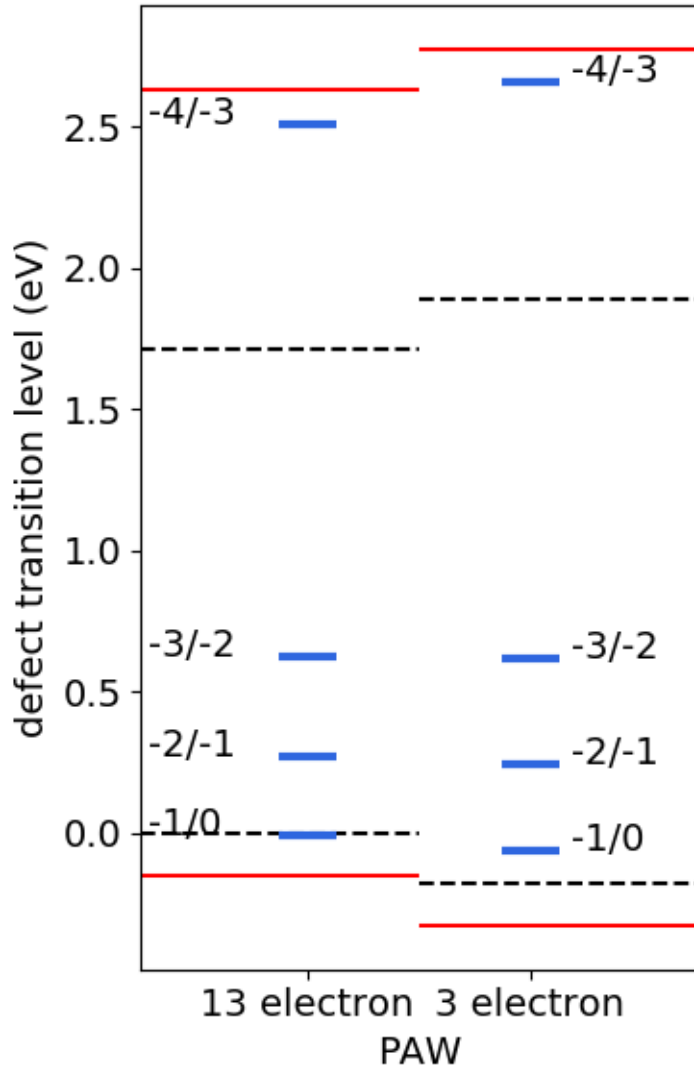
Figure 5.6: Comparison of defect levels between PAWs for the gallium vacancy. The energies for the Abinit PAW are referenced to the valence band edge. The energies for the 3-electron PAW so that the difference in the band gap is split between the valence and conduction bands. The dashed black lines are the band edges and the solid red lines are the bounds.
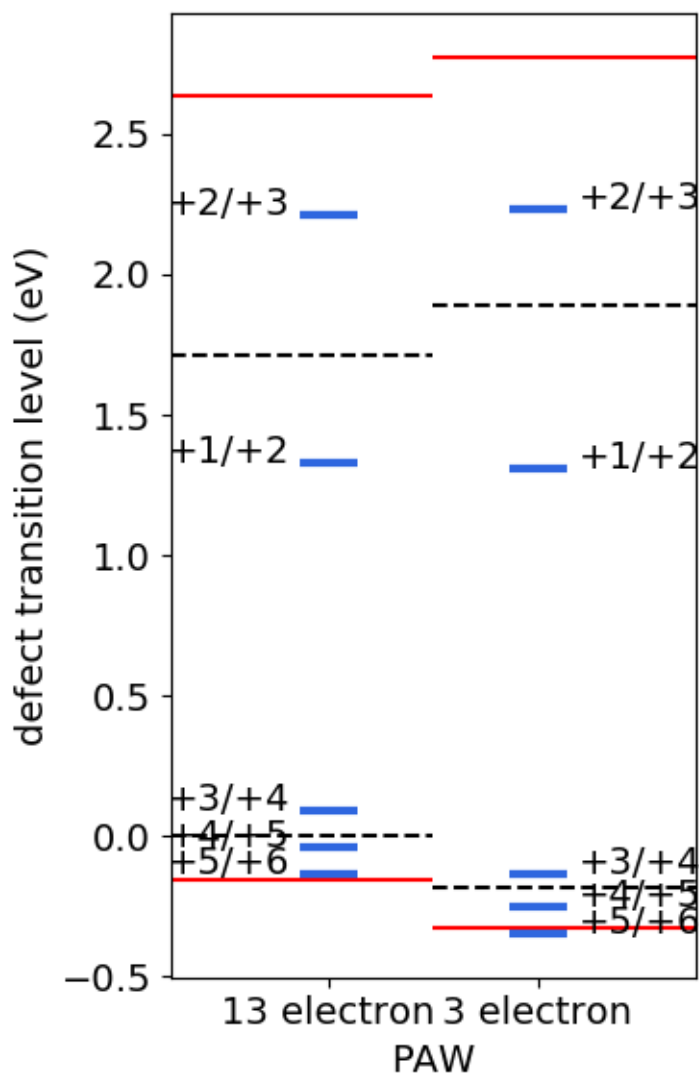
Figure 5.7: Comparison of defect levels between PAWs for the gallium interstitial in the tetrahedral site. The energies for the Abinit PAW are referenced to the valence band edge. The energies