

Understanding density functional theory (DFT) and completing it in practice

Cite as: AIP Advances 4, 127104 (2014); <https://doi.org/10.1063/1.4903408>

Submitted: 29 September 2014 . Accepted: 24 November 2014 . Published Online: 03 December 2014

Diola Bagayoko 



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Perspective: Fifty years of density-functional theory in chemical physics](#)

The Journal of Chemical Physics **140**, 18A301 (2014); <https://doi.org/10.1063/1.4869598>

[Density functional theory: An introduction](#)

American Journal of Physics **68**, 69 (2000); <https://doi.org/10.1119/1.19375>

[Perspective: Kohn-Sham density functional theory descending a staircase](#)

The Journal of Chemical Physics **145**, 130901 (2016); <https://doi.org/10.1063/1.4963168>

AVS Quantum Science

Co-published with AIP Publishing



Coming Soon!



Understanding density functional theory (DFT) and completing it in practice

Diola Bagayoko

Southern University and A&M College in Baton Rouge (SUBR), Baton Rouge, Louisiana 70813, USA

(Received 29 September 2014; accepted 24 November 2014; published online 3 December 2014)

We review some salient points in the derivation of density functional theory (DFT) and of the local density approximation (LDA) of it. We then articulate an understanding of DFT and LDA that seems to be ignored in the literature. We note the well-established failures of many DFT and LDA calculations to reproduce the measured energy gaps of finite systems and band gaps of semiconductors and insulators. We then illustrate significant differences between the results from self consistent calculations using single trial basis sets and those from computations following the Bagayoko, Zhao, and Williams (BZW) method, as enhanced by Ekuma and Franklin (BZW-EF). Unlike the former, the latter calculations verifiably attain the absolute minima of the occupied energies, as required by DFT. These minima are one of the reasons for the agreement between their results and corresponding, experimental ones for the band gap and a host of other properties. Further, we note predictions of DFT BZW-EF calculations that have been confirmed by experiment. Our subsequent description of the BZW-EF method ends with the application of the Rayleigh theorem in the selection, among the several calculations the method requires, of the one whose results have a full, physics content ascribed to DFT. This application of the Rayleigh theorem adds to or completes DFT, in practice, to preserve the physical content of unoccupied, low energy levels. Discussions, including implications of the method, and a short conclusion follow the description of the method. The successive augmentation of the basis set in the BZW-EF method, needed for the application of the Rayleigh theorem, is also necessary in the search for the absolute minima of the occupied energies, in practice. © 2014 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License. [<http://dx.doi.org/10.1063/1.4903408>]

I. INTRODUCTION

We begin by following the works of Hohenberg and Kohn¹ and of Kohn and Sham² to review the derivation of density functional theory (DFT) and of the local density approximation (LDA) of it. The necessity for an overview of this derivation becomes apparent in the last paragraph of Section II below, where we summarize necessary conditions for the results of ab initio, self consistent calculations to represent the physical content of DFT or of LDA. Then, in Section III, we illustrate the “band gap problem,” i.e., the woeful underestimation of measured band gaps, by 30 to 50% or worse, by most DFT and LDA calculations. The tabular presentation of some 99 different calculations with results significantly lower than corresponding, experimental ones also shows DFT BZW and BZW-EF calculations whose findings are in agreement with the corresponding, measured ones. In Section IV, we provide a summary description of the BZW and BZW-EF method. This description straightforwardly leads to the two key features of the method that guarantee its faithful rendition of the physical content of DFT. They are (1) the strict adherence to the conditions inherent to the derivation of DFT, including the verified attainment of the ground state energy, and (2) the application of the Rayleigh theorem to avoid the destruction of the physical content of low energy unoccupied levels. With the introduction of the basis set augmentation that is needed for verifiably

reaching the ground state energy, the first feature adds to or completes DFT in practice. The second feature also does so by identifying, out a potentially infinite number, the calculation that provides the physical content of DFT.

II. THE DERIVATION OF DFT AND OF LDA

In 1964, Hohenberg and Kohn¹ introduced density functional theory (DFT). This introduction entailed the following theorem that they established. *Let an arbitrary number of electrons in a box be subject to an external potential $v(\mathbf{r})$ and the electron-electron Coulomb repulsion, then, this potential is a unique functional of the electron density $n(\mathbf{r})$, except for an additive constant.* In the above and following expressions or equations for the potential, the density, and functionals, the bold signifies that the position variable is a vector. This convention holds throughout this text. Hohenberg and Kohn considered the case of a non degenerate ground state in the derivation and used reductio ad absurdum. It was later established that the theorem is applicable to the case of a degenerate ground state. The Hamiltonian of the system is written as

$$H = T + U + V \quad (1)$$

where T and U are for the kinetic energy and electron-electron interaction operators, with

$$V = \int v(\mathbf{r})\psi^*(\mathbf{r})\psi(\mathbf{r})d\mathbf{r}. \quad (2)$$

The authors concluded that the sum of the kinetic and electron-electron interaction energies is a *universal functional* of the charge density, since Ψ is a functional of $n(\mathbf{r})$:

$$F[n(\mathbf{r})] \equiv (\Psi, (T + U) \Psi). \quad (3)$$

It followed from the above results that, for a given potential $v(\mathbf{r})$, we have the energy functional

$$E_v[n] = \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n]. \quad (4)$$

Hohenberg and Kohn proceeded to establish that this unique functional of the density attains its minimum (from above) for the correct ground state density $n(\mathbf{r})$, provided the total number of particles,

$$N[n] \equiv \int n(\mathbf{r})d\mathbf{r} = N \quad (5)$$

is kept constant. Specifically, the actual quote from the Hohenberg and Kohn article¹ follows.

“It is well known that for a system of N particles, the energy functional of Ψ'

$$E_v[\Psi'] = (\Psi', V\Psi') + (\Psi', (T+U)\Psi') \quad (6)$$

has a minimum at the ‘correct’ ground state Ψ , relative to arbitrary variations of Ψ' in which the total number of particles is kept constant.”

