

Theoretical predictions of electronic materials and their properties

Alex Zunger

In this article, I first define the basic structure of modern 'first-principles theory of real materials' (including old references), and then I review recent applications to electronic materials. I argue that electronic structure theory of real materials has advanced to the point where bold predictions of yet unmade materials and of unsuspected physical properties are being made, fostering a new type of interaction with experimentalists. I review the basic characteristics of this new style of theory, illustrating a few recent applications, and express opinions as to future challenges.

Addresses

National Renewable Energy Laboratory, Golden, Colorado 80401, USA; e-mail: alex_zunger@nrel.gov

Electronic identifier: 1359-0286-003-00032

© Current Chemistry Ltd ISSN 1359-0286

Abbreviations

CBM	conduction band minimum
EMA	effective mass approximation
LAPW	linear augmented plane wave
LDA	local density approximation
LMTO	linear muffin tin orbital
VBM	valence band minimum
ΔH^f	formation enthalpy
ΔH^m	migration enthalpy
ΔE_v	valence band offset

Introduction

The scope and structure of first-principles theory of 'real materials'

aims at understanding material properties and processes from an atomistic quantum-mechanical point of view, retaining the complexity and specificity of actual solids, without losing track of the underlying global trends and basic physics.

What's done?

The basic structure of electronic structure theory is drawn from text-book quantum-mechanics. A 'system' (atom, molecule, solid, nanostructure) is defined via its 'external potential' $V_{ext}(\mathbf{r})$, which includes information on the atomic

comprising the system, plus, when appropriate, external (e.g. electric, magnetic) fields. One then sets up an effective single-particle Schroedinger equation:

$$\left\{ -\frac{1}{2} \nabla^2 + V_{ext}(\mathbf{r}) + V_{screening}(\mathbf{r}, \rho) \right\} \Psi_i(\mathbf{r}) = \epsilon_i \Psi_i(\mathbf{r}) \quad (1)$$

where $V_{screening}$ is the systems response to V_{ext} . The screening potential, depends on the density matrix ρ . Where $\{\psi_j\}$, $\{\epsilon_j\}$ are the single-particle wavefunctions and energies, respectively. Assuming at first a given geometry (hence, a given $V_{ext}(\mathbf{r})$) and using a specific microscopic model of screening (e.g. the local density approximation [LDA] [1]), one solves Equation 1 iteratively, using the boundary conditions appropriate to the system at hand (e.g. periodic bulk solids, isolated quantum dots, etc.). Iterations are needed because $V_{screening}$ depends on ρ and the latter is obtained from the wavefunctions ψ via $\rho(\mathbf{r}, \mathbf{r}') = \sum_j \psi_j^*(\mathbf{r}) \psi_j(\mathbf{r}')$. The sequence is; one assumes ρ , then computes $V_{screening}(\rho)$, then solves Equation 1 to find $\{\psi_j\}$, then recomputes a new ρ from the latter, and continues to self-consistency. When self-consistency is attained, for a

obtains the total electron + nuclear energy [2] of that geometry:

$$E_{tot}\{R_\alpha, \alpha = 1 \dots N\} = \sum_i \epsilon_i - \frac{1}{2} E_H - \frac{1}{4} E_{XC} + \frac{1}{2} E_{i-j} \quad (2)$$

where the first term is the sum of single particle energies, E_H is the electron–electron Coulomb (Hartree) energy, E_{XC} is the exchange–correlation energy, and E_{i-j} is the ion–ion energy. The forces on all N atoms are then obtained from E_{tot} ; one seeks the geometry that produces no forces:

$$\vec{F}_\alpha = \partial E_{tot} / \partial R_\alpha = 0 \quad (3)$$

What's the input and what's the output?