for the 3-electron PAW so that the difference in the band gap is split between the valence and conduction bands. The dashed black lines are the band edges and the solid red lines are the bounds.
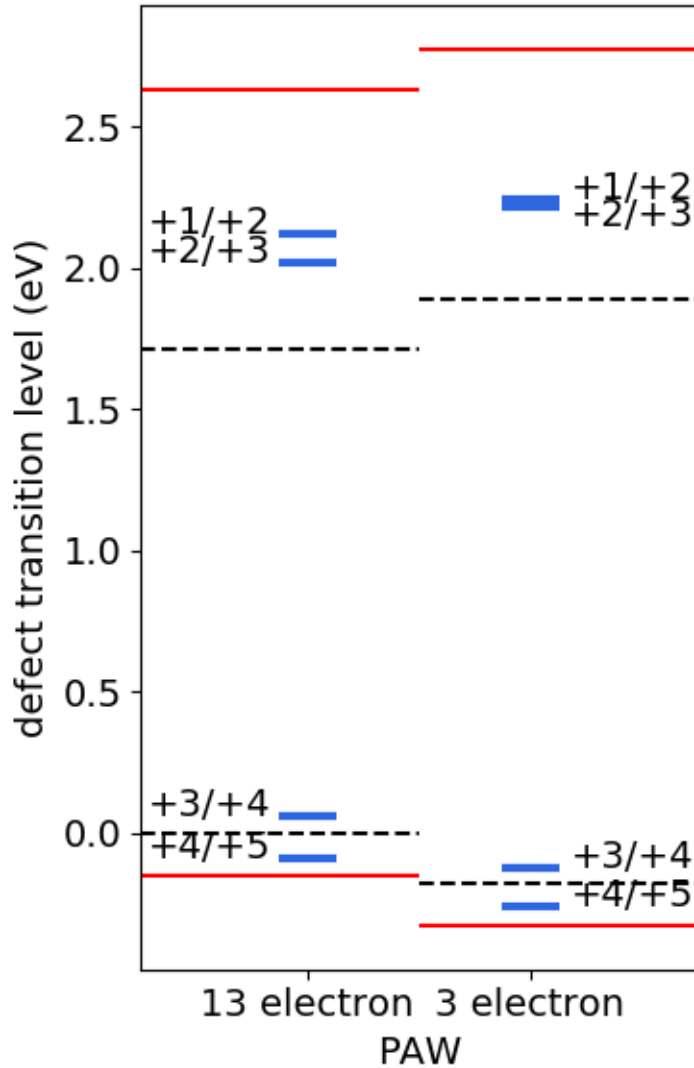
Figure 5.8: Comparison of defect levels between PAWs for the gallium interstitial in the octahedral site. The energies for the Abinit PAW are referenced to the valence band edge. The energies for the 3-electron PAW so that the difference in the band gap is split between the valence and conduction bands. The dashed black lines are the band edges and the solid red lines are the bounds.
