Modeling Materials: Comparison of Two Projector-Augmented Wave (PAW) Datasets

A paper presented to the faculty of the Department of Physics of Wake Forest University in partial fulfillment of the requirements for graduation with Honors in Physics

James Lewis Drewery II

May 9, 2014

Approved by:

_____
Dr. Natalie Holzwarth

_____
Dr. Timo Thonhauser

_____
Dr. Freddie Salsbury Jr.

Table Of Contents

1. Introduction

1.1 Motivation: Why Model Materials

Materials science has become one of the most predominantly studied areas of physics today. Simulating atomic interactions may provide physicists with an experimental advantage. The use of computational modeling provides a quick and efficient means of analyzing properties of materials on the atomic level. Computational modeling of materials is an important tool in the development of new technologies. A potential focus of this technology is in energy storage. The simulation of new and innovative ways of designing batteries may lead to a cleaner and more efficient world. Although it is unlikely that perfect battery designs will be a result of this simulation process alone, it may provide experimentalists with a foundation as to what may in fact yield possible results.

1.2 Computational Requirements

The simulation of materials requires initial data input, or datasets, for each constituent atom. Each dataset contains all of the important information needed to represent the atom that figures in the simulation of materials containing that atom. Upon coming to work with Professor Holzwarth, the creation of these initial datasets was the main focus of my research. With the help of Marc Torrent's user guide for the ATOMPAW computer program, and the watchful eye of my mentor, I came to understand how the physics I had been exposed to until this point reflects in a legitimate research endeavor. Fresh out of PHY 215, the capabilities of applying Modern Physics to our research was amazing. Understanding quantum numbers, electron configurations and placement, atomic radii, and more were things that were at the heart of my research in the beginning. Cameron Kates and myself worked together to update an online database of these atomic inputs. Our process was systematic in nature. We hoped to find an approach that for any given element, we could create a particularly effective atomic input that would work with several materials simulation software packages.

1.3 Road to Thesis

Initially, my work with Dr. Holzwarth consisted solely in the creation of the aforementioned datasets, together with Cameron Kates, Zachary Pipkorn, and Chaochao Dun. This research encompassed my final semester Junior year and the following summer during the Wake Forest Research Fellowship. At the end of the summer, one Dr. Holzwarth's French collaborators, Francois Jollet, compiled a similar database of atomic inputs (Jollet, Torrent, and Holzwarth). Jollet had a very eclectic approach

in his dataset creation. It was in the testing and comparison of these two databases of atomic datasets that my thesis was based. I began testing the French datasets with the ones that we had created and found some very interesting conclusions. I was successful in the comparison of many compound materials. From these comparisons, I was able to draw several conclusions as to what parameters were most likely more pertinent in the atomic simulations. Additionally, a hypothesis was made about which method was more effective in creating these inputs. I found that with a more liberal choice of valence state electrons, versus core state electrons, that our results were generally more similar to what we took as our standard. In choosing more valence state electrons, we found a general trend in that our datasets were better than Jollet's, but at a price. The computational cost of including more valence states was something that Jollet was able to work around and still yield results very to ours, and at times better. Also, his process of using many formats and approaches to build his database showed striking contrast to our systematic approach and raises questions as to which is in fact a better method. Ideally, a universally accepted approach would greatly benefit computational material science research. As a possible step toward this, I offer a detailed analysis of the differences found between our dataset results and that of Jollet's.

2. Computational Methods

2.1 Basic Formalism

The starting point of the simulations is the self-consistent electronic structure of an atom within the framework of Density Functional Theory (DFT) (Hohenberg and Kohn). The idea is that we can model an atom by incorporating First Principles calculations. This means that we are using only established laws of physics and we are not taking into account any outside or predetermined experimental data. First Principles simulations of materials are an important tool for basic research and the development of new techniques. Additionally, because we will want to model elements that will actually have many electrons, we incorporate Mean-Field approximations and also Density Functional Theory (DFT).

With this First Principles basis and use of DFT for our calculations, we are able to illustrate what how the wave function will behave for the all-electron case. To illustrate this, first recall the Schrödinger Equation,

$$Equation\ 1: H\psi = E\psi$$

and that the Hamiltonian represents the kinetic energy plus the potential energy. Here, the Hamiltonian of the atom is defined such that,

$$Equation\ 2: H(r)_{atom} = -\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{r} + V_{ee}(r) + V_{XC}\big(\rho(r)\big).$$

The first term represent the kinetic energy of the electron, the second term represents the Coulomb interaction of an electron with the nucleus of charge Z in cgs units, $V_{ee}(r)$ is the electron-electron potential, and $V_{XC}(r)$ is the exchange correlation due to the DFT calculation. This is the effective Hamiltonian that acts on each electron. With respect to DFT, we assume a local-density approximation (LDA) for our calculations (Perdew and Wang). Mean-Field Theory and DFT are figure prominently in our research as they provide the basis for our calculations. This is simply how our software code is designed to find the All-Electron wave functions for which we are trying to more smoothly approximate with our Pseudowave functions. The main take away from these two theories is rather simple. For Mean-Field theory, we suppose that each electron sees an average of all the others. As such, we replace the individual electrons instead by an electron charge density.

2.1.1 Pseudopotentials and PAW

The All-Electron case takes into account both core and valence state electrons. However, in materials modeling, the important contributions come from the valence electrons only. Pseudopotentials were invented to further approximate the valence electrons so that they may be numerically well represented in solid simulations. The Projector Augmented Wave scheme (PAW) was developed by Peter Bloechl is a more accurate and efficient Pseudopotential method (Blöchl). We need to create the PAW dataset to represent the valence electron properties for each atom.

2.2 Software

2.2.1 ATOMPAW

ATOMPAW is a computer program that we used in the creation of atomic projector and basis functions for our electronic structure calculations (Holzwarth, Tackett, and Matthews). This code incorporates the PAW method as previously mentioned. ATOMPAW is the first code which we run to create our datasets. The resultant output from running the code yields the datasets that are used to model the solid materials. These datasets include the All-Electron and Pseudo basis functions and

corresponding Projector functions. In addition, the datasets include Pseudopotential functions and matrix elements for each atom.