The authors took the next crucial step consisting of showing that a density $n'(\mathbf{r})$, associated with a another potential $v'(\mathbf{r})$ [different from $v(\mathbf{r})$], leads to a minimum value of the above energy functional that is higher than the one obtained with the ground state density $n(\mathbf{r})$ of the system.

Specifically, they established the following inequality:

$$\int v(\mathbf{r})n'(\mathbf{r})d\mathbf{r} + F[n'] > \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n] \quad (7)$$

It should be noted that the potential $v(\mathbf{r})$ is used on both sides of the inequality. This result is generally referred to as the second Hohenberg and Kohn theorem or as the DFT variational principle. Hohenberg and Kohn delved into several other considerations that are outside the scope of this work. These topics include in-depth discussions of the case of a gas with an almost constant density and that of another with slowly varying density. Gradient expansions emerged in connection with the latter case.

In 1965, Kohn and Sham² introduced the local density approximation (LDA) of DFT. These authors rewrote the energy functional in Equations (4) and (6) as

$$E = \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \frac{1}{2} \iint \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + T_s[n] + E_{xc}[n], \quad (8)$$

where $T_s[n]$ is the kinetic energy of a system of non-interacting electrons whose density is $n(\mathbf{r})$, while $E_{xc}[n]$ is defined as the exchange correlation energy of a system of interacting electrons with the same density $n(\mathbf{r})$. In the case of slowly varying densities, Kohn and Sham obtained an expression for $E_{xc}[n]$ in terms of the exchange correlation energy per electron of a uniform gas whose density is n .

$$E_{xc}[n] = \int n(\mathbf{r})\epsilon_{xc}(n(\mathbf{r}))d\mathbf{r}, \quad (9)$$

where $\epsilon_{xc}(n(\mathbf{r}))$ is the exchange and correlation energy per electron of a uniform electron gas of density n . The next pivotal step taken by these authors consisted of using the stationary property of Equation (8) (i.e., a particular form of Equation (6)) to derive an equation that straightforwardly leads to the Kohn Sham (KS) equation,

$$\left\{ -\frac{1}{2}\nabla^2 + \left[v(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + \mu_{xc}(n(\mathbf{r})) \right] \right\} \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \quad (10)$$

with the charge density computed using *only the wave functions of the occupied states* as

$$n(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2. \quad (11)$$

The referenced derivation of the Kohn Sham equation **requires** that the total number of particles (N) be kept constant. Interestingly enough, Equation (11) is sufficient to underscore the fact that N must be constant. Any change in N means a totally new density and hence new values for all quantities that are unique functionals of the density! The above KS equation describes a system of non interacting electrons moving in the potential between brackets in Equation 10. The exchange correlation contribution to the chemical potential of a uniform gas of density n is $\mu_{xc}(n(\mathbf{r}))$.

The above overview of some key features of the derivation of DFT and of LDA was to underscore the following points: **(a)** *In all DFT applications, the total number of particles must be kept constant; and, (b) for the outcomes of an electronic structure calculation using DFT or LDA potentials to represent the physical content of DFT or LDA, it is necessary **either (b.1)** to have, à priori, the “correct,” tridimensional ground state charge density of the system **or (b.2)** to verify (as opposed to assuming) that the calculated, occupied energies have attained their absolute minima (from above) for the system under study.* The content of this paragraph has far reaching implications as shown below.

III. SOME FAILURES OF DFT CALCULATIONS ASCRIBED TO DFT: “THE BAND GAP PROBLEM”

From the onset, we make a difference between failures of calculations using DFT potentials and intrinsic failures or deficiencies of DFT. While calculations using DFT potentials are the most successful and widespread in the last three decades, in several areas of materials studies, most of these calculations have resulted in significant underestimations of the band gaps of semiconductors and of insulators. The content of Table I below amply illustrates “the band gap problem” noted in the Nobel Lecture of Professor Kohn.³

Even though our focus is mainly on the description of electronic and related properties of semiconductors and insulators, with a special emphasis on band gaps, a similar problem exists for the study of atoms, molecules, clusters, and nanostructures. Indeed, for most finite systems in their ground states, there exists a difference between the lowest, unoccupied energy and the highest occupied one. For molecules, the former and the latter energy levels are often referred to as the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital

TABLE I. Band gaps from ab initio, self consistent DFT calculations. None of the 99 different calculations in Column 3, with corresponding ranges of calculated gaps in Column 4, produced results acceptably close to the corresponding, experimental values in Column 5. Details are available on the calculations in the references in Column 6. *In stark contrast, the results of the ab initio, self consistent DFT BZW and BZW-EF calculations in Column 6 agree with corresponding, experimental ones in Column 5.*

Material	DFT Potential	Number of Calculated Gaps	Ranges of Calculated Gaps	Experimental Gaps	DFT BZW or BZW-EF Gaps
Cubic InN	LDA & GGA	10	-0.55 - + 0.08	0.61 eV	0.65 eV ⁴
AlAs	LDA	5	1.20 – 1.44 eV	2.15 – 2.25 eV	2.14 – 2.15 eV ⁵
zb-ZnS	LDA & GGA	5	1.65 – 2.37 eV	3.723 eV	3.725 eV ⁶
w-GaN	LDA & GGA	17	1.68 – 2.52 eV	3.3-3.5 eV	3.20 & 3.29 eV ⁷
zb-GaN	LDA	8	1.80 – 2.10 eV	3.10 eV	2.90 eV ⁷
Rutile TiO ₂	LDA & GGA	18	1.67 – 2.12 eV	3.00 – 3.10	2.95 & 3.05 eV ⁸
w-ZnO	LDA	12	0.23 – 2.26 eV	3.30 – 3.40 eV	3.39 eV ⁹
zb-BP	LDA & GGA	15	1.11 – 1.38 eV	2.02 ± 0.05 eV	2.02 eV ¹⁰
c-BN	LDA & GGA	9	4.20 – 4.47 eV	6.20 – 6.4 ± 0.5 eV	6.48 eV ¹¹

(HOMO), respectively. The difference between these two energies is the energy gap (for discrete energy levels).