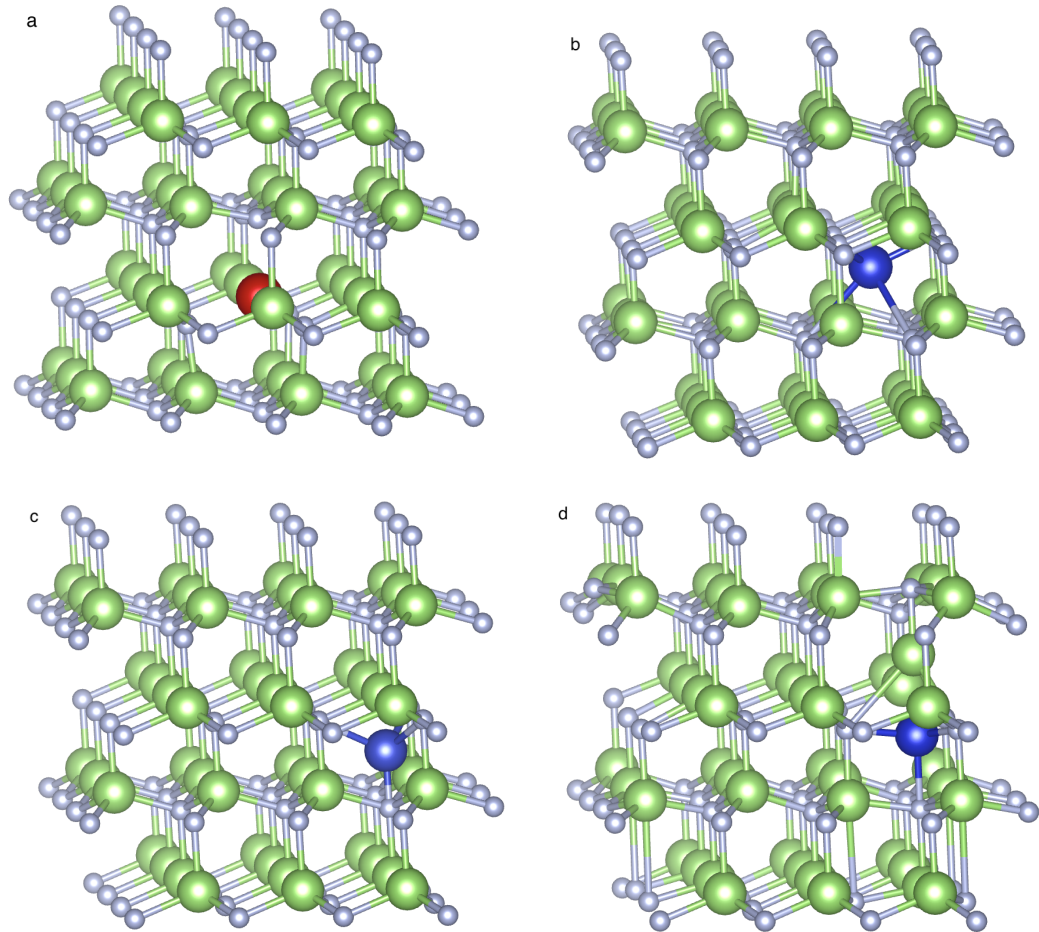
Figure 5.9: Structures for the gallium vacancy (a) and the gallium interstitial (b-d) in w-GaN. The larger spheres are gallium and the smaller spheres are nitrogen. For the gallium vacancy, the location of the missing gallium is indicated with the red sphere. For the interstitials, the interstitial gallium is shown with the blue sphere. The octahedral site for the gallium interstitial is shown in (b). The initial guess for the location of the gallium interstitial in the tetrahedral site is shown in (c). The location of the gallium interstitial in the tetrahedral site after relaxation is shown in (d).