2.2.2 ABINIT

ABINIT is a program that allows us to use our Pseudo functions to test for the total energy, charge density, and electronic structure of our electronic structure (Gonze et al.). After running ATOMPAW, we take our "ideal" cases and we can test them by running them on the ABINIT software. The main point of this program is to check that ATOMPAW outputs or datasets can reliably describe the properties of materials made of those atoms. In particular, our main focus was to compute the Binding Energy versus lattice constant for simple binary compounds and elemental solids with an LAPW standard.

2.2.3 WIEN2k

The WIEN2k plot is derived using a Linearized Plane Wave method, or LAPW (Schwarz and Blaha). This is another technique for generating Binding Energy vs. Lattice constant curves. The WIEN2k software is what we consider to be our "standard" by which we compare both our results and those of the generated datasets by Jollet. We take this to be a standard because it is another process for finding the exact same results, a Binding Energy vs. Lattice Constant plot, using a method completely independent of the use of Pseudopotentials. My main source of research of the Pseudopotential approach came from reading *Electronic Structure: Basic Theory and Practical Methods* (Martin).

2.2.4 Brief Summary of Code Usage

The use of the two previously described codes is to find the best suited Pseudopotential wave function that smoothly approximates the All-Electron case. In the dataset generation phase, one is yielded with many results given the range of input parameters. Taking this range and finding the case with minimum error between the Pseudo case and the All-Electron case yields the choice of input parameters for ABINIT. Running this final code yields a Binding Energy vs. Lattice Constant curve that we can compare to our standard which is the Wien2k, or LAPW, plot.

2.2.5 Binding Energy and Lattice Constant

The result that will be of most importance to our analysis will be that of the Binding Energy vs. Lattice Constant plot. It is important to describe what each of these parameters represents. The

minimum of the Binding Energy curve occurs at the equilibrium lattice constant and the curvature represents the bulk modulus ("Binding Energy (physics) -- Britannica Online Encyclopedia"). The Lattice Constant is a degree of proximity between the atoms in the unit cell of our calculations. We wanted to test our datasets to see if they accurately modeled the near-equilibrium properties of materials containing the atoms in question. Our process was to make these graphs and see if the ABINIT and WIEN2k curves visually coincided. We wanted to keep the calculations simple so we chose several elemental materials and binary compounds for the comparison. We also had to specify the structure of the cell for each compound before we ran ABINIT and WIEN2k. The structures mostly consisted of simple cubic, Body-Centered Cubic, Face-Centered Cubic, Zinc-blende, and Diamond orientations (given in order by Figure 1).
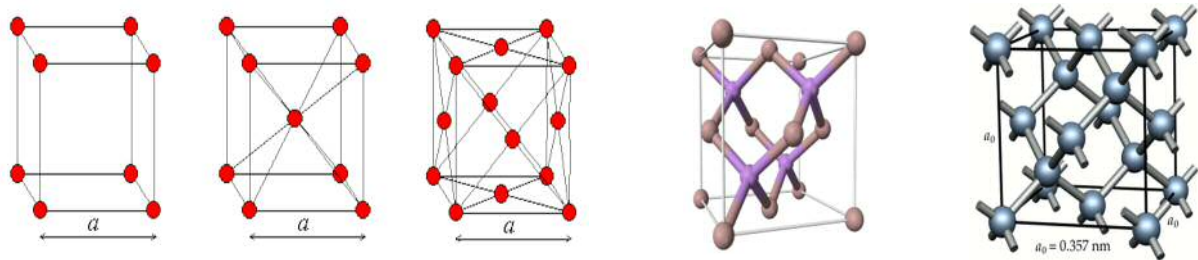

Figure 1: Showing possible cell structures in which we incorporated into our calculations (SC, BCC, FCC, ZB, D (Van Zeghbroeck, B.)

3. Thesis Components

It is very important that a clear distinction is made on the comparison between the Jollet's datasets and ours. Jollet incorporated many schemes and approaches in his database development. This is in contradiction of our adherence to a strict process by which we created our datasets. The differences and similarities to follow may provide insight as to the efficiency of either approach. Additionally, it is noteworthy to mention just how many parameters there are to consider in comparing these datasets. It is not simply "change parameters X and Y to see how Z is affected." It would be more appropriately to describe it such that one is to "change parameters A, B, C... and Y to see how Z is affected." There were many parameters that I handled differently than did Jollet and it is with analyzing these distinctions that may provide insight as to what is the better method. The parameters involved were choice of pseudization schemes, matching radii, valence state choices, basis function energies, and others.

