A Mathematical Solution to the Theoretical Band Gap Underestimation: *Predictive Calculations of Properties of Semiconductors*

D. Bagayoko, Ph.D.

Southern University System Distinguished Professor of Physics
Chairman, Department of Physics
Southern University and A&M College in Baton Rouge (SUBR),
Baton Rouge, Louisiana 70813, USA

A Presentation at the Institute of Condensed Matter and Material Physics, School of Physics, Peking University

March 25, 2010

ACKNOWLEDGMENTS

- **We profoundly thank Professor Guang-Shan Tian for the invitation and Professor Wei Guo for Arrangements relative to my stay in Beijing.**

- My collaborators for the work presented here include: **Dr. G. L. Zhao, Ms. Lashounda Franklin, Dr. Troy D. Williams, Mr. Y. Luo, Dr. Hua Jin, and Mr. E. C. Ekuma.**

- Research work funded in part by the Louisiana Optical Network Initiatives (**LONI**), the Department of the Navy, Office of Naval Research (**ONR**, Award Nos. N00014-05-1-0009 and N00014-05-1-0830), the National Science Foundation (**NSF**), and **NASA** [Award Nos. NCC13-00010, NCC2-1344, and NNG 05G146G].

OUTLINE

I. INTRODUCTION: A brief Historical Overview of Electronic Property Calculations and Related Approximations and Limitations

II. THE BAND GAP UNDERESTIMATION: Serious discrepancies between calculated and measured values of energy & band gaps of materials & between Schemes aimed at resolving the problem

III. MOTIVATIONS: To resolve the above discrepancies and to predict, *from first principle calculations*, E-gaps and other electronic properties of materials

IV. METHOD: THE BAGAYOKO, ZHAO, AND WILLIAMS (BZW) METHOD
General: LCAO (LCGO) and **DFT** & particularly **LDA** Potentials
Specific: the Bagayoko, Zhao, and Williams **(BZW)** Method

J. Phys. Condens. Matter 10, 5645, 1998 for BaTiO3
Phys. Rev. B60, 1563, 1999 - - Diamond, Si, and GaN

V. SOME ILLUSTRATIVE RESULTS: For GaN, Si, Diamond, SWCNT, SiC (3C & 4H), ZnO, ZnSe, AlN, Si_3N_4, Ge, InN, InAs, GaAs, AlAs (E(**k**), Gaps, Effective Masses, etc.)

VI. CONCLUSION: **DFT and LDA potentials,** contrary to some previous and understandable beliefs, correctly describe electronic and related properties of semiconductors, including band gaps, if one implements the BZW Method.

I. HISTORICAL OVERVIEW: THE FUNDAMENTAL EQUATION

Note well that these equations are non-relativistic – hence, they are not the ultimate equations. They are approximations, the best we know!

$$\frac{}{} \quad {}_{i} \quad {}_{A\,i}\,\frac{Z_A}{r_{Ai}} \quad {}_{A\;\;B}\,\frac{Z_A Z_B}{R_{AB}} \quad {}_{i\;\;j}\,\frac{}{r_{ij}} \quad \vec{r}\;\;\vec{R} \quad E_{el} \quad \vec{r}\;\;\vec{R}$$

The Born-Oppenheimer approximation applied to the above equation leads to the following (note the absence of R in “Hamiltonian” and in the wave function):

$$\frac{}{} \quad {}_{i} \quad {}_{A\,i}\,\frac{Z_A}{r_{Ai}} \quad {}_{i\;\;j}\,\frac{}{r_{ij}} \quad \vec{r} \quad E_{el} \quad \vec{r}$$

HISTORICAL OVERVIEW (early 1930s): THE HARTREE FOCK EQUATIONS

$$- \qquad {}_{A} \; \frac{Z_A}{\left|\vec{r} \quad \vec{R}_A\right|} \; {}^{N}_{i} \qquad {}_{i}\,\vec{r} \qquad \frac{}{\left|\vec{r} \quad \vec{r}\right|} \; {}_{i}\,\vec{r} \qquad d\,\vec{r} \qquad {}_{k}\,\vec{r}$$

$$ {}^{N}_{i} \qquad {}_{i}\,\vec{r} \qquad \frac{}{\left|\vec{r} \quad \vec{r}\right|} \; {}_{i}\,\vec{r} \qquad {}_{k}\,{}_{H}\vec{r} \qquad d\,\vec{r} \qquad {}_{k}\;{}_{k}\,\vec{r}$$

These non-linear, integro-differential equations embed the following approximations [not eigenvalue equations of the form H **]**