Given the inputs, the N atomic number $\{Z_\alpha\}$ and initial

the 'system' equilibrium geometry $\{R_\alpha\}$, its eigenvalue spectrum $\{\epsilon_i\}$, and total energy. Second derivatives of the total energy with respect to unit cell deformations yield elastic constants $\{C_{ij}\}$, whereas second derivatives with respect to collective atomic displacements yield force constants, hence the phonon spectrum. Bonding information is gleaned from the charge-density.

$$\rho(\mathbf{r}) = \sum_j \psi_j^*(\mathbf{r}) \psi_j(\mathbf{r})$$

Response to external perturbations (pressure, strain,

rotation for the perturbed system, or the linear response methods [3]. This yields quantities such as pressure equation of states, strain deformation potentials, and polarizabilities. When the eigenvalues $\{\epsilon_i\}$ are interpreted

the systems energy level structure, the transition matrix

What made it possible?

from conceptual to computational breakthroughs, as well as from progress in computer technology. The main enabling steps are as follows:

1. The formulation of $V_{\text{screening}}$ in terms of the density ρ by Kohn and Sham [1].
2. The evaluation of the functional form of useful approximation to $V_{\text{screening}}(\rho)$, for example, the early exchange and correlation functional of Singwi *et al.* [4] and of Hedin and Lundqvist [5], and the most accurate one by Ceperley and Alder [6], and Perdew and Zunger [7].
3. The simplification of V_{ext} in terms of atomic pseudopotentials [8] that are calculable from well-defined (e.g. LDA) atomic models, [9–11] rather than empirically [12].
4. The formulation of nondivergent methods for evaluating E_{tot} and F_{α} of infinite systems (e.g. in momentum-space) [2].
5. The development of efficient computational strategies to solve Equation 1, once general forms of V_{ext} and $V_{\text{screening}}$ are formulated. This includes plane-wave methods (e.g. [2]), LAPW (linear augmented plane wave) [13] and LMTO (linear muffin tin orbital) [14,15].

matrix problems underlying Equation 1, for example, iterative-diagonalization, [16,17] and conjugate gradient [18].

7. The development of strategies for displacing atoms (Equation 3) simultaneously with refining charge densities

the latter [19] with molecular dynamics by Car and Parrinello [20].

parallel architectures.

The characteristics of current first-principles approaches to prediction of materials properties

From the traditional band structure models (inferred from the classic solid state text books in a number of important ways.

The focus is on the physics

gral evaluation approximations), but focuses instead on

structure of matter.

really done in the calculation, or to suspect that one is 'getting the right answer for the wrong reason,' because there are now a number of independent approaches that systematically give the same answers. These are the LAPW and converged pseudopotential calculations (unfortunately, there are still quite a few LDA-based methods around that are poorly implemented producing unreliable results.)

Realism is key

Text-book simplifications including one-dimensional models; spherical-potentials; nearly free-electron models; simple tight-binding and other 'elegant' and 'back of the envelope' constructs have given way to an approach that tackles the real complexity and diversity of matter, without losing sight of the underlying physics.

The approach is reflective

Discrepancies with experiments are analyzed by searching for specific inappropriate physical inputs/assumptions, rather than by using empirical adjustments to cover-up our basic ignorance. For example, when a 'wrong' band gap was calculated early on [21] for LiF (9.8 eV, instead of the measured and previously calculated 14.2 eV) using a 'right' method (LDA with no adjustments [21]), the analysis of this shocking discrepancy led to a deeper understanding of fundamental (self-interaction) corrections [21] to the underlying formalism. Similar discrepancies led later on

to the prediction of quantum-particle corrections to the band structure.

The attitude is bold and interactive

This type of theory sometimes predicts unsuspected stable structures, or materials with previously unknown prop-

erties. The time between the theoretical prediction and its experimental testing is often shorter than the duration of a PhD thesis.

described above characterize mostly the electronic structure theory of 'electronic materials,' that is semiconductors. Indeed, the structure–function relationships underlying semiconductority lead themselves to this type of materials more readily than, for example, superconductivity, or and dynamic correlations play a more crucial role.