3.1 Atomic Configuration

A primary component of the dataset is in regards to the distribution of the electrons of the atomic element. At times, the delegation of electrons was quite easy. We could simply follow the choice of electron configuration given by the periodic table and this would yield really good results with smooth wavefunctions. However, many times this approach would fail and we had to be a bit clever with the electron states. At times, we could add in additional partially occupied, or even unoccupied, electron states. Doing this would give an additional valence electron state and a bit more flexibility when looking for bound states of the Schrödinger equation. An instance where such a technique is employed is shown in Figure 2. As illustrated there were at times vast differences in the choice of electron distribution. To make sense of this, the description of Jollet's Barium dataset may prove beneficial. There are 56 electrons total. The first line of the first box gives the maximum of each quantum number for the material. The dataset reads max S is 6, max P is 5, and max D is 5. The next two lines below actually tell how many electrons are allotted to that specific quantum number. In essence, this tells the program the cutoff of the number of electrons present. Here, Jollet has chosen for the 5D state to have 2 electrons and for the 6S state to have zero electrons. This choice of parameter is very different from what we chose on the right, as we stuck to a more basic and straightforward assortment of the electrons. I predict that his choice of adding in the unoccupied 6S state introduced a degree of flexibility that was not present in ours. As illustrated, the choice of electron configuration we have from the summer adheres strictly to that of the periodic table. However, Jollet's file puts 2 electrons in the 5D state and uses an unoccupied 6S state. In doing so, he introduces 2 additional valence electron states. Upon comparing the Binding Energy vs. Lattice Constant curves, it is obvious that the French dataset is in fact a better choice. This conclusion is supported by the fact that the choice of matching radius for the pseudofunction to the all-electron function was very close together with a difference of only 0.2 atomic units. It is of mention to note that our compound BaS, where the Sulfur atomic datasets were identical, was one of the only instances in which Jollet's results were vastly better than our own. I also would mention that I did not personally have a hand in our Barium dataset creation. However, it does illustrate the difference in electron placement while explicitly representing the effect of allowing additional valence states.
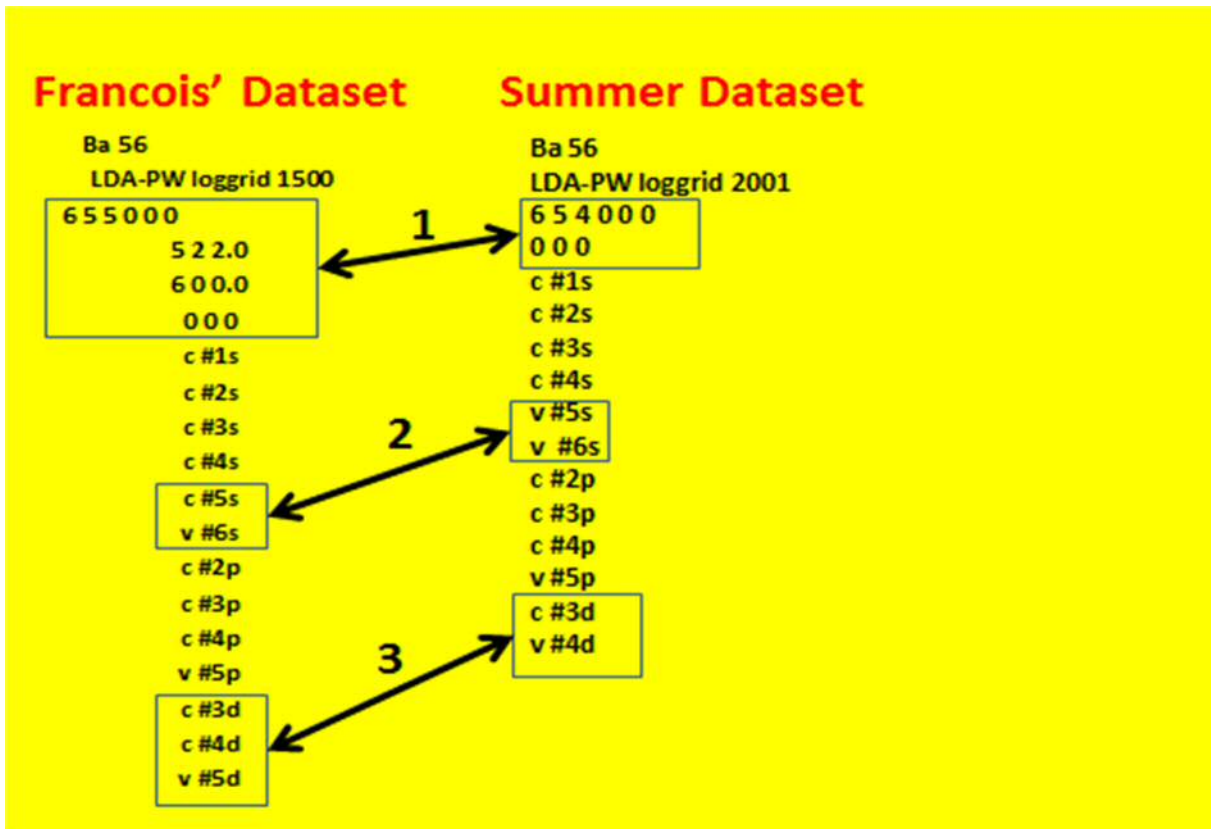
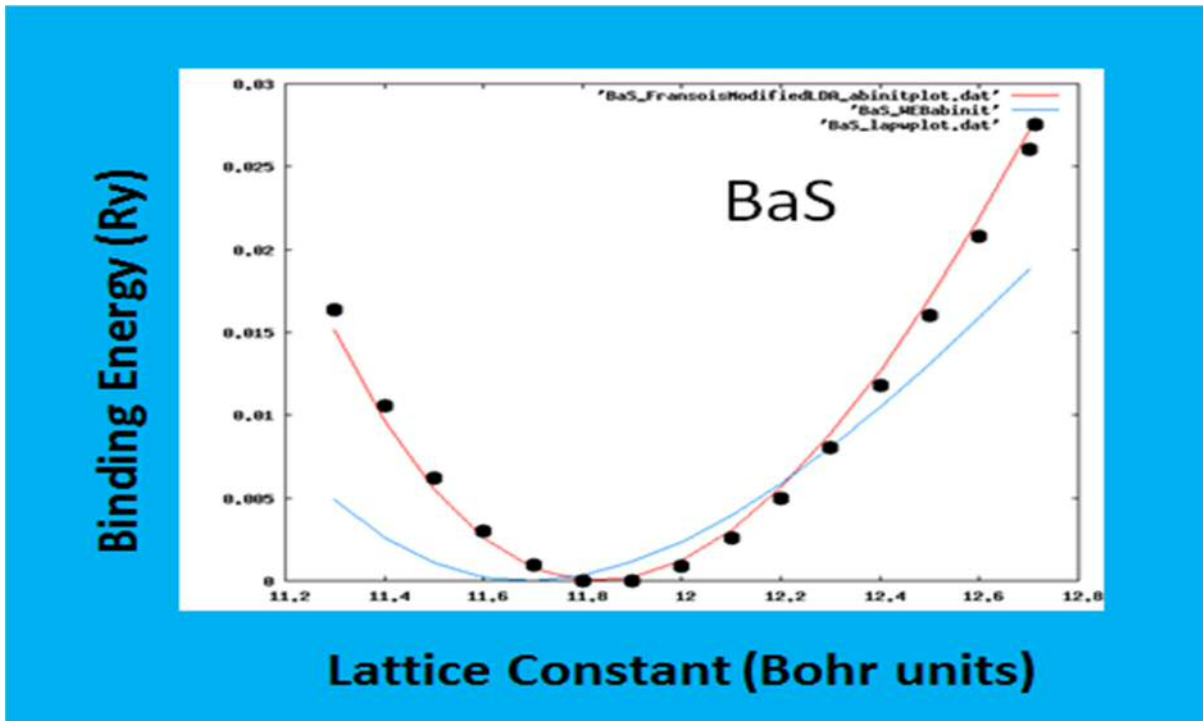Figure 2: Showing differences of electron distribution in either dataset.