Chapter 6

Conclusions and Outlook

We have developed methods for automatically optimizing pseudopotentials for speed and accuracy improvements in plane wave DFT calculations. In Chapter 3, we used a multi-objective genetic algorithm to optimize pseudopotentials for interatomic force accuracy and computational work. Our method produced Pareto sets of pseudopotentials, which offer varying compromises between accuracy and efficiency. We have seen that our force-based optimization can produce pseudopotentials of comparable accuracy compared to pseudopotentials from the literature when predicting forces or equilibrium lattice constant and bulk modulus, while improving the computational efficiency of the pseudopotentials. To enable and improve our optimizations, we developed new pseudopotential metrics, which we have compared to existing metrics. As demonstrated in Chapters 4 and 5, our arctangent metric for scattering properties is the most effective metric for pseudopotential scattering properties currently formulated, and is key to enabling efficient pseudopotential optimizations. Finally, we leveraged our optimization techniques and metrics to enable tractable studies of point defects in GaN.

Currently the predominant resource for DFT users in need of pseudopotentials is pseudopotential libraries. The long term vision of our research is to 1) improve and expand on these current libraries and 2) offer a method to automatically generate new pseudopotentials when libraries are not sufficient. This work has demonstrated that new pseudopotentials can offer speed and accuracy advantages over those currently available in these libraries. The reduction in force errors for LiF in Chapter 3, for example, was from about 7% to about 2%. Perhaps the more important finding is that computational work represents an opportunity for significant improvement of pseudopotentials. For SiGe, we were able to find pseudopotentials with accuracies comparable to the reference pseudopotentials in

the literature, but faster by a factor of three according to our computational work objective. This improvement in speed was obtained by tuning only the radii and projector matching energies and required no reduction of the number of electrons in the valence. We have also demonstrated significant speedup in a Ga PAW by reducing the number of electrons in the valence. In this case, although we did vary radii and projector matching energies, we did not attempt to find the most efficient Ga PAW within our search space (in other words, we didn't search for the fastest of the fast).

Because of these results, further speed gains for pseudopotentials across the periodic table are likely in the future. These gains will come both from optimizing the pseudopotential parameters and from reducing the transferability of the pseudopotentials in anticipation of use in specific materials or regimes.

A tradeoff between accuracy and efficiency is possible by leveraging reductions in transferability for speed gains (i.e. reducing the relevant materials space or applicable physical regimes). In contrast, recent libraries have been designed primarily with transferability in mind. Speed gains are likely possible in comparison to those libraries given the transferable nature of those pseudopotentials, whereas we can improve speed gains by neglecting transferability. The advantage of a selection of pseudopotentials is the availability of appropriate pseudopotentials optimal for a user's problem.

This tradeoff in accuracy and speed encourages a much more diverse and versatile pseudopotential library. Although it is possible that libraries could eventually mature to supply needs of every researcher, especially with the recent surge in new library development, much more development is still needed. Having a large selection of pseudopotentials makes calculations more accessible by preventing situations where a non-expert user has to generate their own pseudopotential appropriate pseudopotentials aren't available.

One could imagine a new paradigm of pseudopotential generation where pseudopotentials are generated as needed to be optimal for a specific problem. This approach could be beneficial for longer term studies of a small class of materials or structures, where the time

investment for optimizing a new pseudopotential is offset by the eventual time savings, or when strict accuracy is needed. The material-specific optimization allows for generation of pseudopotentials that work well for the chosen system without needing to worry about transferability to unrelated materials, allowing softer pseudopotentials. Our technique suggests that pseudopotentials can be generated as needed for users' specific applications to achieve gains in computational efficiency. It would be especially beneficial for long term studies of a specific material or related materials, or when speed is critical.