Theory predicts defect properties in III-V nitrides

Using the same first-principles approach, Niemi and Nieminen [27] have used the pseudopotential LDA

approach to study the properties of defects in GaN. The

(via acceptors) and n-type doping (via donors). The theory gives access to five basic quantities: firstly, the lowest-energy location and geometry of the defect in the lattice; secondly, the formation enthalpy $\Delta H_f(A^q)$ for defect A in charge state q , as a function of the Fermi energy ϵ_f . The lower $\Delta H_f(A^q, \epsilon_f)$ is, the larger the equilibrium concentration of A^q . Positively-charged defects (donors) form more easily in p-type materials (ϵ_f near the VBM), but their formation releases electrons, thus contributing to n-type behavior. Negatively-charged defects (acceptors) form more easily in n-type materials (ϵ_f near the CBM [conduction band minimum: 'shallow donor']), but their formation

ly, the donor (0/+) [or acceptor (0/-)] transition 'energy level' required to change the charge state. If the (0/+) level is near the CBM, the system becomes n-type, whereas if

'shallow acceptor', the system becomes p-type. Deep donors or acceptors do not contribute carriers; fourthly, the migration barrier energy $\Delta H_m(A^q, \epsilon_f)$ for diffusion of A^q (the larger it is, the less mobile is A^q), and fifthly, the identification of defects which will naturally passivate a given

Regarding 'n-type doping' via donors, theory found the following, firstly, V_N^+ (nitrogen vacancy) is a shallow donor, (producing upon ionization, electrons), but its formation energy $\Delta H_f(V_N^+; n\text{-type})$ is too high to warrant an appreciable carrier density, so V_N^+ is not the source of the observed n-type behavior. However, secondly, Si_{Ga}^+ and O_N^+ (Si on Ga site; oxygen on N site) are much easier to

shallow donor states. Thus, these impurities are likely to be the reason for the observed n-type behavior. Thirdly, H^+ is a shallow donor, and its equilibrium lattice location

mation enthalpy permits a large equilibrium concentration in p-type samples. The migration barrier ΔH_m is low (~ 0.7 eV), so H^+ is highly mobile. It compensates Mg_{Ga}^- (see below).

Regarding 'p-type doping' via acceptors, theory found the following, firstly, V_{Ga}^{3-} (Ga vacancy) has a low formation energy, but its acceptor level is deep (~ 1 eV above VBM),

ered by attachment to O_N^- (forming V_{Ga}^- acceptor-donor pair). Secondly, H^- has a shallow acceptor above the VBM, but its formation enthalpy in n-type material is high (leading to limited solubility). It is almost immobile, due to a large ΔH_m . Hence, its ability to compensate donors such as Si_{Ga}^+ is limited. Thirdly, Mg_{Ga}^- and Be_{Ga}^- are shallow

enthalpies, and are thus predicted to lead to p-type

low formation enthalpy H^+ center. High temperature and

donor-acceptor pair, and recover the p-type behavior.

Opinion

These calculations are executed at a high level of accuracy, show by and large a significant internal agreement, and have greatly contributed to interaction with experiments. However, experimenting seems to lead the way by coming up first with practical dopants in GaN.

Theory predicts the band offsets between semiconductors

When two semiconductors AC and BC form an AC/BC

$\Delta E_v(AC/BC) = E_v(AC) - E_v(BC)$. The magnitude of the offset is an important device characteristic, as it determines the ability of electrons (in the conduction band) and

junction to the other. In the past few years, first-principles electronic structure theory has been applied to predict valence band and conduction band offsets between various materials either assuming an unstrained interface (leading to a 'natural band offset') or assuming a hetero-

trial calculations, and an-electron LAPW calculations produce quite similar predictions, whereas the LMTO method sometimes gives somewhat different values. Let me first illustrate the situation, as well as the extent of agreement/disagreement with experiment by considering nitrides. LAPW calculations [28*] give for the 'natural' offset $\Delta E_v(\text{InN/GaN}) = 0.26$ eV, whereas the pseudopotential values is 0.3 [29]. The LMTO value [30] is 0.51 eV.

experimental values is generally excellent, often within 0.1 eV or so. This is the case for II-Vs [31] and most III-Vs [32]. Even in 'complicated' cases, where one calculates

ternary alloy ($\text{Ga}_x\text{In}_{1-x}\text{P}$) that can have different degrees η of long range order, the agreement with experiment [34] is excellent. (In fact, the agreement with more recently measured values is even better than with the older values. The calculated value did not change.)