Figure 3: Binding Energy vs. Lattice Constant curve for BaS.

In Figure 3, we show the Binding Energy curve from Jollet's dataset compared with LAPW and our summer result. It is worth noting that Jollet's more successful dataset did not use the 4D state as valence, but included a bound 5D state. This may be responsible for the better result. With the example of Jollet's dataset, we see how to improve our own initial dataset. As such, we chose not to publish the data on the web. However, the fact that Jollet's dataset very accurately represents the WIEN2k plot gives us something to study in future endeavors.

3.2 Matching Radius Selection (along with Electron Distribution)

It may be the case that our choice of matching radius between the Pseudofunction and the All-Electron function may play a role in the final result. Initially, it is at times difficult to determine a good choice for the initial radius of the atom with which we are concerned. To get a relative idea, we consulted a copy of *Quantum Theory and Molecular Solids* to find approximate values for the radius value (Slater). Again, this just gets us a range to work with and the true value for the matching radius should be determined from refining the ATOMPAW outputs. The choice of the radius value may have important consequences when comparing our files with those of our French collaborators.
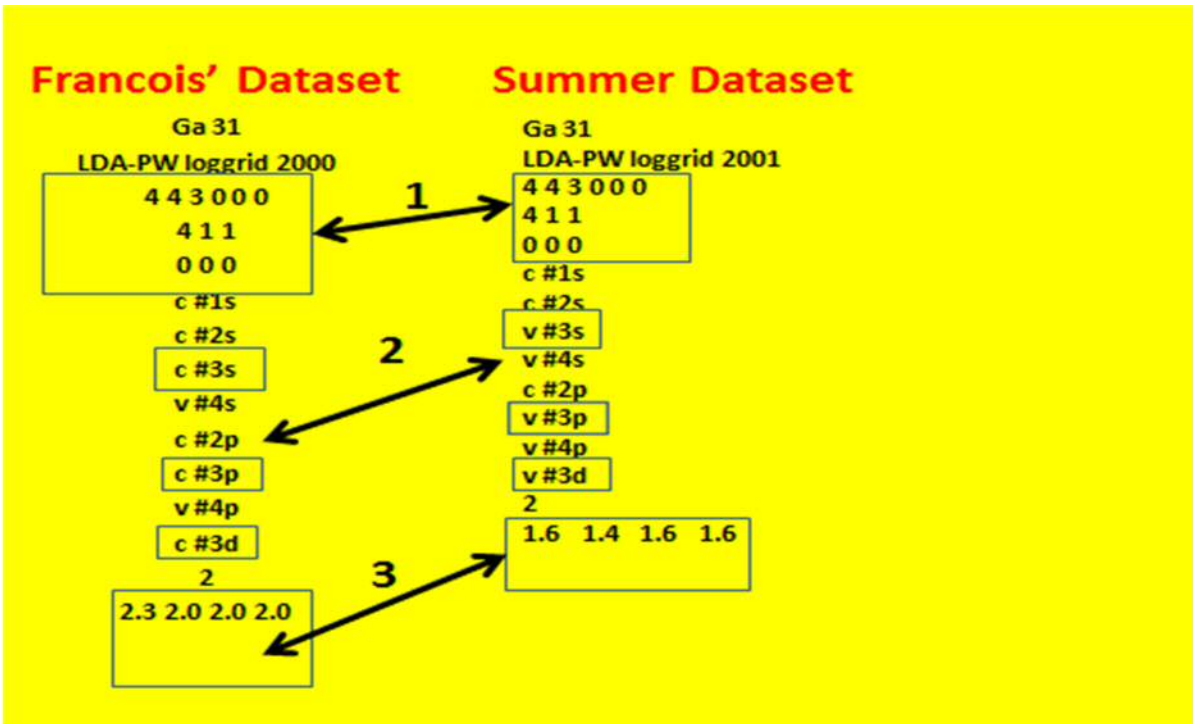

Figure 4: Atomic inputs for Gallium highlighting Electron Distribution and Radius difference.