- Neglect of relativistic effects
- Born-Oppenheimer approximation (no nuclear motion, decoupling of phonon & electronic properties)
- Single Slater determinant assumed sufficient (see Configuration Interactions – CI)
- An electron is assumed to move in the means fields created by the others
- The central field approximation (in practice)

HISTORICAL OVERVIEW: DENSITY FUNCTIONAL EQUATIONS – LOCAL DENSITY APPROXIMATION (LDA) - Starting in 1964 & 1965

EQUATION 1 (Only for the ground state)

$$- \qquad V \ \vec{r} \qquad \frac{n \ \vec{r}}{\left| \vec{r} \qquad \vec{r} \right|} d\vec{r} \quad V_{xc} \ n \ \vec{r} \qquad {}_{i} \ \vec{r} \qquad {}_{i} \quad {}_{i} \ \vec{r}$$

Equation 2 (Sum over occupied states only)

$$n \ \vec{r} \qquad {}^{N} \qquad \vec{r} \qquad \vec{r}$$

Sum over occupied states only

II. THE BAND GAP UNDERESTIMATION: DEFINITION OF TERMS

In finite systems, like atoms, molecules, clusters, etc., the energy levels occupied by electrons are discrete. For the hydrogen atom, they are given by the negative of a constant over n^2, where n is the principal quantum number.

The energy gap of a discrete system is the lowest, unoccupied energy value minus the highest occupied energy value.

In crystalline solids, by virtue of the Bloch theorem, the energy levels of electrons are functions of a continuous parameter $\vec{k}$. Hence, they form bands.

The band gap for solids is the lowest, unoccupied energy value ($E_{min,c}$) minus the highest occupied energy value ($E_{max,v}$). It can be a **direct** gap (If $E_{min,c}$ and $E_{max,v}$ occur at the same $\vec{k}$), it is otherwise **indirect** .

Metals are the only materials with no band gap. Finite systems (atoms, etc.), semiconductors, and insulators have energy or band gaps that are important for most of their electronic properties and related properties (i.e., optical ones).

THE BAND GAP UNDERESTIMATION (1a)

Illustrative Discrepancies between the Measured and Calculated Band Gaps of ZnO in the Wurtzite Structure

THE BAND GAP UNDERESTIMATION (1b)

THE BAND GAP UNDERESTIMATION (1c)

Illustrative Discrepancies Between the Theoretical Values of the Band Gap of Cubic Indium Nitride (c-InN)

Potential	Computational Method	a (Å)	B(GPa)	E_g (eV)
Local Density Approximation (LDA) Potentials	Pseudopotential Method (PP)	4.95 5.004	145 140	**-0.36[a], -0.40[b]** **-0.35[e], -0.18[g]**
	LAPW	4.94	145	
	Full Potential LAPW	5.03	138	**-0.11[h] & -0.48[h]**
	Full Potential LMTO			**-0.4[i]**
		4.92	139[j]	
	Atomic Sphere Approximation			**-0.1[i], +0.02 and +0.08[k]**
Generalized gradient approximation (GGA)	PP	5.06 & 5.109	120[c] 118	**-0.55[b]**
LDA plus SIC		5.05[d]		**+0.43[a]**
QP Calculation	PP			**+0.52[a]**

PLEASE SEE MORE ON THE NEXT PAGE

THE BAND GAP UNDERESTIMATION (1c’)

Illustrative Discrepancies Between the Theoretical Values
of the Band Gap of Cubic Indium Nitride (c-InN)

Potential	Computational Method	a (Angstroms)	B(GPa)	E_g (eV)
QP+SIC	PP			**+1.31**[a]
DFT Exact Exchange				**+1.4**[g]
DFT, SX	ASA			**+1.3**[i]
Estimate of the bulk modulus of c-InN			137[l]	
Empirical Pseudopotential Calculations (EMP)				**+0.592**[m]
Experimental: Measured lattice constants		4.97 ± 0.01[n] 4.98[0] & 4.986[p]		

NOTE: Other LDA Values of Eg: from -0.40 to + 0.08 for C-InN
QP, EXX, SX Values: From +1.3 to 1.4 eV
LDA-BZW Values: **0.65 eV a = 5.017** Angstroms
Experimental Values: **0.61** eV and a = **5.01 +/- 0.01** Angstroms

THE BAND GAP UNDERESTIMATION (1d)

Illustrative Discrepancies Between **(a)** Measured Values and **(b)** Calculated Values, and **(c)** Measured and Calculated Values of **the Band Gap of Wurtzite Indium Nitride (w-InN).**