One can even imagine a situation where multiple pseudopotentials are used for the same element in different parts of the same system. For example, a silicon pseudopotential optimized for bulk silicon could be used far away from an interface while a different pseudopotential optimized to interact with the interface material could be used for silicon atoms closer to the interface.

Automated optimization is a clear route towards smarter generation of new pseudopotentials, offering a more thorough and intelligent search of the design space than would be allowed with traditional hand-tuning methods. Because more pseudopotentials are tested, the probability of finding a high-quality pseudopotential is increased. Additionally, the optimization reduces the need for human time and expertise. We demonstrated the effectiveness of automated optimizations in Chapters 3 and 5, and other recent works have seen positive results using pseudopotential optimizations as well.[10,11] In Chapter 5 we compared a genetic algorithm to a parameter sweep for the same search space of PAWs and found that the genetic algorithm required an order of magnitude fewer calculations, which justifies our use of genetic algorithms earlier in Chapter 3.

In this work, we have not focused on maximizing the efficiency of the optimization methods, instead focusing primarily on the metrics and methods for distinguishing a good pseudopotential from a bad one using metrics vs. human inspection. Therefore, improvements can be made to our optimization methods, including several immediate ones. First, the scattering properties of each angular momentum channel are independent of the projec-

tor matching energies of other angular momentum channels. This reduces the search space significantly, which we have not leveraged. Second, the parameters governing the genetic algorithm can be tuned for this type of problem. We have used default values for the genetic operations (crossover, mutation), and what we considered to be reasonable values for population sizes. However, it's possible significant gains could be made by choosing these parameters more rigorously.

A major contribution to the effectiveness of automated optimizations from this work is the exploration and evaluation of the pseudopotential quality metrics, which enable optimizations. These, plus contributions from other groups,[18,12,11] will enable much more effective optimizations. In each chapter we have used various metrics for the accuracy of pseudopotentials. For scattering accuracy in the atom, we have used our arctangent metric, which we have compared to a metric based on the logderivatives. For accuracy in the solid, we have used the lattice constant (Chapters 3 and 5), the Δ-factor (Chapter 4), and interatomic forces (Chapter 3). We can now draw some conclusions about the advantages and disadvantages of each metric.

The lattice constant and the Δ-factor are both derived from similar energy-volume calculations. Thus we could expect them to be correlated to some extent, excluding cases where the bulk modulus is extremely wrong. In fact, we have seen that they provide similar information, and this has been independently confirmed in Reference 12. Furthermore, the interatomic forces, can also be expected to provide the same information as the lattice constant or the Δ-factor, since they are related to the energies by the Hellman-Feynman theorem, provided the atomic positions are sufficiently near equilibrium. All other things being equal, there is an advantage to the Δ-factor in that it also captures energy-volume curvature disagreement while lattice constant does not. The advantage of the Δ-factor over a force-based metric is that it is much simpler to implement. Additionally, the reporting of the Δ-factor is now common for new pseudopotentials, thus providing a standard that can be used to compare all pseudopotentials.

To evaluate pseudopotential scattering properties, we have developed the arctangent metric, S_A. This metric is related to the logderivatives but can be used as a metric for automated evaluation. We have demonstrated that this metric can be used to reliably capture ghost states without resorting to calculations in the solid such as lattice constants, density of states, or band structures. Additionally, it provides a meaningful representation of the overall quality of scattering properties in pseudopotentials and PAWs. Thus S_A is key to enabling efficient pseudopotential optimizations. We describe below possible applications of this metric by the pseudopotential communities.

For those designing pseudopotentials by hand, using S_A can streamline the process by easily screening out pseudopotentials with poor logderivative agreement or ghost states without requiring visual inspection. It also requires no calculations in the solid using the pseudopotential as would be necessary for other methods of detecting ghost states such as band structure calculations.