In Figure 4, representing the case of Gallium, we see that there is a vast difference in the parameter of matching radius (arrow 3). Jollet's choice of radius was nearly 50% times our choice. This vast difference in the radius may be a potential reason for the difference in results. This larger radius in Jollet's dataset may have caused issues with the presence of excess room between the atoms when ABINIT was executed. Additionally, it is again pertinent to observe the choice of the electrons. We have many more valence state allocations than does Jollet this may either overshadow the difference in radii or it may also complement an inherently better choice. Again, it is really hard to pinpoint just exactly where the certain parameters contribute to better results. Generally, Jollet's results were very similar to our own, even if ours were consistently more precise to the LAPW plot. However, the amount of processing power each method requires may be the more important variable to consider when one is considering the best approach for dataset creation.
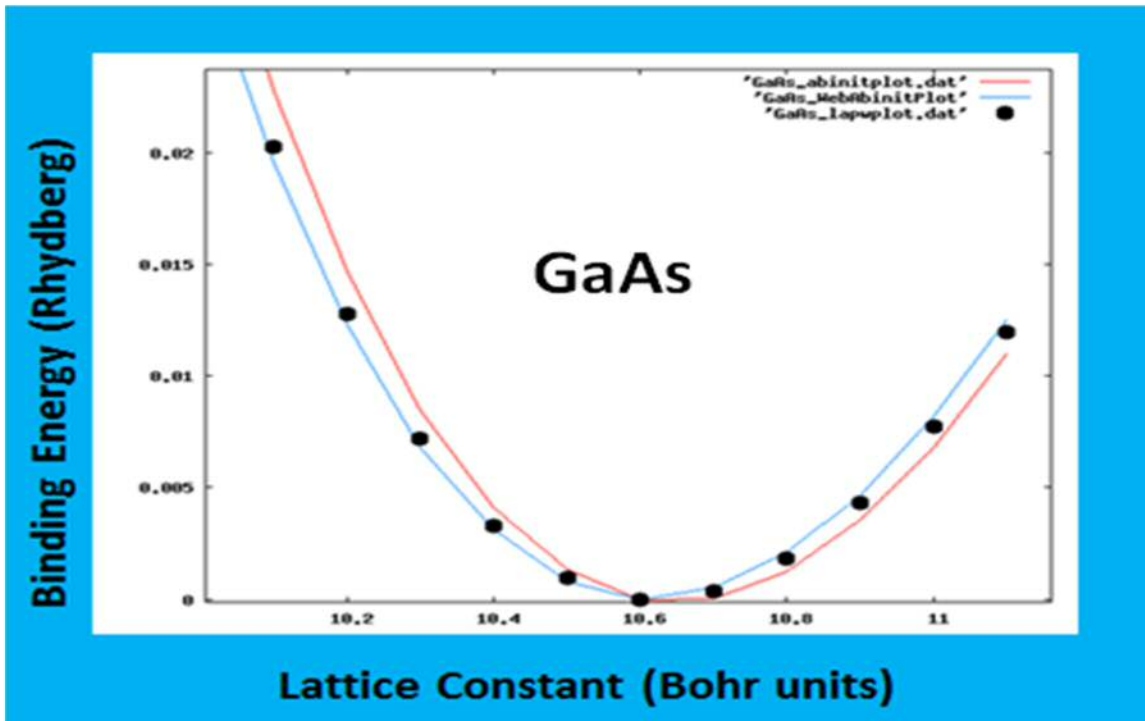

Figure 5: Binding Energy vs. Lattice Constant for Gallium-Arsenide.

Figure 5 shows the Binding Energy curve results to have a slightly better result for our choice of parameters than those of Jollet. It is an important observation to see that the lowest binding energy occurs at the same lattice constant value as the LAPW.

4. Tools for Comparison

4.1 Wavefunction Analysis and Logderivatives

The fundamental aspect of the PAW method is the creation of the Pseudofunction to more smoothly represent the All-Electron wave function. This relationship between the two functions is shown in the transition between Figure 5 and Figure 6. It is important to understand that there is two distinct Wavefunctions that we are dealing with. The All-Electron Wavefunction (Red) should be the very similar for both Jollet's dataset and our dataset. There may be very small differences, but this is only if the electron configuration is different as in the Barium case. This had better be the case because there should be the same total number of electrons in for both inputs. The Pseudo Wavefunction (Green) is, however, subject to change and may vary with our differing choice of parameters. A fundamental difference between the two datasets was the choice of basis functions in creating the Pseudofunction.
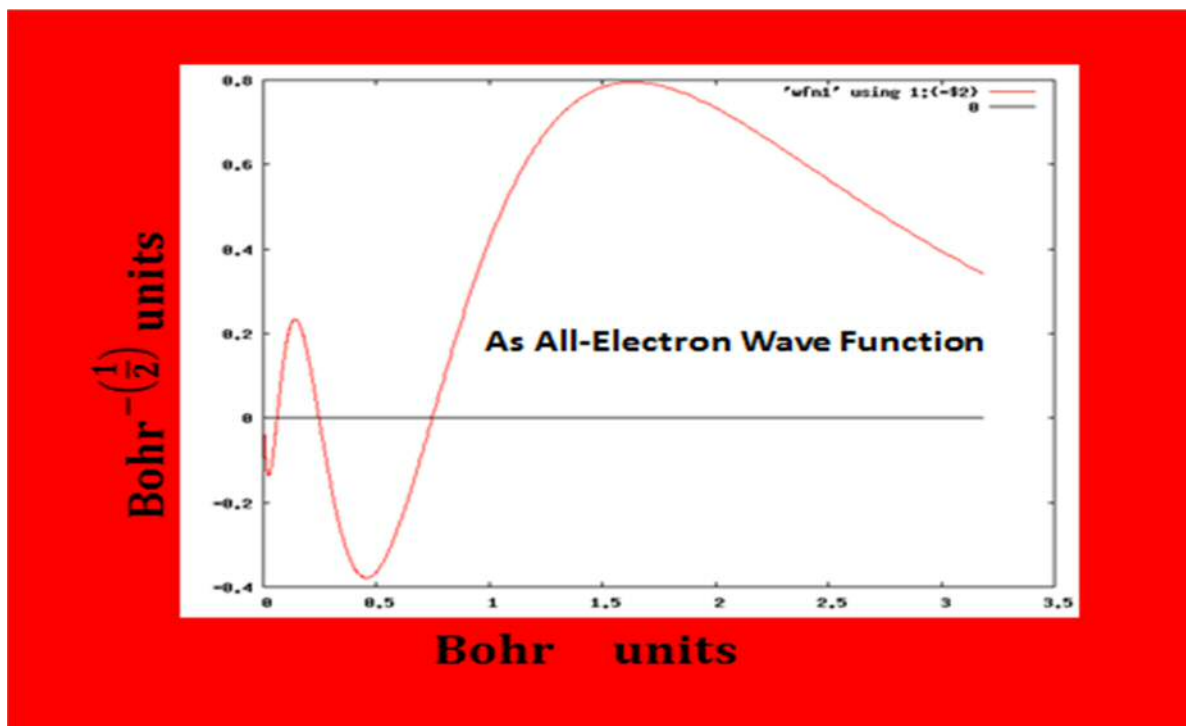