I would like to draw attention to three cases where the agreement between theory and the currently available

on these cases, it is essential that the conflict be resolved. Firstly, the measured [35*] $\Delta E_v(\text{GaN/InN}) = 1.05$ eV is considerably different from all first-principles values (0.26, [28*] 0.3, [29] 0.5 [30]). Secondly, the currently measured value for $\Delta E_v(\text{InAs/GaAs})$ strained on an InAs substrate is -0.57 eV [with $E_v(\text{InAs})$ below $E_v(\text{GaAs})$] whereas LAPW

is $-0.04 E_v(\text{InAs})$ being below $E_v(\text{GaAs})$ whereas the LAPW value is -0.28 eV ($E_v(\text{InAs})$ being above $E_v(\text{GaAs})$)

Thirdly, indirect experimental evidence discussed in [32] (see Fig. 1) (CLC: 0.84, V_h: 1.0 eV, E_v: 1.4 eV) where the theory [32] produces a reverse order of band edges.

Opinion

The three discrepancies between theory and experiment for GaN/mnN, GaAs/mnAs and InSb/mnSb are much larger than the stated theoretical error bars, so the experimental determination (e.g. the degree of interfacial coherence maintained in the sample, and the accuracy of determination of the valence band edge) needs to be re-examined.

Spectroscopy of quantum dots explained

Semiconductor ‘quantum dots’ with typical dimensions of 20–100 Å can be synthesized as free-standing objects, or as semiconductor-embedded objects. Free-standing dots (e.g. InP, CdSe) are grown by colloidal methods, are strain-free, nearly spherical and have chemically passivated sur-

are grown by molecular beam epitaxy, are highly (but

Progress made in the growth of ‘free-standing’ quantum dots and in the growth of semiconductor-embedded (‘self-assembled’) dots has opened the door to new and exciting spectroscopic studies of quantum structures. These have

red (red) shifts between absorption and emission emis-

ions, ‘exchange-splitting’, strain-induced ‘splitting’, and Coulomb blockade transitions. These new observations have created the need for developing appropriate theoretical tools capable of analyzing the electronic structure of 10³–10⁶ atom objects. The main challenge is to understand firstly, the way the one-electron levels of the dot and surface effects, and secondly, the nature of many-particle interactions such as electron–hole exchange (underlying the ‘red shift’), electron–hole Coulomb effects (underlying excitonic transitions), and electron–electron Coulomb effects (underlying Coulomb blockade effects).

In response to the challenges posed by these developments, a few theoretical models of the electronic structure of quantum dots were recently advanced. They are all based on the k-p method (see [37]), in which the wavefunction of the dot is expanded in terms of N_Γ bands of the surrounding bulk material. The method range from the ‘effective mass

Norris and Bawendi [38] and Norris, *et al.* [39] and to N_Γ = 8 of Jiang and Singh [40].

Most impressively, in their application to free-standing CdSe dots, Norris and Bawendi [38], succeeded in explaining, via the 6 × 6 k-p method, the origin of as many as eight observed excitonic transitions in various size dots, and Jiang and Singh [40], and Grundmann, Stier and

Bimberg [41] have explained the origin of the two main emission peaks in CdSe-embedded InAs dots.

In addition to explaining such ‘one-electron properties,’ an extended theory was recently able to explain the Stokes shift between absorption and emission as being due to electron–hole exchange interactions [42] (excitation can create either antiparallel, or parallel spins of the excited electron and the particle left behind. The splitting between these ‘singlet’ and ‘triplet’ configuration is due to exchange interactions, and is found to be dramatically enhanced in dots relative to bulk solids). An impressive success of theory in this regard is the identification of enhanced exchange interaction (as opposed to surface trapping) as the origin of the observed red shift.