	EXP Group 1	**EXP** Group 2	LDA PP	LDA FP-LMTO	GGA PP	GWA or QP (pp)
w-InN →	**1.9-2.0 eV**	**0.7-1.0 eV**	-0.4 to -0.2 eV	0.2 to 0.43 eV	-0.37 eV	.74 &.5 eV
c-InN →	Not Available	Not Available	-0.18 to –0.40 eV	-0.1 eV	-0.55 eV	0.52 &1.31

<u>**NOTE**</u>: LDA-BZW RESULTS: **0.88 eV,** exactly as experiment with same charge carrier concentration (Burstein-Moss Effect)

Exp: Experiment (group 1 before 2000, Group 2 after 2000);
LDA: Local Density Approximation;
GWA: Green Function and Screened Coulomb Approximation;
PP: Pseudopotential Method;
FP-LMTO: Full potential linear muffin tin orbital method

THE BAND GAP UNDERESTIMATION (2a): Explanations of the sources of discrepancies

In addition to the approximations inherent to the above discussed equations, theorists put forward explanations of the sources of the discrepancies for DFT calculations:

- DFT is reported to contain an error stemming for **having an electron interact** with itself (Perdew and Zunger, 1981)–They devised a ***self-interaction* correction** *.*

- **Derivative discontinuities** (John Perdew et al.1982) of the exchange correlation energy (Exc) in DFT (i.e., discontinuity of Vxc).

- **Derivative discontinuities** of Exc and of the of the kinetic energy (Ks) of non-interacting particles as given in the Kohn-Sham equation (Perdew and Levy, 1983). They assert that the discontinuity of Vxc is non-zero in semiconductors. **(We could not find a proof of this assertion.)**

- **Derivative discontinuities** of Vxc and of Ks (Sham and Shlüter, 1983). They state that they do not know if the discontinuity of Vxc is non-zero. They do not know the value of the discontinuity and do not claim it to be non-zero.

THE BAND GAP CATASTROPHE (2b): The proliferation of schemes (mostly ad hoc)

As we stated in some J. Appl. Phys. publications, the continuing proliferation of schemes purporting to resolve the band gap problem resembles that of epicycles for the Ptolemaic “Earth” system. (Well known for ~2000 years!)

It seems that by citing the above papers, authors can do practically anything and get it published: for them **“It is well known that density functional theory underestimates the band gap of materials.”** This practice continues even though a mathematical solution to the problem was provided by BZW in 1998 and 1999 – **<u>This situation and the gap problem constitute a catastrophe</u>.**

Some articles at the end of this presentation describe some of these schemes that include:

GW Approximation (mostly ab-initio)

SIC

LDA +C

EXX

TDDFT (for excited states, ~ ab-initio)

GGA and **meta-GGA** (mostly ab-initio)

LDA + Wigner Interpolation

LDA + U (U is from Hubbard theory)

sX (X screened)

WDA (weighted DF Approximation)

Hybrid Functionals (Chemistry), **Scissors approximation, gKS, and more**

III. MOTIVATIONS

The key motivation for our work was simply to resolve the above described energy gap and band gap problems and related used of <u>EXTENSIVE</u> talents and times to produce ad-hoc results with little or no predictive value.

The quintessential importance of correct **energy gaps**, correct **effective masses**, and of correct **optical,** and **related properties** of materials simply warrants any effort to get these quantities from first-principle calculations – TO USHER IN AN ERA OF TRUE PREDICTIVE CALCULATIONS OF PROPERTIES OF MATERIALS (ATOMS, MOLECULES, CLUSTERS, SEMICONDUCTORS, INSULATORS, & NUCLEI).

An added motivation was that, **at the nanoscale**, quantum effects are both ubiquitous and non-negligible. Hence, getting these effects correctly was deemed to be pivotal for further progress.

IV. METHOD

A. THE LCAO (LCGO) FORMALISM

THE EIGENVALUE EQUATION, $H \quad E$, IS SOLVED SELF-CONSISTENTLY BY TAKING

$$\overset{N}{\underset{i}{}} \; a_i \quad {}_{i},$$ i = 1, N

The $_i$ are obtained from calculations the atomic or ionic species that are present in the system (including radial & angular features)

We employed **Gaussian orbitals** (LCGO) for 0 28e2i (funclculatinvolvmploi (5.3

IV. METHOD (cont’d)

C.1. THE RAYLEIGH THEOREM

Let

The linear combination of atomic orbital (LCAO) method leads to:

$$^{N}_{i} \quad a_i \qquad _i$$, i = 1, N

Let $^{N}_{i}$ be the **self-consistent** eigenvalues, **ordered from the lowest** ^{N} **to** the highest $^{N}_{N}$

Let the above equation be solved with *(N+1) orbitals* *<u>obtained by adding</u>* *one orbital to the* *<u>previous set</u>* *of N orbitals*. And let be the new self-consistent

eigenvalues, $_i^{N}$, be ordered from the lowest, ^{N} ,to the highest. These eigenvalues are obtained with the <u>larger</u> basis set of (N+1) orbitals.