For those designing pseudopotentials using automated or semi-automated optimization techniques, the metric has two advantages. First, as a metric alone it has the ability to reliably guide optimizations. An optimization will only perform correctly if the metrics used as objectives are meaningful, so that the value of the metrics truly indicate the quality of the pseudopotentials, which is the case for the arctangent metric. The second advantage is that S_A can reduce optimization time when used as a screening metric by reducing the number of necessary calculations in the solid.

Furthermore, in anticipation of users performing calculations in different physical regimes, S_A could be defined for different energy ranges (e.g. -5 to +10 Ry vs. -5 to +5 Ry). A version with a higher energy limit could be useful for ensuring accuracy in excited state calculations, for example. In problems where speed is important and scattering agreement at high energies is not critical, a lower energy range upper limit would be sufficient. This approach of tailoring the energy range to the specific application is reminiscent of an approach used by Sarkar et al. in which they modify the pressure range of the Δ-factor to account for high

95

energy density calculations of materials at high pressures.

For users of plane wave DFT, the S_A metric gives an immediate and easy-to-interpret indicator of the quality of a pseudopotential's scattering properties. Reporting this metric to the community, similar to the way the Δ-factor is reported, would give valuable information to users. Currently, it is the responsibility of the designers of pseudopotentials to ensure there are no ghost states in their libraries, or the responsibility of users to carefully check the pseudopotential scattering properties. And as we have seen, good Δ-factor does not necessarily imply good scattering properties. Using a full test set of structures when calculating the Δ-factor, as is done in other works, may improve the predictive power of the Δ-factor, but the improvement depends on the completeness of the test set of structures, and increasing the test set increases the expense of the Δ-factor calculations. Additionally, Prandini et al.[12] have urged the use of a wider array of tests in the solid, pointing out that Δ alone does not necessarily correlate to the accuracy of other metrics such as those based on phonons, band structures, and cohesive energies. The inclusion of further tests as suggested would even further increase the expense of tests in the solid, and their findings support our claim that Δ alone does not represent the true accuracy of a pseudopotential. Therefore, S_A stands out as an obvious choice for a new pseudopotential metric, providing a valuable supplement to the information given by the Δ-factor and other metrics in the solid.

In conclusion, we summarize the key takeaways of this research. First, new pseudopotentials will be faster and more accurate than currently existing pseudopotentials, and can offer plane wave DFT users a user-selectable tradeoff between accuracy and efficiency. Next, the development of these new pseudopotentials will benefit from automated optimization techniques, especially genetic algorithms. Finally, the arctangent metric we have developed will enable these optimizations and provide a straightforward and robust method for evaluating scattering properties of pseudopotentials. This metric reliably captures pseudopotential ghost states with negligible computational expense and could provide a standardized scattering property metric for the plane wave DFT community.

BIBLIOGRAPHY

[1] Marc Torrent and Natalie A. W. Holzwarth. A users guide for *atompaw* code. http://users.wfu.edu/natalie/papers/pwpaw/atompaw-usersguide-MarcTorrent.pdf.

[2] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[3] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 1965.

[4] James C Phillips and Leonard Kleinman. New method for calculating wave functions in crystals and molecules. *Physical Review*, 116(2):287, 1959.

[5] DR Hamann, M Schlüter, and C Chiang. Norm-conserving pseudopotentials. *Physical Review Letters*, 43(20):1494, 1979.

[6] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Physical Review B*, 41(11):7892, 1990.

[7] Peter E Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953, 1994.

[8] Pseudopotentials supplied with the VASP package. https://cms.mpi.univie.ac.at/vasp/vasp/Pseudopotentials_supplied_with_VASP_package.html.

[9] François Jollet, Marc Torrent, and Natalie Holzwarth. Generation of Projector Augmented-Wave atomic data: A 71 element validated table in the XML format. *Computer Physics Communications*, 185(4):1246–1254, April 2014.

[10] Martin Schlipf and François Gygi. Optimization algorithm for the generation of ONCV pseudopotentials. *Computer Physics Communications*, 196:36–44, 2015.