In general, however, the exchange interaction contains a

with the e-h separation $S = |r_e - r_h|$, and a long-range (LR)

wisdom [42] suggests that the LR exchange interaction in quantum dots originates, as in bulk semiconductors, from dipole–dipole coupling of the transition density between unit cells. Under this assumption, the LR contribu-

tion to the red shift is $\propto 1/R^3$, whereas the short-range (SR) contribution is

whereas the LR contribution is set to zero. This approach fits well the observed red shift in CdSe nanocrystals. However, in the case of spherical zinc-blend quantum dots, the predicted 1/R³ scaling of the red shift with size is not observed in either InP [43] or InAs nanocrystals. In both cases, the observed scaling is R⁻².

Opinion

The currently used ‘theoretical technology’ for understanding the electronic structure of quantum dots is based on the traditional k-p envelope-function approach with a limited basis of N_Γ < 8 bands. Despite spectacular successes, it is unlikely that the details of the electronic structure of quantum-dot systems which lack periodicity in all three dimensions and exhibit surface effects as well as interfacial strain would be adequately described using a small number of 3D-periodic Γ-like bulk orbitals. This approach will ultimately be limited by the number of orbitals needed at the surfaces of the dots, non-Γ-like dot states

the exchange interaction. However, the global features of quantum confinement for sufficiently large dots will continue to be described very well by such approaches.

Conclusions

Electronic structure theory of real materials has reached maturity, in that it is no longer preoccupied exclusively

with methodological issues, but is actively engaged in explaining a wide range of materials properties, predicting unsuspected structures, and even challenging experimental data for correctness of assignment or interpretation.

tations), the current attitude promises to signal the beginning of a new friendship between experiment and theory.