Then, by the Rayleigh Theorem, $_i^{N}$ $^{N}_{i}$ **for** $i \quad N$

IV. METHOD (cont'd)

C.2. THE BASIS SET AND VARIATIONAL EFFECT (source of gap probl.)

Using the Rayleigh theorem, BZW identified an **intrinsic, basis set and variational effect as follows.**

This effect is partly due to the fact that **only the wave functions of the occupied states** are including in the construction or reconstruction of the Hamiltonian (**i.e., the physics**) in going from one iteration to the next.

The potential is basically unchanged once the occupied energies converge (in values, branching, & curvatures). However, adding more orbitals will still lower some *unoccupied* levels by virtue of the Rayleigh theorem!

Such **an "extra" lowering is the effect.** It has been ascribed, for decades, to some limitations of LDA. *For details and illustrations, please see Inter. J. Quant. Chem., 17, 527, 1983; J. Phys. Condensed Matter 10, 5645, 1998; and Phys. Rev. B60, 1563, 1999.*

Explanation: *Researchers have been solving one equation out of a system of* ***two coupled equations****. Mathematically, that practice is simply incorrect.*

[redacted]

C.3. IMPLEMENTING THE BZW METHOD

Three or more self-consistent calculations are generally needed. Calculation I employs the minimum basis set (MB Set) needed to just account for all the electrons.

In Calculation II,.1(2o3(th( )6 (M5(sen)9.asiaugmsteedewithee bital(sđ))6 represstei

V. ILLUSTRATIVE RESULTS

Figure 1: **LCGO-LDA-BZW band structures of ZnO** resulting from Calculation III (_____) and Calculation IV (------), *with the Fermi levels from the two calculations superimposed.* **The calculated band gap of 3.47 eV is practically the same as the experimental one.** The extra lowering of the conduction bands in Calculation IV sems from the Basis Set and Variational Effect. Larger basis sets lower them further, while the occupied energies remain unchanged (i.e., at their minima).