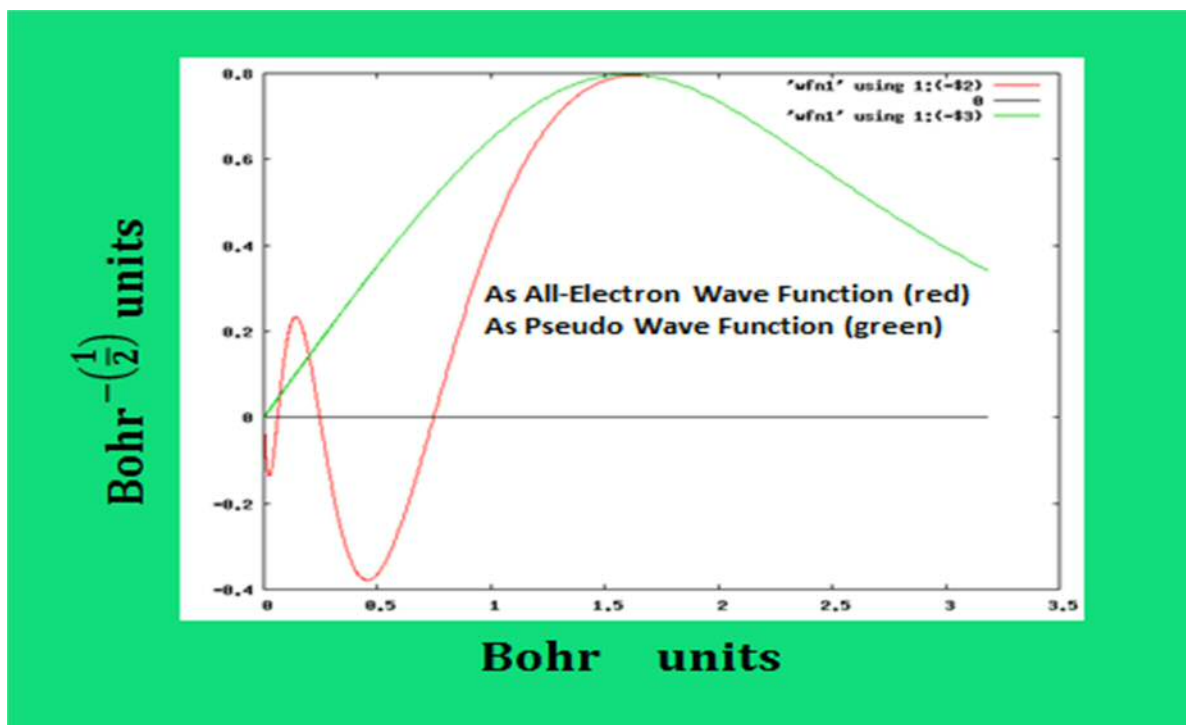Figure 6: Graph of the All-Electron Wavefunction.

Figure 7: Graph of the All-Electron Wavefunction with the newly created Pseudo Wavefunction.

Mostly, Jollet used polynomial functions in creation of his Pseudopotentials whereas we used Bessel functions. Now after we have the Pseudofunctions created, we must be sure that they accurately represent the All-Electron Wavefunction past the choice of matching radius. The Pseudofunction does not model the behavior of the All-Electron case until *after* the matching radius (see Figure 6). After this point, we hope the Pseudofunction accurately describes the All-Electron case. However, we need a means to test the accuracy between the two wavefunctions. We do this by looking at the logderivatives of both functions simultaneously:

$$\text{Equation 3:}\quad \text{Logderivative} \equiv \frac{\left(\frac{dR}{dr}\right)}{R}\Big|_{r_c}.$$

Notice the nature of this calculation is exactly that of how one takes derivatives of logarithmic functions. Doing this calculation for both the All-Electron wavefunction and the Pseudo-wavefunctions should give very similar results. We expect this at least to some degree as for later states, there may be a bit of disagreement, but the thing we are most interested in is using this as a tool for analysis. We may look the logderivatives of the S, P, D, etc. state and see if there is any huge disagreement between the two functions. If there is, then perhaps this specific state is where we need to refocus our attention and alter

the dataset accordingly. For a case in which the functions show little logderivative error, or disagreement, can be seen in Figure 8.
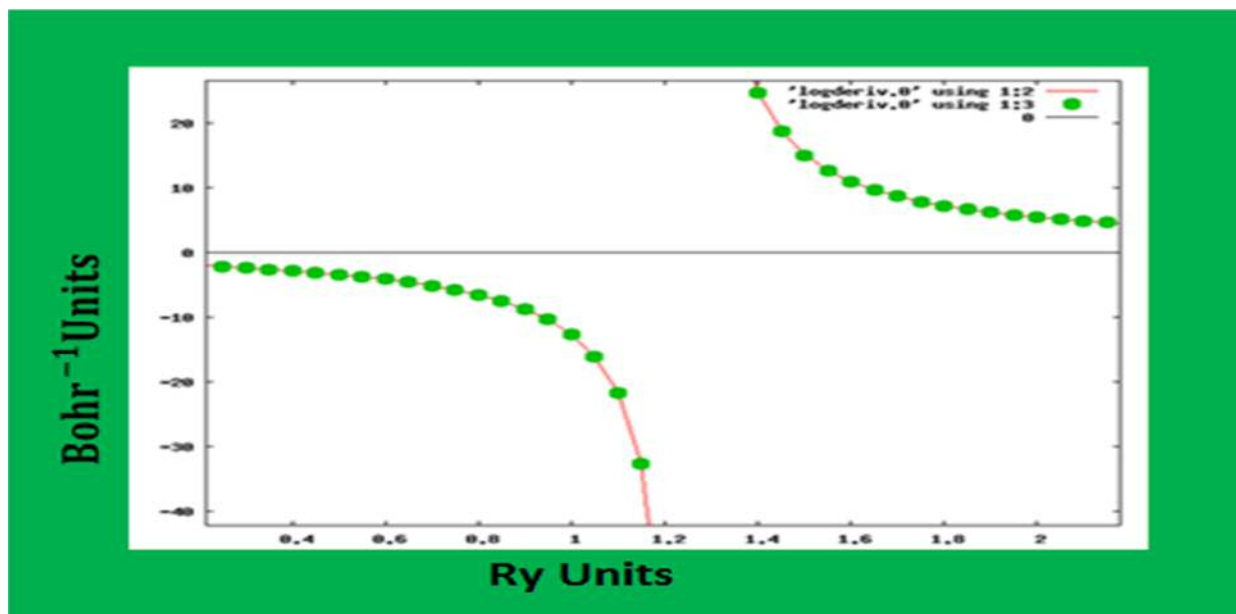

Figure 8: Showing the All-Electron logderivatives are in red, while the Pseudofunction logderivatives are shown as green dots.