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Kohn W, Sham LJ: **Self-consistent equations including exchange and correlation effects.** *Phys Rev A* 1965, **140**:1133-1138.
2. Ihm J, Zunger A, Cohen ML: **A momentum: space formalism for the**
3. Baroni S, Gianozzi P, Testa A: **Green's function approach to linear response in solids.** *Phys Rev Lett* 1987, **58**:1861.
4. Singwi KS, Sjolander A, Tosi PM, Land RJ: **Electron correlations at metallic clusters. IV.** *Phys Rev B* 1970, **1**:1044-1053.
5. Hedin L, Lundqvist BI: **Explicit local exchange-correlation potentials.** *J Phys C* 1971, **4**:2064-2083.
6. Ceperley DM, Alder BJ: **Ground state of the electron gas by a stochastic method.** *Phys Rev Lett* 1980, **45**:566-569.
7. Perdew JP, Zunger A: **Self-interaction correction to density functional approximations for many-electron systems.** *Phys Rev B* 1981, **23**:5048-5079.
8. Phillips JC, Kleinman L: **New method for calculating wave functions in crystals and molecules.** *Phys Rev* 1959, **116**:287-294.
9. Topiol A, Zunger A, Rutner M: **Local density pseudopotentials for first-row atoms.** *Chem Phys Lett* 1997, **49**:367-373.
10. Zunger A, Cohen ML: **A first principles non local pseudopotential approach.** *Phys Rev B* 1978, **18**:5449-5472.
11. Hamann DR, Schluter M, Chiang C: **Norm-conserving**
12. **form factors for fourteen semiconductors of the diamond and**
13. Wei SH, Krakauer H: **Local-density-functional calculation of the pressure-induced metallization of BaSe and BaTe.** *Phys Rev Lett* 1985, **55**:1200-1203.
14. Andersen OK: **Linear methods in band theory.** *Phys Rev B* 1975,
15. Methfessel M: **Elastic constants and phonon frequencies of Si calculated by a fast full-potential linear-muffin-tin-orbital method.**
16. Davidson ER: **The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices.** *J Comput Phys* 1975, **17**:37-94.
17. Wood DM, Zunger A: **A new method for diagonalizing large matrices.** *J Phys A* 1965, **16**:1343-1359.
18. Payne MC, Teter MP, Allan DC, Arias TA, Joannopolus JD: **Iterative molecular dynamics and conjugate gradients.** *Rev Mod Phys* 1992, **64**:1045.
19. Bendt P, Zunger A: **Simultaneous relaxation of nuclear geometries and electronic charge densities in electronic structure theories.** *Phys Rev Lett* 1983, **50**:1684-1688.
20. Car R, Parrinello M: **Unified approach for molecular dynamics and density-functional theory.** *Phys Rev Lett* 1985, **55**:2471-2474.
21. Zunger A, Freeman AJ: **Ground and excited state properties of LiF in the local density model.** *Phys Rev B* 1977, **16**:2901-2926.
22. Neugebauer J, Van de Walle CG: **Atomic geometry and electronic**
23. Neugebauer J, Van de Walle CG: **Role of hydrogen in doping GaN.** *Appl Phys Lett* 1996, **68**:1829-1831.
24. Boguslawski P, Briggs EL, Bernholc J: **Amphoteric properties of substitutional carbon impurity in GaN and AlN.** *Appl Phys Lett* 1996, **69**:233-235.
25. Boguslawski P, Briggs EL, Bernholc J: **Native defects in GaN.** *Phys Rev B* 1995, **51**:17255-17258.
26. Bergardini F, Fiorentini V, Bosin A: **Theoretical evidence for efficient**
27. **in GaN (small migration enthalpy, i.e. highly mobile, shallow donor, located at high density regions near the surface) and H⁺ in GaN (large formation**
28. **located near Ga antibonding site), it explains that Mg_{Ga} is compensated and passivated by H⁺.**
29. **The paper introduces the basic physics of Ga-substituted Be acceptors in GaN: it is a shallow acceptor (a good thing for p-doping), but Be_3N_2 is a**
30. **suggested that introduction of hydrogen would assist solubility.**
31. **•**
32. **•**
33. **•**
34. **•**
35. **•**
36. **•**
37. **•**
38. **•**
39. **•**
40. **•**
41. **•**
42. **•**
43. **•**
44. **•**
45. **•**
46. **•**
47. **•**
48. **•**
49. **•**
50. **•**
51. **•**
52. **•**
53. **•**
54. **•**
55. **•**
56. **•**
57. **•**
58. **•**
59. **•**
60. **•**
61. **•**
62. **•**
63. **•**
64. **•**
65. **•**
66. **•**
67. **•**
68. **•**
69. **•**
70. **•**
71. **•**
72. **•**
73. **•**
74. **•**
75. **•**
76. **•**
77. **•**
78. **•**
79. **•**
80. **•**
81. **•**
82. **•**
83. **•**
84. **•**
85. **•**
86. **•**
87. **•**
88. **•**
89. **•**
90. **•**
91. **•**
92. **•**
93. **•**
94. **•**
95. **•**
96. **•**
97. **•**
98. **•**
99. **•**
100. **•**

39. Norris DJ, Efros AL, Rosen M, Bawendi MG: **Size dependence of exciton fine structure in CdSe quantum dots.** *Phys Rev B* 1996

42. Efros S, Rosen M, Kuno M, Nirmal M, Norris DJ, Bawendi M: **Band-edge exciton in quantum dots of semiconductors with a**

40. Jiang H, Singh J: **Strain distribution and electronic spectra of InAs/GaAs self-assembled dot and ring heterostructure.** *Phys Rev B* 1997, 56:4696-4701.

account of the effects of short range exchange, shape and size of dot on the fine structure of the spectrum.

43. Nirmal M, Norris DJ, Bawendi MG, Efros A, Rosen M: **Observation of the "dark exciton" in CdSe quantum dots.** *Phys Rev Lett* 1995

• **dots: strain distribution, optical phonons, and electronic structure.**

44. Nirmal M, Norris DJ, Bawendi MG, Efros A, Rosen M: **Band-edge luminescence from dark and bright exciton states.** *Phys Rev B* 1996, 54:4843-4856.

confined state in an InAs pyramidal quantum dot, embedded in GaAs. It formed the basis for extensive comparisons with experiments being the first realistic description of the electronic structure of such dots. However, the neglect of the effect of the

45. Micic OI, Cheong HM, Fu H, Zunger A, Sprague JR, Mascarenhas A,