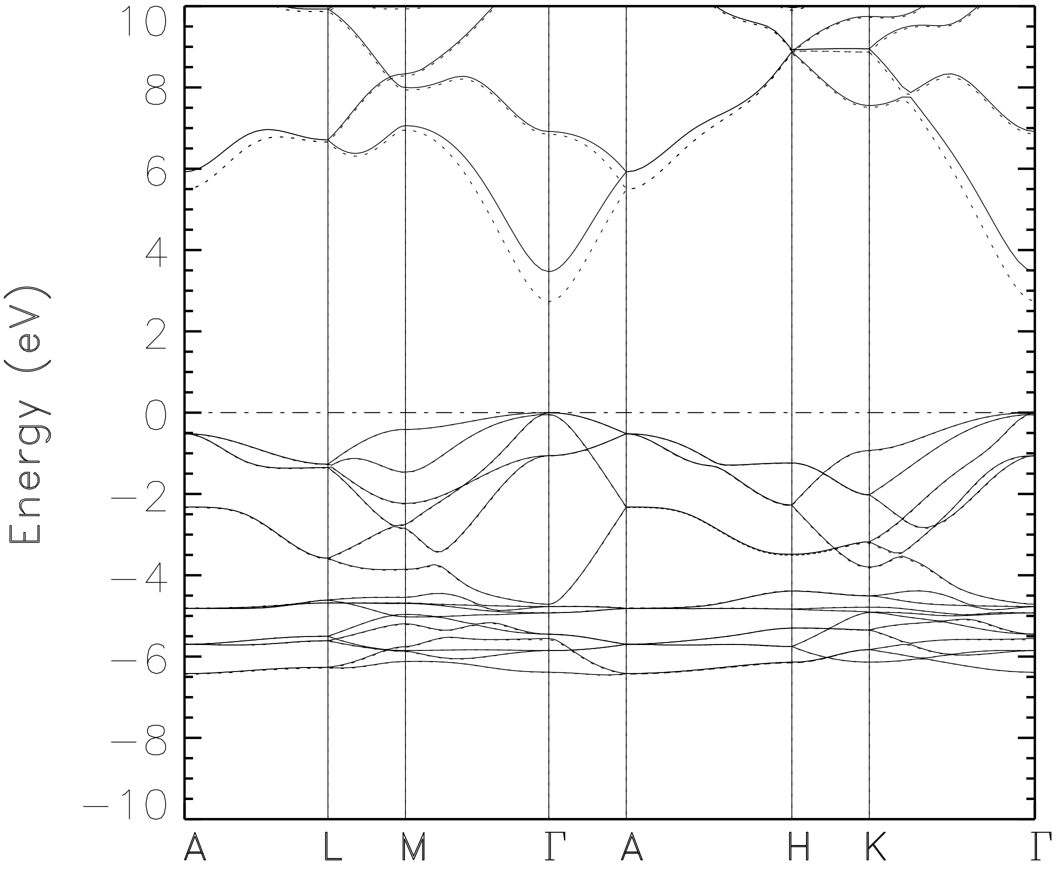


TABLE 1. Calculated Properties of Semiconductors versus Measured Values. [Computations done in accordance with the BZW procedure (i.e., optimal basis sets were obtained and utilized)]

	Calculation	**Measurement**
$BaTiO_3$		
M_p		
M_p		
M_n		
M_n		
GaN		
M_n		

	Calculation	**Measurement**
Diamond		
M_n		
M_n		
3C-SiC		

TABLE 2. Calculated Properties of Semiconductors versus Measured Values. [Computations done in accordance with the BZW procedure (i.e., optimal basis sets were obtained and utilized)]

Si	**Calculation**	**Measurement**
M_{nt}		
M_{nl}		
ZnO		
AlN		

4H-SiC	**Calculation**	**Measurement**
M_n		
M_n		
M_M		
M_{MK}		
c-Si_3N_4		
ZnSe		

TABLE 3. Calculated Properties of Semiconductors versus Measured Values. [Computations done in accordance with the BZW procedure (i.e., optimal basis sets were obtained and utilized)]

w-InN	**Prediction**	**Measurement**	**c-InN**	**Predictions**	**Measurement**
	0.88 eV			**0.65 eV**	
				a=5.017 Å	
AlAs	**Calculation**	**Measurement**	**InAs**	**Calculation**	**Measurement**

w-CdS: Calc. Eg = 2.47 Exp. 2.5 eV **zb-CdS**: Calc. 2.39 eV Exp. 2.42 eV

Sodium Nitrite: Calc. Eg = 3.18 eV Exp. 3.14 – 3.22 eV

TiO_2	**Calculation**	**Measurement**	**Ge**	**Calculation**	**Measurement**
Eg	2.7 eV	3.0 eV	Eg	0.62	0.66-0.74
GaAs	**Calculation**	**Measurement**	**c-Si_3N_4**	**Prediction**	**Measurement**
Eg	1.24	~ 1.4 – 1.5 eV	Eg	**3.68 eV**	3.66 – 3.7 eV

Imaginary Part of the Dielectric Function – xy Component – for Experiment with Eg < 1 eV

Figure 2.