Bagayoko et al., using the original BZW method and LDA potentials, predicted the band gaps and other properties of cubic Si₃N₄¹² and c-InN.⁴ These predictions were confirmed by experiment in 2003^{13,14} and 2006,¹⁵ respectively. Ekuma and Bagayoko⁹ predicted the indirect band gap of rutile TiO₂ to be 2.95 eV while the direct one was found to be 3.05 eV. Before this DFT work, TiO₂ was reported to be a direct gap semiconductor. This prediction that TiO₂ is an indirect gap material was confirmed¹⁶ in 2014. Bagayoko and Franklin¹⁷ were the first to settle definitively (with ab-initio computational results) the discrepancy between measured band gaps of 2 eV, mostly before the year 2000, and of less than 1 eV, mostly after 2000, for wurtzite InN. The calculated LDA-BZW band gap of 0.88 eV confirmed the experimental value of 0.883 and 0.889 eV¹⁸ for w-InN with carrier concentrations in the range of 10¹⁹ cm⁻³. The Burstein Moss effect changes the gap of this material drastically. *These examples are to illustrate the fact that BZW and BZW-EF calculations not only describe, but also predict accurately electronic and related properties of materials with DFT potentials.*

The underestimates in Table I, from the 99 calculations in Column 3, are mostly explained in the literature in terms of the effects of self-interaction¹⁹ and particularly of the derivative discontinuity^{20–22} of the exchange correlation energy. The reader is referred to these sources for details. It should be noted here that Perdew and Levy²⁰ and Sham and Schlüter^{21,22} established the derivative discontinuity for semiconductors and insulators, respectively, by considering systems where the number of particles is not constant. While this discontinuity should be expected to contribute prominently to the band gaps in such systems, it should not affect, at least by much, our solution for the aforementioned coupled Kohn-Sham and charge density equations whose validity rests on a constant number of particles. For our purposes, the crucial point is the following: any correct assessment of the contributions of self-interaction and of the derivative discontinuity of the exchange correlation energy has to be carried out using computational results that truly represent the physical content of DFT (or LDA). Hence, these calculations must adhere to both conditions (a) and (b) at the end of Section II. As far as we can determine, none of the referenced 99 calculations in Table I did so, as per the contents of the publications of the corresponding results. The reason for the agreement between all DFT or LDA BZW and particularly BZW-EF results and corresponding, experimental ones becomes apparent in the description of the method below.

IV. THE BZW AND BZW-EF METHOD

The following description of the method is in two parts. The first one (Part I), which entails the use of successively larger basis sets (with richer angular and radial characteristics), adheres strictly

to the mathematical and physical conditions inherent to the derivation of DFT. This part of the method allows the attainment of the absolute minima of the occupied energies, as required by DFT, while keeping the number of particles constant.

The second part (Part II) of our method, begotten from our prior experiences with the contraction of basis sets,²³ enunciates the Rayleigh theorem relative to the eigenvalues of the same equation solved with two basis sets, with the smaller basis set entirely included in the larger one. It utilizes this theorem to select, among the potentially infinite number of calculations with large basis sets leading to the same absolute minima of the occupied energies, the one with the smallest basis set. This calculation, due to the attainment of the ground state energy, provides the physical content of DFT not only for the occupied states, but also for low energy unoccupied ones. *The basis set for this particular calculation is referred to as the optimal basis set.* The use of basis sets much larger than the optimal one does not lead to changes in the occupied energies; however, it often lowers some unoccupied energies, including the lowest ones. We address below the mathematical artifact responsible for this unphysical, extra-lowering of unoccupied levels while the occupied ones do not move.

Our calculations have features in common with many others. Specifically, we utilize DFT potentials, with emphasis on LDA and generalized gradient approximation (GGA)²⁴ ones, like other calculations. We employ the linear combination of atomic orbital (LCAO) formalism, as done by most electronic structure calculations. Most of our calculations have been carried out with the program package developed and refined at the Ames Laboratory of the US Department of Energy, Ames, Iowa.²⁵ The radial parts of atomic orbitals include Gaussian functions. As done in other calculations, we begin by obtaining self consistent electronic energies and related wave functions for the atomic or ionic species present in the material under study. We generally perform preliminary solid state calculations whose outputs allow an assessment of charge transfers, if any, between the atomic species. This preliminary work is important and can be partly guided by the octet rule of chemistry. It is not proven that iterative processes, entailing only linear variations, can correct for an input charge distribution (in three dimensions) that is drastically different from the actual one. The Bagayoko, Zhao and Williams (BZW) method and its enhanced version, following the works of Ekuma and Franklin (BZW-EF), are described below. Our strict adherence to this method, we contend, is the key reason our results agree with experimental ones, on the one hand, and disagree with findings from DFT calculations that mostly used single trial basis sets, on the other hand. *Part I of the method follows.*

In the BZW^{25,26,30} or BZW-EF^{9,27-29} method, calculations for materials start with a small basis set no smaller than the minimum one. The latter is defined as the basis set just large enough to account for all the electrons in the system (i.e., material) under study. Completely self consistent calculations are performed with this basis set that can be the minimum basis set. A second, self consistent calculation is carried out with a basis set comprising that of the previous calculation as augmented by one orbital. Depending on the s, p, d, or f character of the added orbital, the number of functions (and the dimension of the Hamiltonian) increases by 2, 6, 10, or 14, respectively. The criterion for adding the referenced orbital makes the difference between the BZW and the BZW-EF method. In the original BZW^{25,26,30} we added orbitals in the increasing order of the excited energies they represent in the atomic or ionic species. In the BZW-EF method,²⁷⁻²⁹ for a given principal quantum number and for a given site, we add p, d, and f orbitals, if applicable, before the corresponding s orbital of that quantum number. The realization that the polarization of p, d, and f orbitals has primacy over the spherical symmetry of s orbitals, for *the valence electrons* in binary molecules and other systems comprising two or more atomic sites, led to this Ekuma and Franklin (EF) enhancement of the method. This counter-intuitive approach recognizes the fact that in systems varying from binary molecules to solids, the redistribution of the cloud of *valence electrons* is expected to deviate drastically from those around isolated atoms or ions.