4.2 Projector Wave Functions

Another key component in analyzing our datasets was by looking at the projector wavefunctions associated with our Pseudofunction and All-Electron Wavefunction. It was important that these projectors be smooth and go to zero after the matching radius value.
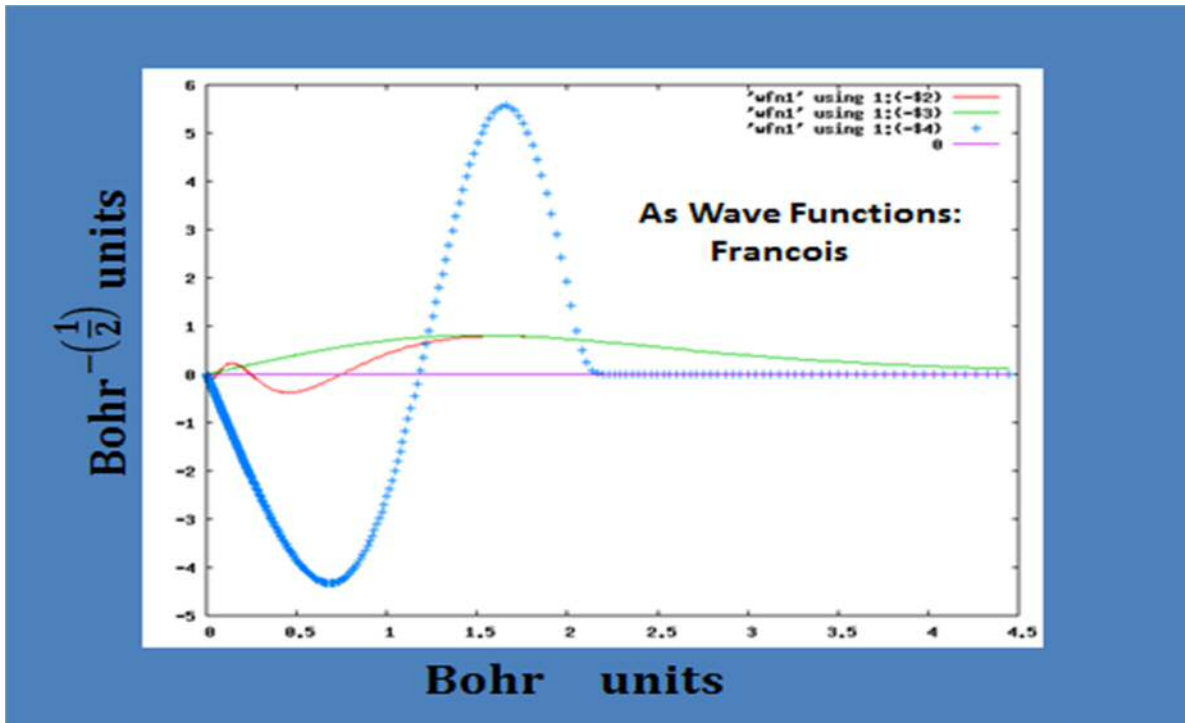
Figure 9: Graph showing all Wavefunctions for Francois Jollet's Arsenic dataset.

In Figure 9, the All-Electron wavefunction is shown in red, the Pseudofunction in green, and projector shown in blue (pink represents the line y=0). I observed these wavefunctions for both our cases and Jollet simultaneously to get a better idea about how our inputs related to each other. As such, I found that much of the time, the wavefunctions resembled each other, with differences in amplitude being the main difference. Predominantly speaking, I found that our results were better and the most likely reason is due to the more liberal placement of valence states in the dataset. As such, I observed smoother functions in general, especially the projectors, for our datasets compared to Jollet's. Although, I found our results were only better by a small margin.

5. Problems

5.1 Ghost States

The presence of "Ghost States" was perhaps the most problematic occurrence in my research. A "ghost state" occurs when the computing software finds numerical solutions to the Schrödinger Equation where there should not be solutions. In such cases, the self-consistent field step would not converge. When this happened, we would end up using inaccurate data when running ABINIT. This would result in a divergent trial, even though our chosen data was supposedly the optimum choice.

However, sometimes we were able graph the logderivatives and find where the problem arose.
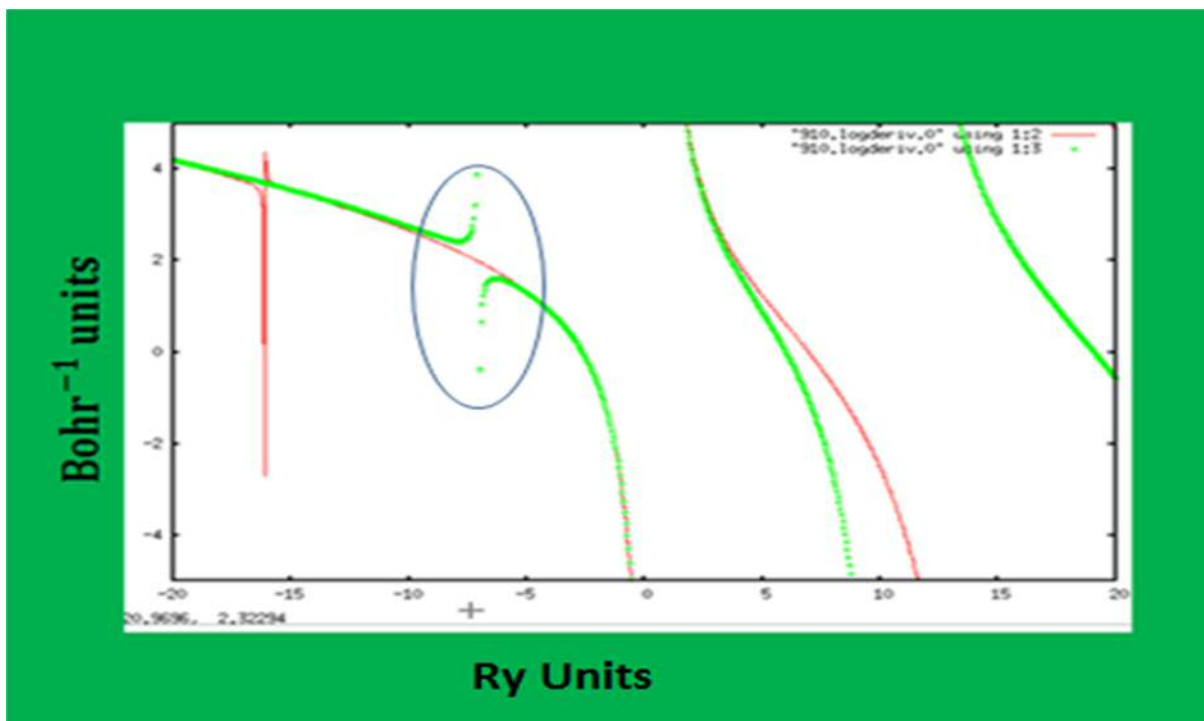