Experiment with Eg < 1 eV

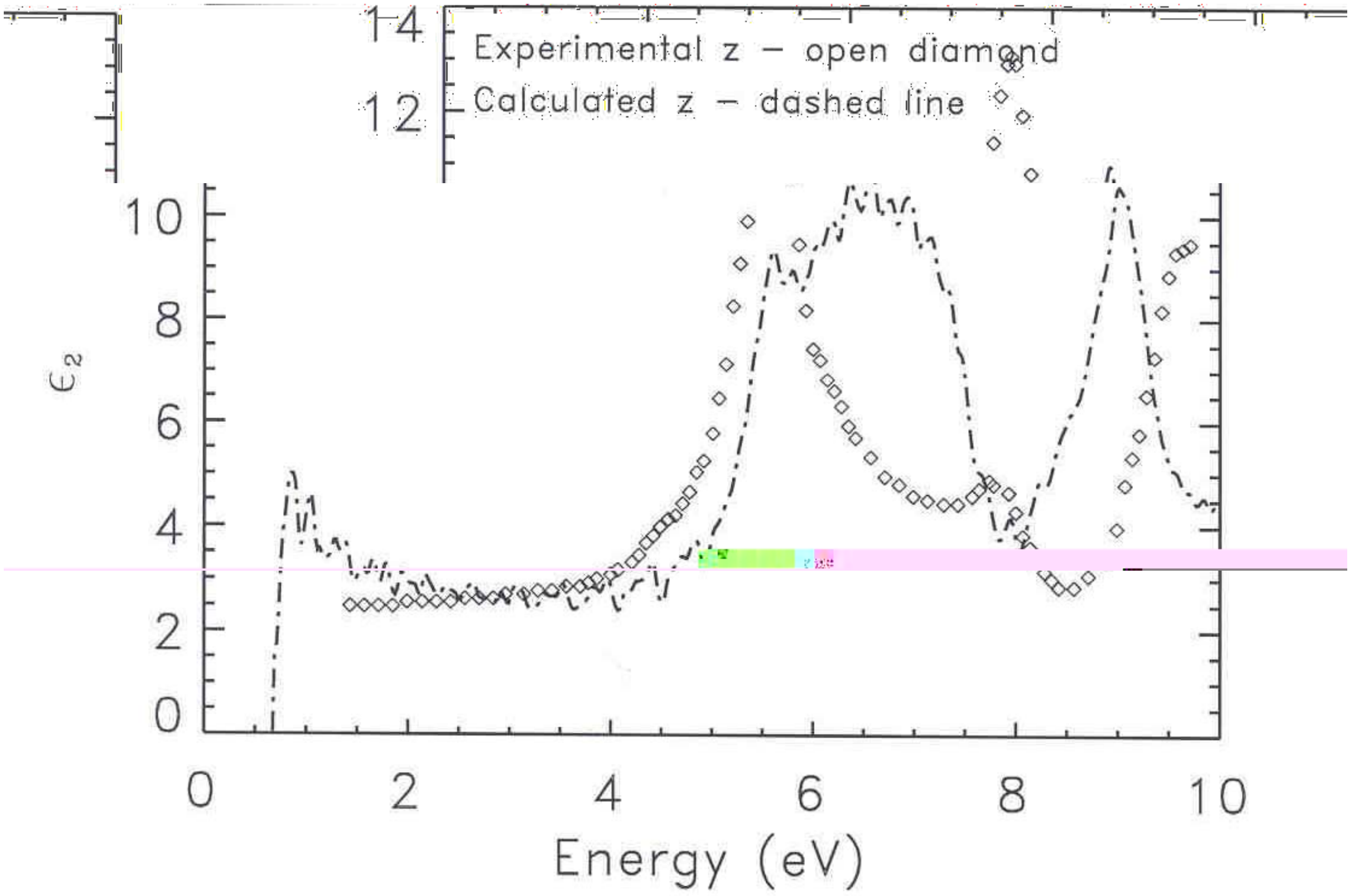


Figure 3.