While several BZW calculations agreed very well with experiments, as attested to by the predictions for cubic Si₃N₄ and cubic InN, some calculated band gaps were 0.1 to 0.2 eV lower than corresponding, experimental values – after taking uncertainties into account. In contrast, the BZW-EF calculations performed to date led to band gaps in full agreement with experiment. This

difference of 0 to 0.2 eV between BZW and BZW-EF results is another manifestation of the limitations of iterative processes. Iterations cannot correct for an inadequate angular representation in the trial functions, irrespective of the degree of convergence attained by the said iterations. We elaborate on this point in the discussion section in connection with the work of Franklin et al. on wurtzite ZnO. Unless otherwise indicated, our calculations henceforth utilize the enhanced BZW-EF version of the method. With the above understanding of the augmentation procedure for the basis set, the rest of Part I of the method follows below.

After setting the Fermi levels to zero, the occupied energies from calculations I and II are compared numerically and graphically. They are generally different, with some *occupied energies* from calculation II being lower than their corresponding ones from calculation I. The above lowering of some occupied energies from calculation II indicates that the initial, small basis set is inadequate for the rendition of the wave function of the system - in size (dimension), angular symmetry, and radial functions - for the description of the ground state. Equivalently, the above difference means that the basis set in calculation I is not complete for the description of the ground state. With no proof that the outcomes of Calculation II provide the physical description of the system, we proceed with calculation III, after augmenting the basis set of Calculation II as described above. The occupied energies of Calculations II and III are compared.

This process of augmenting the basis set and of carrying out self consistent calculations continues until three successive calculations produce the same occupied energies. At this point, the calculations are completed, given that we reached the absolute minima of the occupied energies. In early applications of the method, we used to require that two consecutive calculations produce the same occupied energies. After our work on zinc blende GaP and BP, where Calculations II and III led to the same occupied energies (and unoccupied ones up to 10 eV), but Calculation IV produced occupied energies lower than those from III, we required three consecutive calculations to avoid taking local minima for absolute minima, as the latter are the ones required by DFT.

The above paragraphs provide the full description of Part I of the BZW method. The calculations keep the total number of particles, N , constant. Further, they produce the absolute minima of the occupied energies. In DFT language, the optimal basis set can be utilized to construct the “correct” charge density of the system and to calculate ground state properties, including energy gaps and band gaps and widths. In retrospect, we realized that Hohenberg and Kohn actually suggested this approach (Part I of our method): Once $F[n]$ is known, “*the problem of determining the ground state energy and density in an external potential would be rather easy since it requires merely the minimization of a functional of the three-dimensional density function.*”¹ The energy functional minimized by the BZW or BZW-EF method is the one in Equations (4), (6), and (8) above. The only difference between Equations (4) and (6), on the one hand, and Equation (8), on the other hand, is that the latter contains the local density approximation of some quantities in (4) and (6).

Despite the explicit reference to *arbitrary variations* of the trial wave functions (i.e., basis sets) in connection with Equation (6) above, the community seems to have so far believed that the “linear” minimization through the iterative process was sufficient. By “linear” minimization we mean a minimization process (i.e., iterations) where changes are made only to “linear expansion” coefficients. Such a minimization, as demonstrated by several tables of successive calculations of our group^{9,10,31} is simply inadequate. Indeed, it cannot correct for a serious deficiency of a single trial basis set in (a) size (dimension of the Hamiltonian), (b) angular symmetry (for the redistribution of the cloud of valence electrons), (c) richness of the radial features of the orbitals, and (d) other input conditions for some materials. While (a) through (c) are self-explanatory, (d) requires some comments. Nwigboji et al.³² recently performed several sets of ab initio, self consistent LDA BZW-EF calculations for wurtzite aluminum nitride. With the three positive and negative oxidation states of Al and N, respectively, the BZW-EF method required three different sets of calculations with input of Al^{1+} and N^{1-} , Al^{2+} and N^{2-} , and Al^{3+} and N^{3-} , respectively. Wurtzite AlN is described with the optimal basis set calculation for Al^{3+} and N^{3-} as input species. This choice was made based on the binding energy that is 1.5 V larger in magnitude than the results from the calculations with the other ionic inputs. Also, the upper valence band width is 1 eV larger than the corresponding results from calculations with the other input ionic species.

Another perspective on the solution of the DFT equations has been suggested by Bagayoko et al.³³ who noted the fact that it is a system of equations that has to be solved. Kohn and Sham explicitly stated so.² Once the exchange correlation potential is known, the system of equations is reduced to the KS Schrodinger type equation and the equation giving the ground state charge density in terms of the wave functions of the occupied states. For this system, the second equation is similar to a constraint on the KS equation. With a single trial basis set, with the changes to linear expansion coefficients only, the iterations solve self consistently the KS equation. However, these iterations do not solve the system of equations self consistently. To do so, one has to change the initial input to the charge density equation in terms of the size, angular symmetry, and radial characteristics of the basis set as done by the BZW-EF method. Additionally, the issue is not to change the initial input to the charge density equation using arbitrarily selected basis sets. As per the BZW-EF method, the minimization of the energy functional in Equations (6) or (8) requires that the basis sets in question be such that, except for the first and the last ones, each one contains the basis set immediately preceding it in size and is in turn included in the basis set that immediately follows it in size. Essentially, this minimization that solves the system of equation can be viewed to comprise iterations for the charge density equations, with progressively augmented basis sets, and iterations for the KS equation for each choice of basis set. *The above embedding of increasing basis sets is actually needed for the minimization of the energy functional in Equations (6) or (8); as such, it adds to DFT in practice.* These points complete the description of Part I of the BZW-EF (and BZW) method. Part II of the BZW-EF method follows.