[11] Kanchan Sarkar, Mehmet Topsakal, NAW Holzwarth, and Renata M Wentzcovitch. Evolutionary optimization of paw data-sets for accurate high pressure simulations. *Journal of Computational Physics*, 347:39–55, 2017.

[12] Gianluca Prandini, Antimo Marrazzo, Ivano E Castelli, Nicolas Mounet, and Nicola Marzari. Precision and efficiency in solid-state pseudopotential calculations. *Computational Materials*, 4(1):72, 2018.

[13] Richard M Martin. *Electronic structure: basic theory and practical methods*. Cambridge university press, 2004.

[14] Kevin F. Garrity, Joseph W. Bennett, Karin M. Rabe, and David Vanderbilt. Pseudopotentials for high-throughput DFT calculations. *Computational Materials Science*, 81:446–452, May 2013.

[15] N.A.W. Holzwarth, A.R. Tackett, and G.E. Matthews. A projector augmented wave (PAW) code for electronic structure calculations, Part I: atompaw for generating atom-centered functions. *Computer Physics Communications*, 135(3):329–347, April 2001.

[16] Paolo Giannozzi. Notes on pseudopotential generation. https://www.quantum-espresso.org/Doc/pseudo-gen.pdf.

[17] Pablo Rivero, Víctor Manuel García-Suárez, David Pereñiguez, Kainen Utt, Yurong Yang, Laurent Bellaiche, Kyungwha Park, Jaime Ferrer, and Salvador Barraza-Lopez. Systematic pseudopotentials from reference eigenvalue sets for dft calculations: Pseudopotential files. *Data in brief*, 3:21–23, 2015.

[18] Kurt Lejaeghere, Veronique Van Speybroeck, Guido Van Oost, and Stefaan Cottenier. Error estimates for solid-state density-functional theory predictions: an overview by means of the ground-state elemental crystals. *Critical Reviews in Solid State and Materials Sciences*, 39(1):1–24, 2014.

[19] Kurt Lejaeghere, Gustav Bihlmayer, Torbjörn Björkman, Peter Blaha, Stefan Blügel, Volker Blum, Damien Caliste, Ivano E Castelli, Stewart J Clark, Andrea Dal Corso, et al. Reproducibility in density functional theory calculations of solids. *Science*, 351(6280):aad3000, 2016.

[20] E Kucukbenli, M Monni, BI Adetunji, X Ge, GA Adebayo, N Marzari, S De Gironcoli, and A Dal Corso. Projector augmented-wave and all-electron calculations across the periodic table: a comparison of structural and energetic properties. *arXiv preprint arXiv:1404.3015*, 2014.

[21] R. Car and M. Parinello. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, 55(22):2471, 1985.

[22] P. Ehrenfest. Bemerkung über die angenäherte gültigkeit der klassischen mechanik innerhalb der quantenmechanik. *Zeitschrift für Physik A Hadrons and Nuclei*, 45(7–8):455–457, 1927.

[23] RA Hansel, Casey N Brock, Brandon C Paikoff, Alan R Tackett, and D Greg Walker. Automated generation of highly accurate, efficient and transferable pseudopotentials. *Computer Physics Communications*, 196:267–275, 2015.

[24] Brian M. Adams, Keith R. Dalbey, Michael S. Eldred, David M. Gay, Laura P. Swiler, William J. Bohnhoff, John P. Eddy, Karen Haskell, Patricia D. Hough, and Sophia Lefantzi. DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis, December 2009. Sandia Technical Report SAND2010-2183, Updated 21 January 2011.

[25] Socorro. http://dft.sandia.gov/socorro.

[26] Elk. http://elk.sourceforge.net/.

[27] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865, 1996.

[28] Miguel A. L. Marques, Micael J. T. Oliveira, and Tobias Burnus. Libxc: a library of exchange and correlation functionals for density functional theory. March 2012.

[29] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Physical Review B*, 13(12):5188, 1976.

[30] F. D. Murnaghan. The compressibility of media under extreme pressures. *Proceedings of the National Academy of Sciences of the United States of America*, 30(9):244–7, September 1944.