Experiment with Eg < 1 eV

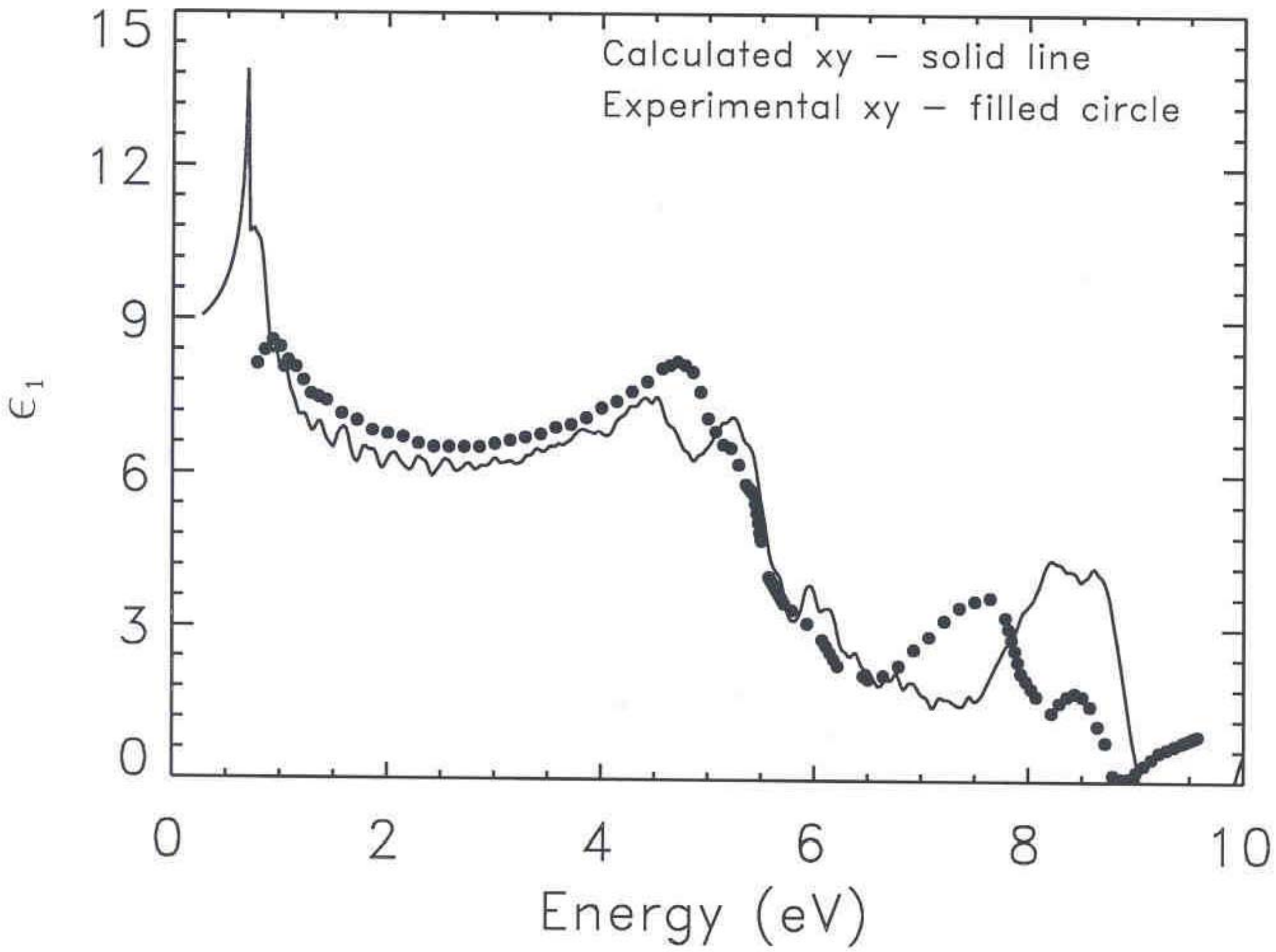
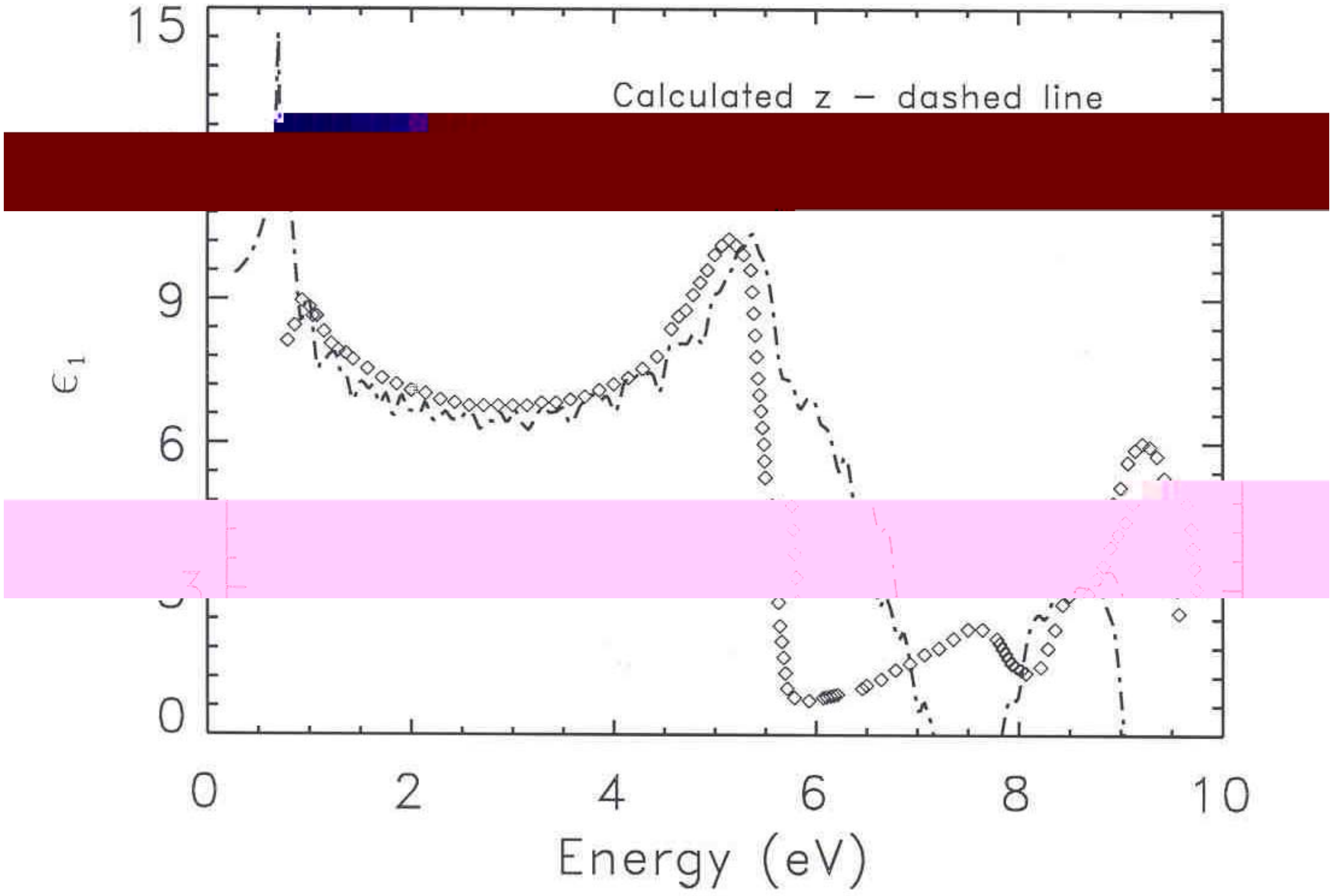


Figure 4.