Figure 10: Presence of a "Ghost State" in the logderivative plot.

Notice how it is the Pseudofunction (green) that is causing the problem here. The asymptote shown in the blue circle in Figure 10 represents a bound state that is not reflected by the All-Electron Wavefunction (Red). This was the main cause of problems for us this summer as it was particularly difficult to pinpoint the cause of such behavior.

6. Conclusions

In summary, we focused on several interesting relationships between our datasets and that of Jollet's. These relationships were determined by choice of initial parameters of the either input file. While each party had their own means of dataset creation, there seems to be agreement between the results. Although our datasets were found to yield more precise results, our data runs were generally more computationally expensive. This is a reasonable consequence of our choice of liberal assignment of valence states in the datasets. However, a question may be posed as to whether this extra computational cost is legitimate. After all, we are using a series of approximations to model how one may expect these materials to behave. Perhaps the extra precision is not necessary. If so, then Jollet's datasets that generally incorporate less valence states may in fact be a more efficient

approach. Another honorable mention of this research is that it was the first time Jollet's datasets were tested in compound form. All of his other trials were dealing with elemental calculations. The fact that we saw such agreement between his datasets and ours, even though he only dealt with singular elements, speaks to the efficiency of both methods. Additionally, the agreement found between our final results, given that, at times, our initial datasets were greatly different, leads one to believe that there may in fact be a range of "acceptable" input parameters that may yield good results. Perhaps with the study and refinement of these input parameters, and debugging of the computer codes, one may eventually come to an acceptable hypothesis as to the best format in which to computationally model materials. This research was very important because we must always test datasets to ensure that they will comply in computational modeling applications.

6.1 Delta Factor

In our work, we have used the Binding Energy curve qualitatively to distinguish and test our calculations. More recently, there has been a development to make this analysis more quantitative. The idea comes with this notion of a Delta Factor. This measurement is found by integrating the difference between the standard Binding Energy curve and any other calculation result (Lejaeghere et al.). Jollet's datasets were designed to have small Delta Factors. Up until now, this Delta Factor has only been used with elemental materials.

However, the authors are now enhancing the features of the Delta factor such that it modifies the comparison to include binary compounds. If we computed the Delta Factor for the materials we did this summer, we expect them to be very small. This new comparison may provide a standard that we can very easily use based on our previous work.

Acknowledgements and Collaborators:

References

"Binding Energy (physics)." *Britannica Online Encyclopedia*. N. p., n.d. Web. 2 May 2014.

Blaha, P. et al. "Full-Potential, Linearized Augmented Plane Wave Programs for Crystalline Systems."
    Computer Physics Communications 59.2 (1990): 399–415. *ScienceDirect*. Web. 3 May 2014.

Blöchl, P. E. "Projector Augmented-Wave Method." *Physical Review B* 50.24 (1994): 17953–17979. *APS*.
    Web. 3 May 2014.

Gonze, X. et al. "ABINIT: First-Principles Approach to Material and Nanosystem Properties." *Computer
    Physics Communications* 180.12 (2009): 2582–2615. *CrossRef*. Web. 3 May 2014.

Hohenberg, P. and W. Kohn. "Inhomogeneous Electron Gas." *Physical Review* 136.3B (1964): B864–B871.
    APS. Web. 3 May 2014.

Holzwarth, N. A. W. et al. "A Projector Augmented Wave (PAW) Code for Electronic Structure
    Calculations, Part I: Atompaw for Generating Atom-Centered Functions." *Computer Physics
    Communications* 135.3 (2001): 329–347. *ScienceDirect*. Web. 3 May 2014.

Jollet, F. et al. "Generation of Projector Augmented-Wave Atomic Data: A 71 Element Validated Table in
    the XML Format." *Computer Physics Communications* 185.4 (2014): 1246–1254. *ScienceDirect*.
    Web. 3 May 2014.

Lejaeghere, K. et al. "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 (2014): 1–24. *Taylor and Francis+NEJM*. Web. 4 May 2014.

Martin, Richard M. *Electronic Structure : Basic Theory and Practical Methods /*. Cambridge, UK ; New
    York :: Cambridge University Press,, 2004. Print.

Perdew, John P. and Y. Wang. "Accurate and Simple Analytic Representation of the Electron-Gas
    Correlation Energy." *Physical Review B* 45.23 (1992): 13244–13249. *APS*. Web. 3 May 2014.

Schwarz, Karlheinz. and P. Blaha. "Solid State Calculations Using WIEN2k." *Computational Materials
    Science* 28.2 (2003): 259–273. *ScienceDirect*. Web. 3 May 2014. Proceedings of the Symposium
    on Software Development for Process and Materials Design.

Slater, John C. 1900-1976. *Quantum Theory of Molecules and Solids /*. New York :: McGraw-Hill,, 1963.
    Print. International Series in Pure and Applied Physics.

Van Zeghbroeck, B. "Chapter 2: Semiconductor Fundamentals." *Semiconductor Fundamentals*. N.p., n.d.
    Web. 12 May 2014.