The second part (Part II) of the BZW or BZW-EF method, as given below, does not follow directly from any derivation of DFT known to us. Further, it turns out to be necessary for most DFT calculations utilizing the LCAO formalism. The Rayleigh theorem is a key basis for this Part II of our method. This theorem states that when the same eigenvalue equation is solved with N and $(N+1)$ functions, with the N functions entirely included in the $(N+1)$, then, the first N ordered (from the smallest to the largest) eigenvalues obtained with $(N+1)$ functions are lower or equal to corresponding ones from the calculation with N functions. Stated differently, the theorem means that the increase of the size of the basis set does not increase a given eigenvalue; rather, an eigenvalue is decreased if it is not at its absolute minimum value, otherwise, it is unchanged. This theorem shows that when three successive calculations produce the same occupied energies, the one with the smallest basis set provides the DFT description of the material under study. The other two calculations, and other calculations with much larger basis sets that contain the optimal basis set, do not change the occupied energies; however, as per the Rayleigh theorem, they continue to lower some unoccupied energies, including the lowest lying ones. This extra lowering of some unoccupied energies is a mathematical artifact^{26,27} and not the manifestation of a physical interaction contained in the Hamiltonian. It destroys the physical content of the low lying unoccupied energies as obtained with the optimal basis set.

A subtle point about the Rayleigh theorem that has been missed by some previous readers of our work follows. Before reaching the absolute minima of the occupied energies, the successive calculations of the BZW-EF method do not solve the same eigenvalue equation! Indeed, the charge densities, and hence the potentials and the Hamiltonians for two consecutive calculations, are respectively different, until the occupied energies reach their absolute minima. *So, we ascribe the lowering of the occupied and low, unoccupied energies to physical interactions contained in the Hamiltonian, up to the attainment of the absolute minima of the occupied energies.*

A DFT based proof of Part II of our method we are articulating for the first time follows. Upon the attainment of the absolute minima of the occupied energies, as stated by Zhao et al.²⁷ and others,^{9,31} there is practically no difference between the charge densities from two self consistent calculations that both lead to the absolute minima of the occupied energies. The basis sets for these calculations include the optimal basis set. Intuitively, one can understand the reason the charge density no longer changes, given that it is calculated using only the wave functions of the occupied states. *Once the two charge densities from said calculations are the same, then the potentials and other quantities that are different, unique functionals of the density must remain unchanged.* Even though the Hamiltonian matrices are different in terms of size and of some matrix elements for two such calculations that yield the same absolute minima of the occupied energies, *the energy*

content of these Hamiltonians, as in Equation (6) (or (8) for LDA), that is a unique functional of the density, remains the same as that obtained with the optimal basis set. Hence, the lowering of unoccupied energies cannot be attributed to the content of the Hamiltonians. As noted above, it is a mathematical artifact that is not an ill-defined “basis set effect”; rather, it is a “basis set and variational” effect that can be invoked only after verifiably attaining the absolute minima of the occupied energies.

The above description of Part II of the BZW or BZW-EF method shows that it adds to or completes DFT, inasmuch as its correct application is concerned. Indeed, the above referenced extra-lowering destroys the physical content of the lowest laying unoccupied energies in atoms, molecules, some nanostructures, semiconductors, and insulators. It provides a simple explanation of the rampant underestimations of measured band gaps by most single basis set DFT (and LDA) calculations. The importance of Part II of the method resides in the preservation of the physical content of low laying unoccupied energies for non metals. For binding energies and band widths, Part I of the BZW-EF method is needed even for the description of metals. As noted above, the embedding of basis sets, as done in the BZW-EF method, seems to be necessary for the minimization of the energy content of the Hamiltonian as in Equations (6) and (8).

V. DISCUSSIONS AND IMPLICATIONS

An important feature of BZW-EF calculations, as compared to some BZW ones, consists of the agreement between the low, unoccupied energies obtained from the calculation with the optimal basis sets and those resulting from the calculation immediately following it, up to 10 eV. This has been the case for ZnO,⁹ BN,¹¹ and BP¹⁰ as examples. It is this feature that explains the exact reproduction of measured band gaps by the BZW-EF method, while some BZW results are sometimes 0.1 to 0.2 eV smaller than corresponding, experimental ones. Hence, in the vicinity of the Fermi energy (E_F), the Ekuma and Franklin (EF) enhancement makes a significant difference for non metals. Of course, the higher an excited energy, the less likely it is to be obtained from a ground state calculation; this fact stems in part from the use of finite numbers of basis functions. It is also due to the basis set and variational effect that precludes the use of very large basis sets containing the optimal ones, if one desires to preserve the physical content of low, unoccupied energies.

The thesis work of Williams^{27,34} indicated that the BZW method (or its enhancement) may not be needed for the description of the highest occupied and lowest unoccupied energies of materials known experimentally to be metals. The crossing of the Fermi energy by at least one band, in metals, simply explains this fact. The work of Williams showed that by the time the occupied energies reach their absolute minima, so do the low, unoccupied energies, beginning with the band that crosses the Fermi level. It is presumed that the particularly successful applications of DFT to the study of metals are partly due to this situation.

The work of Franklin et al.⁹ showed, with an explicit example, the fact that the energies from self consistent calculations may not represent the physical content of DFT. As per the last paragraph of Section II, the physical content of the results rests on the adherence of the calculations to the conditions inherent to DFT. Self consistent results for w-ZnO, with and without the 4d orbital on Zn^{2+} , are drastically different, even for the occupied bands. Indeed, the width of the upper most groups of valence band increases by 1.25 eV with the 4d orbital in the valence on zinc. Further, the zinc 3d bands are lowered by more than 0.25 eV.