Experiment with Eg < 1 eV



VI. CONCLUSION

LDA-BZW calculations eliminate much of **the believed or perceived limitations** of density functional theory (DFT), with emphasis on the local density approximation (LDA).

Specifically, **low-energy conduction bands** and the **band gaps** are correctly described, mostly within experimental uncertainties, and **predicted by LDA-BZW calculations.**

The

VI. CONCLUSION (Cont’d)

IMPLICATIONS OF THIS CAPABILITY:

Molecular engineering (including the prediction or design of novel molecules, small or large, inorganic or organic)

Semiconductor engineering (including the prediction or design of materials with desired band gaps)

Enhancement of simulations (partly through accurate inter-atomic potential or empirical pseudopotential parameters derived from BZW results)

Theoretical exploration, using the nuclear shell model, of possibilities for gamma ray amplification by stimulated emission of radiation (graser).

Fundamental progress in the understanding of materials: i.e., a determination of the true capabilities and limitations of DFT and LDA and of schemes purporting to correct LDA or to go beyond DFT altogether.

THANK YOU VERY MUCH FOR YOUR ATTENTION

Annotated Bibliography

Annotated Bibliography (Cont'd)

"Predictive Ab-initio Computations of Properties of Ferroelectric Materials," D. Bagayoko and G. L. Zhao, International Journal of Modern Physics B, Vol. 13, Nos. 29, 30, and 31, Pages 3767-3773 (1999). The industrial application of these materials partly underscores the utility of accurately predicting their electronic and related properties.

"Electronic Structure and Charge Transfer in 3C- and 4H-SiC," **G**. L. Zhao and D. Bagayoko. New Journal of Physics (NJP), 2, 16.1-16-12, 2000. The BZW method successfully described the named materials within the local density approximation (LDA). This article is available free of charge at www.njp.org (search for Bagayoko or Zhao).

"Effective masses of charge carriers in selected symmorphic and nonsymmorphic carbon nanotubes," G. L. Zhao, D. Bagayoko, and L. Yang, Phys. Rev. B 69, 245416, June 2004. As intimated in the conclusion of this presentation, the BZW method is believed to be indispensable in the theoretical description of nanomaterials for which (a) not only quantum effects are ubiquitous (a) but also are non-negligible.

"Predictions of Electronic, Structural, and Elastic Properties of Cubic InN." D. Bagayoko, L. Franklin, and G. L. Zhao, Journal of Applied Physics, Vol. 96, No. 8, Pages 4297-4301, October 2004. As per the title of this article, the theoretical results therein are predictions of experimental values that have not yet been measured!

Annotated Bibliography (Cont'd)

"Density Functional Theory Band Gap of Wurtzite InN." D. Bagayoko and Lashounda Franklin, Journal of Applied Physics 97, 123708, 2005.

"Density functional band gaps of AlAs." H. Jin, G. L. Zhao, and D. Bagayoko, Phys. Rev. B 73, 245214(2006).

"Calculated Optical Properties of Wurtzite InN." H. Jin, G. L. Zhao, and D. Bagayoko, J. Appl. Phys. 101, 033123 (2007). The excellent agreement with experiment, for energies up to 5.6 to 6 eV, denotes not only the correct description of the band gap, but also of the low-lying conduction bands! This conclusion also stems from agreements between our calculated optical transition energies and corresponding experimental ones for $BaTiO_3$ and other semiconductors.

Comment on "Band structures and optical spectra of InN polymorphs: Influence of quasiparticle and excitonic effects." D. Bagayoko, L. Franklin, H. Jin, and G. L. Zhao, Phys. Rev. B 76, 037101 (2007).

Comments on "Band gap and band parameters of InN and GaN from quasiparticle energy calculations based on exact-exchange density-functional theory" [Appl. Phys. Lett. 89, 161919 (2006)]. D. Bagayoko, L. Franklin, and G. L. Zhao. Accepted for publication in the Journal of Applied Physics.

THANK YOU VERY MUCH, AGAIN, FOR YOUR INTEREST