[31] PAW atomic datasets – ABINIT. http://www.abinit.org/downloads/PAW2.

[32] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47(1):558, 1993.

[33] Georg Kresse and Jürgen Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B*, 49(20):14251, 1994.

[34] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science*, 6(1):15–50, 1996.

[35] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.

[36] X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Ct, T. Deutsch, L. Genovese, Ph. Ghosez,

M. Giantomassi, S. Goedecker, D.R. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M.J.T. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M.J. Verstraete, G. Zerah, and J.W. Zwanziger. Abinit: First-principles approach to material and nanosystem properties. *Computer Physics Communications*, 180(12):2582 – 2615, 2009.

[37] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, et al. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter*, 21(39):395502, 2009.

[38] Casey N Brock, Brandon C Paikoff, Muhammad I Md Sallih, Alan R Tackett, and D Greg Walker. Force-based optimization of pseudopotentials for non-equilibrium configurations. *Computer Physics Communications*, 201:106–118, 2016.

[39] James R Chelikowsky and Steven G Louie. *Quantum theory of real materials*, volume 348. Springer Science & Business Media, 1996.

[40] DR Hamann. Optimized norm-conserving Vanderbilt pseudopotentials. *Physical Review B*, 88(8):085117, 2013.

[41] Ann E. Mattsson. A lithium projector augmented wave potential suitable for use in VASP at high compression and temperature. Technical Report SAND2012-7389, Sandia National Laboratories, Albuquerque, New Mexico (United States) and Livermore, California (United States), 2012.

[42] Ann E. Mattsson. Assessing the standard molybdenum projector augmented wave VASP potentials. Technical Report SAND2014-15971, Sandia National Laboratories, Albuquerque, New Mexico (United States) and Livermore, California (United States), 2014.

[43] Seth Root, Rudolph J Magyar, John H Carpenter, David L Hanson, and Thomas R Matts-

son. Shock compression of a fifth period element: Liquid xenon to 840 gpa. *Physical review letters*, 105(8):085501, 2010.

[44] Seth Root, Rudolph J Magyar, David Lester Hanson, Ann Elisabet Mattsson, Thomas Kjell Rene Mattsson, and Kyle Robert Cochrane. Shock Compression of Liquid Noble Gases to Several Mbar. Technical Report SAND2011-8008C, Sandia National Laboratories, Albuquerque, New Mexico (United States), 2011.

[45] Martin Kiffner, Dieter Jaksch, and Davide Ceresoli. A polynomial ansatz for norm-conserving pseudopotentials. *Journal of Physics: Condensed Matter*, 2018.

[46] Alan F Wright and JS Nelson. Explicit treatment of the gallium 3d electrons in gan using the plane-wave pseudopotential method. *Physical Review B*, 50(4):2159, 1994.

[47] S-H Jhi and J Ihm. Study of the electronic structure and the role of gallium 3d electrons in gallium nitride. *physica status solidi (b)*, 191(2):387–394, 1995.

[48] K Karch, F Bechstedt, and T Pletl. Lattice dynamics of gan: Effects of 3d electrons. *Physical Review B*, 56(7):3560, 1997.

[49] WRL Lambrecht, B Segall, St Strite, G Martin, A Agarwal, H Morkoc, and A Rockett. X-ray photoelectron spectroscopy and theory of the valence band and semicore ga 3d states in gan. *Physical Review B*, 50(19):14155, 1994.

[50] Normand A Modine, Alan F Wright, and Stephen Roger Lee. Bounds on the range of density-functional-theory point-defect levels in semiconductors and insulators. *Computational Materials Science*, 92:431–438, 2014.

[51] GD Watkins. Negative-u properties for point defects in silicon. *MRS Online Proceedings Library Archive*, 2, 1980.

[52] P Bogusławski, Emil L Briggs, Jerzy Bernholc, et al. Native defects in gallium nitride. *Physical Review B*, 51(23):17255, 1995.