There are other limitations of calculations that contribute to the deviation of their results from corresponding, experimental ones. The case of indium nitride illustrates this point. LDA BZW calculations of Bagayoko et al. obtained the experimental band gaps for wurtzite and cubic indium nitrides, w-InN¹⁷ and c-InN⁴, respectively. For w-InN, our calculated total width of the valence bands also agrees with experiment. We did not invoke self-interaction correction and saw no evidence of pd repulsion (other than the content of the Hamiltonian). We recall that Bagayoko et al. found 0.88 and 0.65 eV for the band gaps of w-InN and c-InN, respectively. In contrast, Furthmüller et al.³⁵ found -0.21 and -0.36 eV as the gaps for the former and the latter, respectively. Their calculations with the 4d electrons frozen in the core, using self interaction correction (SIC) pseudopotentials, led to gaps of 0.58 and 0.43 eV for w-InN and c-InN, respectively. These values

are less than half of corresponding results from similar SIC calculations with the 4d electrons in the valence.³⁵ *The drastically different outcomes of two similar SIC calculations suggest that issues other than DFT may affect computational results.* We do not discuss the adjustments of LDA results by Furthmüller et al. for some quasiparticle properties in light of our focus on DFT, a ground state theory. Additionally, Furthmüller et al. found differences of 0.76 and 0.79 eV between the band gaps at Γ , for w-InN and c-InN, depending on whether or not the 4d electrons are explicitly treated (i.e., included in the valence). However, the corresponding changes in band gaps, for DFT BZW calculations,³⁶ are just 0.13 eV for both w-InN and c-InN. *The rather large (0.76 & 0.79 eV) and small (0.13 eV) changes in the band gaps, as calculated by Furthmüller et al.³⁵ and Bagayoko et al.³⁶, respectively, strongly suggests differences between the results of these calculations other than those due to DFT.* In stark difference with the small change (0.13 eV) in the band gaps, when the 4d electrons are moved from the core to the valence, LDA BZW valence band widths increase by more than 1 eV for both materials. The inclusion of the 4d electrons in the valence is needed for a correct description of InN^{4,27,36}. This need to include the d electrons in the valence is a key result obtained by Wei and Zunger³⁷ in their study of II-VI semiconductors. *This case of InN illustrates our contention, borne out of the derivation of DFT, that there may be shortfalls of some calculations, and not of DFT, that are ascribed to DFT.*

The importance of the band gap partly explains our emphasis on it in this overview of DFT and our description of the BZW-EF method. Many applications of semiconductors and insulators hinge on the nature (direct or indirect) and numeric value of the band gap. Additionally, once calculations produce an inaccurate band gap, then their outputs cannot be utilized to calculate dielectric functions, densities of states, or effective masses that agree with experiments – without adjustments. Peaks in the two former quantities will disagree with corresponding, experimental ones and the underestimation of the band gap is known to reduce the electron effective mass that is consequently underestimated. As early as 1998, the BZW method led to calculated dielectric functions in agreement with experiment for BaTiO₃.²⁶ For most of the semiconductors studied with the BZW or BZW-EF method, total and partial densities of states and effective masses (for electron and holes) have been calculated. Structural predictions for cubic materials^{4,12} have been confirmed by experiment. For several of these materials, peaks in the density of states agreed well with experimental ones.

Two important, recent developments in the utilization of density functional theory are not discussed above. The first one is the introduction of the very successful modified Becke and Johnson potential (mBJ) by Tran and Blaha,³⁸ with subsequent improvements,³⁹ and the second consists of simulations of excited states by Jarlborg et al.^{40,41} using constrained total-energy calculations. There are LDA and GGA versions of the modified Becke and Johnson potential. For more than 40 different semiconductors, the Tran and Blaha mBJ (TB-mBJ) potentials have led to band gaps much closer to experimental ones as compared to results from most single basis set DFT calculations. As per our results in Table I, ab-initio DFT BZW and BZW-EF calculations generally yield results that agree more with experiment than corresponding ones obtained with TB-mBJ potentials. Several authors^{42–45} have examined the performance of TB-mBJ potentials. The good results for the band gaps seem to be accompanied by serious underestimations of the absolute value of the binding energy of valence d electrons in several systems, including ZnO,⁴⁵ GaN, GaP, ZnS, ZnSe, and CdS.⁴⁴ *As noted by the group of Blaha⁴² and by Jiang,⁴⁴ TB-mBJ cannot be obtained as the derivative of an energy functional, due in part to the definition of the parameter “c” utilized in the potential; its value is generally determined by fitting to experimental band gaps of selected materials so as to minimize the absolute relative error. Hence, in a fundamental sense, TB-mBJ potentials are not totally DFT potentials. As underscored by Jiang, the existence and uniqueness of a minimum energy resulting from calculations with these potentials are not necessarily guaranteed by the second Hohenberg Kohn theorem. For this reason, the content of this work, as in Sections I-IV above, does not necessarily apply to calculations with TB-mBJ potentials, despite the significant progress they have made possible.*

The simulations of the excited states of electrons or holes by Jarlborg et al. employ constrained total-energy calculations for systems with (N-1), N, and (N+1) electrons to arrive at relaxation energies which permit the correct description of the excitation process. Their calculations utilize

a local spin density approximation (LSDA) potential and linear muffin tin orbitals (LMTO). A central difficulty in determining the relevance of our discussions to this important development resides in the fact that we only addressed ab-initio calculations for the ground states of systems with fixed numbers of particles, while the intricate determination of relaxation energies, as described by Jarlborg et al., necessarily entails calculations of the total energies of systems with $(N-1)$, N , and $(N+1)$ electrons. The relaxed excitations energies for an electron and a hole are respectively the differences between E_t^{N-1} and E_t^N and between E_t^{N+1} and E_t^N , where the E 's are total energies and the superscripts are the numbers of electrons. The relaxation energy for the excitation process is defined as the difference between these excitation energies and ϵ_n^0 , the n th Kohn Sham orbital energy. Despite the noted comparison difficulties, it appears that future work could aid in elucidating a possible convergence. Indeed, we expect the relaxation energies obtained by using the BZW-EF method throughout the above calculations to be different from the ones resulting from calculations that utilize a single basis set. Even if the BZW-EF method is not employed in the calculations of the total energies, significant differences are expected depending on whether or not ϵ_n^0 is from a BZW-EF calculation or not. The above suggested comparisons are expected to be particularly informative if calculations with and without the BZW-EF method utilize the same DFT potential. We have studied magnetic and semi metallic materials in the past. We can therefore carry out spin polarized calculations in the future.

The two developments discussed serve in part to delineate the domain of validity of the content of this article. Indeed, as for the TB-mBJ potentials, this work does not apply to calculations using potentials that are not derivable entirely from DFT. Other examples of such potentials are LDA+C,⁴⁶ LDA+U,⁴⁷ DFT+U+V,⁴⁸ and several others. *This work, as per the derivation of DFT reviewed in Section II, only applies strictly to studies of the ground states of materials.* So, it does not apply to quasi particle energy calculations in any shape or form. As noted in connection with the work of Jarlborg et al., any comparison of our results with excited state energy calculations should be done very cautiously. Given the agreements between DFT BZW-EF calculated densities of states for both the occupied and the unoccupied states (up to 10 eV, at most, for the latter) and corresponding, experimental ones and that between the calculated imaginary and real parts of dielectric functions and measured ones, it appears that these calculations produce some results that are pertinent to excitation, beginning with optical transition energies. In this limited sense, we noted above how future calculations with the method could interface with the work of Jarlborg et al, on excitation and relaxation energies.

The implications of the accurate description and prediction of electronic and related properties of materials by DFT BZW and particularly BZW-EF calculations are diverse and some are profound. For the advancement of our fundamental understanding of materials, the method calls for revisiting some previous interpretations of DFT results, using only the findings of ab-initio, self consistent calculations whose results truly represent the physical content of DFT. In that process, the development of new potential schemes aimed at solving the energy gap and band gap problem is expected to be drastically reduced. With the capability for accurate descriptions of properties of materials, DFT BZW-EF calculations can inform and guide the design and fabrication of semiconductor based devices. The predictive capability of DFT BZW-EF calculations portends many benefits for the Materials Genome Initiative (MGI) in terms of the search for materials (i.e., molecules, nanostructures, semiconductors, and insulators) with desired properties. Specifically, this capability enables theoretical band gap engineering, for ternary and quaternary systems as examples. With the ubiquity and non-negligible nature of quantum effects at the nanoscale, the method is expected to play a key role in the growth of both the science of nanostructures and the market driven nanotechnology.

VI. CONCLUSION

We followed the works of Hohenberg and Kohn and of Kohn and Sham to articulate the derivation of density functional theory (DFT) and of its local density approximation (LDA). The derivation elucidated clear conditions, summarized at the end of Section II, to which electronic

structure calculations and related ones have to adhere in order for their results to represent the physical content of DFT, in general, or of LDA, in particular. Stark differences between results of some 99 calculations that woefully underestimated band gaps of materials and those obtained with the BZW or BZW-EF method posed a question. This question was partly answered by the lack of any evidence, as per the contents of the 99 publications, of a strict adherence to the referenced, necessary conditions. Hence, disagreements between the results of these calculations and corresponding, experimental ones do not necessarily signify any failures or shortfalls of DFT. The referenced, necessary conditions have to be met by a calculation before its results can represent the physical content of DFT. As to the reasons that DFT BZW and BZW-EF calculations accurately reproduce experimental results and make predictions confirmed by experiment, the first part of the answer was the strict adherence of these calculations to the conditions inherent to the derivation of DFT and of LDA. In retrospect, Part I of the method relative to the minimization of the energy was prescribed by Hohenberg and Kohn in 1964, except for the needed use of successively augmented basis sets! *By showing the necessity for the variations associated with that of size and other attributes of the basis set, in addition to variations of linear expansion coefficients in the self-consistency process, Part I add to DFT in practice.* Part II of the BZW and BZW-EF method consists of utilizing the Rayleigh theorem to avoid mathematical artifacts that arise when basis sets much larger than the optimal ones are employed in ab-initio, self consistent calculations. This Part II of the method is critical in the study of all materials that are not known to be metals, due to their energy or band gaps. Hence, the study of atoms, molecules, clusters and other finite structures, nanostructures, semiconductors, and insulators rests in part on this second part of our method that complements its first part. *By preserving the physical content of low laying energy levels, Part II of the method also adds to or completes DFT in practice.*

ACKNOWLEDGMENTS

Work funded by the US National Science Foundation (NSF) and the Louisiana Board of Regents, through LASiGMA [Award Nos. EPS- 1003897, NSF (2010-15)-RII-SUBR] and NSF HRD-1002541, the US Department of Energy, National Nuclear Security Administration (NNSA) (Award Nos. DE-NA0001861 and DE-NA0002630), LONI-SUBR, and the Ministry of Expatriate Malians, through the Project of Transfer of Knowledge through Expatriate Nationals (TOKTEN).

- ¹ P. Hohenberg and W. Kohn, *Phys. Rev.* **136**(No.3B), B 864 (1964).
- ² W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ³ W. Kohn, *Rev. Mod. Phys.* **71**, 1253 (1999).
- ⁴ Diola Bagayoko, Lashounda Franklin, and G. L. Zhao, *J. Appl. Phys.* **96**, 4297–4301 (2004).
- ⁵ H. Jin, G. L. Zhao, and D. Bagayoko, *Phys. Rev.* **B73**, 245214 (2006).
- ⁶ B. Khamala, L. Franklin, Y. Malozovsky, A. Stewart, H. Saleem, and D. Bagayoko, manuscript to be submitted for publication and to be available on ArXiv.
- ⁷ Yacouba Issa Diakité, Sibiry D. Traoré, Yuriy Malozovsky, Bethuel Khamala, Franklin Lashounda, and Diola Bagayoko, manuscript submitted for publication and available on ArXiv, September 2014.
- ⁸ C. E. Ekuma and D. Bagayoko, *Japanese Journal of Applied Physics (JJAP)* **50**, 101103 (2011).
- ⁹ L. Franklin, G. L. Zhao, C. E. Ekuma, and D. Bagayoko, *Journal of Physics and Chemistry of Solids* **74**, 729–736 (2013).
- ¹⁰ J. I. Ejembi, I. H. Nwigboji, L. Franklin, Y. Malozovsky, G. L. Zhao, and D. Bagayoko, *J. Appl. Phys.* **116**, 03711 (2014).
- ¹¹ A. Stewart, D. Hart, B. Khamala, Y. Malozovsky, and D. Bagayoko. Manuscript submitted for publication and to be available on ArXiv in October 2014.
- ¹² D. Bagayoko and G. L. Zhao, *Physica C* **364-365**, 261–264 (2001).
- ¹³ R. G. Eggedell, V. E. Henrich, R. Bowdler, and T. Sekine, *J. Appl. Phys.* **94**, 6611 (2003).
- ¹⁴ J. Z. Jiang, H. Lindelov, L. Gerward, K. Stahl, J. M. Recio, P. Mori-Sanchez, S. Carlson, M. Mezouar, E. Dooryhee, A. Fitch, and D. J. Frost, *Phys. Rev. B, Rapid Communications* **65**, 161202(R) (2002).
- ¹⁵ J. Schörmann, D. J. As, K. Lischka, P. Schley, R. Goldhahn, S. F. Li, W. Löffler, M. Hetterich, and H. Kalt, *Appl. Phys. Lett.* **89**, 261903 (2006).
- ¹⁶ Batakushna Santara, P K Giri1, Kenji Imakita, and Minoru Fujii, *J. Phys. D: Appl. Phys.* **47**, 215302 (2014).
- ¹⁷ Diola Bagayoko and Lashounda Franklin, *J. Appl. Phys.* **97**, 123708 (2005).
- ¹⁸ J. Wu *et al.*, *J. Appl. Phys.* **94**, 4457 (2003) and *Appl. Phys. Lett.* **80**, 3967 (2002).
- ¹⁹ J. P. Perdew and Alex Zunger, *Phys. Rev.* **23**(No. 10), 5048 (1981).
- ²⁰ John P. Perdew and Mel Levy, *Phys. Rev. Lett.* **51**(No. 20), 1884 (1983).
- ²¹ L. J. Sham and M. Schlüter, *Phys. Rev. Lett.* **51** No. 20, 1888 (1983).
- ²² L. J. Sham and M. Schlüter, *Phys. Rev. B* **32**, 3883 (1985).
- ²³ D. Bagayoko, *Inter. Jour. Quant. Chem.* **17**, 527 (1983).

- ²⁴ David C. Langreth and John P. Perdew, *Phys. Rev. B* **21**, 5469 (1980).
- ²⁵ B. N. Harmon, W. Weber, and D. R. Hamann, *Phys. Rev. B* **25**(2), 1109–1115 (1982).
- ²⁶ D. Bagayoko, G. L. Zhao, J. D. Fan, and J. T. Wang, *J. of Physics: Condensed Matter* **10**, 5645 (1998).
- ²⁷ G. L. Zhao, D. Bagayoko, and T. D. Williams, *Physical Review* **B60**, 1563 (1999).
- ²⁸ C. E. Ekuma, Master's Degree Thesis in Physics (Southern University and A&M College in Baton Rouge (SUBR, Baton Rouge, LA 70813).
- ²⁹ C. E. Ekuma, M. Jarrell, J. Moreno, and D. Bagayoko, *AIP Advances* **2**, 012189 (2012).
- ³⁰ D. Bagayoko, in *Proceedings, Malian Symposium of Applied Sciences (MSAS), Bamako, Mali* (2004) pp. 53–58.
- ³¹ C. E. Ekuma, M. Jarrell, J. Moreno, G. L. Zhao, and D. Bagayoko, *Phys. Lett. A* **Vol. 377**(34-36), 2172–2176, 1 November 2013 - <http://authors.elsevier.com/sd/article/S0375960113005318>.
- ³² I. H. Nwigboji, J. I. Ejembi, Y. Malozovsky, B. Khamala, L. Franklin, G. L. Zhao, C. E. Ekuma, and D. Bagayoko, manuscript submitted for publication and available on ArXiv, September 2014.
- ³³ D. Bagayoko, L. Franklin, G. L. Zhao, and H. Jin, *J. Appl. Phys.* **103**, 096101 (2008).
- ³⁴ T. D. Williams, M. S. thesis (Southern University and A&M College in Baton Rouge (SUBR), Louisiana, 1998).
- ³⁵ J. Furthmüller, P. H. Hahn, F. Fuchs, and F. Bechstedt, *Phys. Rev. B* **72**, 205106 (2005).
- ³⁶ D. Bagayoko, L. Franklin, and G. L. Zhao, *Phys. Rev. B* **76**, 037101 (2007).
- ³⁷ S. H. Wei and A. Zunger, *Phys. Rev. B* **37**, 8958 (1988).
- ³⁸ F. Tran and P. Blaha, *Phys. Rev. Lett.* **102**, 226401 (2009).
- ³⁹ D. Koller, F. Tran, and P. Blaha, *Phys. Rev. B* **85**, 155109 (2012).
- ⁴⁰ T. Jarlborg, B. Barbiellini, H. Lin, R. S. Markiewicz, and A. Bansil, *Phys. Rev. B* **84**, 045109 (2011).
- ⁴¹ T. Jarlborg, B. Barbiellini, C. Lane, Y. J. Wang, R. S. Markiewicz, Z. Liu, Z. Hussain, and A. Bansil, *Phys. Rev. B* **89**, 165101 (2014).
- ⁴² D. Koller, F. Tran, and P. Blaha, *Phys. Rev. B* **83**, 195134 (2011).
- ⁴³ J. A. Carmago-Martinez and R. Baquero, *Phys. Rev. B* **86**, 195106 (2012).
- ⁴⁴ H. Jiang, *J. Chem. Phys.* **138**, 134115 (2013).
- ⁴⁵ B. Ul Haq, R. Ahmed, S. Goumri-Said, A. Shaari, and A. Afaq, *Phase Transitions: A Multinational Journal* **Vol. 86**(No. 12), 1167 (2013).
- ⁴⁶ Su-Huai Wei and Alex Zunger, *Phys. Rev. B* **57**, 8983 (1998).
- ⁴⁷ V. I. Anisimov, J. Zaanen, and Ole K. Andersen, *Phys. Rev. B* **44**, 943 (1991).
- ⁴⁸ V. L. Campo, Jr. and M. Cococcioni, *J. Phys.: Condens. Matter* **22**, 055602 (2010).