# Trends

## The quest for dilute ferromagnetism in semiconductors: Guides and misguides by theory

**Alex Zunger and Stephan Lany**
*National Renewable Energy Laboratory, Golden, CO 80401, USA*

**Hannes Raebiger**
*Yokohama National University, Department of Physics, 240-8501 Yokohama, Japan*

Published June 28, 2010

*Theoretical methods have greatly influenced experiment in search of the elusive marriage between semiconductor electronics and magnetism, and the development of spintronics. The path has not always been a straight one, but realizing the limitations and strengths of theoretical approaches promises a straighter course.*

Subject Areas: **Magnetic Semiconductors, Spintronics**

## Introduction

Magnetism mandates that the electrons in a material collectively align their spins. In a ferromagnet there is an imbalance in the occupancy of spin-up and spin-down states; such a collective spin alignment leads to a macroscopic spontaneous magnetization that persists up to a Curie temperature $T_C$. Many common ferromagnets (such as iron or nickel) are metals, exhibiting a set amount of free electrons that make them conduct electrically. In semiconductors, on the other hand, the concentration and type of electronic carriers is controllable externally, for instance, by doping with select atom types or directly by injecting carriers. Most semiconductors are not magnetic. A material that exhibits both ferromagnetic and semiconductor properties offers the exciting prospect of combining nonvolatile magnetic storage and conventional semiconductor electronics in a single device. Magnetic semiconductors offer a number of interesting possibilities in the pursuit of "spintronics,"a branch of science and technology that exploits the spin dimension of the electron in addition to its charge, for novel electronic devices. These materials combine the properties of a semiconductor and a magnetic material, providing, for instance, a way to create 100% spin-polarized currents, and by the same token, the promise of electrical control of magnetic effects. While in some magnetic semiconductors, for example, magnetite, all of the material's constituent ions are intrinsically magnetic ("concentrated magnets"), the most recent focus has been on nonmagnetic semiconductor host materials that can be doped by a small amount of magnetic transition-metal ions or by defects that promote magnetism ("dilute magnets").

The first experiments on dilute magnetic semiconductors focused on host semiconductors with narrow band gaps, such as GaAs and InAs, which are infrared-absorbing black substances. When doped with manganese, $Mn^{2+}$ substitute for $Ga^{3+}$ ions, thus releasing "holes" (a positive charge resulting from the absence of an electron) that are said to collectively align the $Mn^{2+}$ spins ferromagnetically when the sample is cooled below $T_C$. Ferromagnetism in GaMnAs cannot, however, be enhanced indefinitely by adding more Mn ions, as some of the additional manganese will occupy interstitial sites or form some clustered phases that inhibit a global ferromagnetic order. This has limited the ferromagnetism in GaMnAs and other doped III-V semiconductors to experimentally relevant, but not technologically useful, temperatures. For this reason, the possibility of doping common insulators like CaO, ZnO, $In_2O_3$, GaN, and $HfO_2$ instead was seized upon with great enthusiasm. These materials have the additional bonus feature of a wide band gap, promising the control of magnetism through charge carriers in an optically transparent medium. Reports on ferromagnetism above room temperature in such wide-gap materials quickly appeared—even in samples doped with elements that were not in themselves magnetic, or not doped at all! The excitement brought back fond memories of the early days of room-temperature superconductivity, when unusual phenomena were discovered in a group of unsuspected oxides and nitrides. What was special about these developments in pushing the $T_C$ boundaries in dilute magnets was the role of *theory*, which in many cases preceded experiments and influenced the type of systems that were studied. The nature of the delicate tango danced by theory and experiment in this field is the subject of this article.

The tools of solid-state theory of magnetic systems have played a central role in this rapidly developing field, offering not only *ex post facto* explanations of measured phenomena, but very often leading the field by *predicting* specific combinations of host insulators with impurity atoms that would lead to room-temperature ferromagnetism. In this pursuit, two proactive cultures

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53

© 2010 American Physical Society

of predictive theories have emerged. The initial basic philosophy was to use *model Hamiltonians*, in which certain specific magnetic interactions between ions are postulated (and others are excluded). For this technique, the "user" needs to guess at the outset the *type* of magnetic interactions that will dominate. This style of model-Hamiltonian theory has had a glorious past in numerous areas of condensed matter physics. The most prominent supposition for the interaction type between spin-polarized 3*d* impurity ions (such as Mn, Co, and other common dopants from the transition-metal series) in insulators assumed it was similar to the interaction of nuclear spins mediated by the conduction electrons in metals, and the model also assumes that the host crystal is largely unperturbed by the presence of the 3*d* impurity ion. This type of Ruderman-Kittel-Kasuya-Yosida (RKKY) [1] model-Hamiltonian approach [2] went all the way to predict ferromagnetism in various 3*d* impurity-doped compound semiconductors. In particular, the prediction that ferromagnetism persists to a high Curie temperature with $T_C$'s well above room temperature in various wide-gap oxides and nitrides [2]—has resulted in thousands of papers in the past decade, trying to capitalize on these exciting predictions offering high $T_C$ in common insulators.

Alongside the culture of model-Hamiltonian theories of dilute magnets, first-principles calculations have rapidly penetrated this field. The philosophy here is to describe each host and impurity system by considering its fundamental electronic structure, thereby obtaining the magnetic interaction type as the answer, not the question. Instead of prejudging the preferred type of magnetic interactions between ions, one articulates fundamental electron-ion and electron-electron interactions and lets the magnetic interactions emerge as solutions to the fundamental many-electron Hamiltonian. The availability of density-functional approximations (exchange and correlation functionals for the inter-electronic interactions), accurate pseudopotentials (simplifying the calculation greatly), and the ubiquitous computer packages encoding such developments into friendly interfaces freed the theorist from having to guess at the outset how a particular impurity would affect a particular host crystal. At the same time, this apparent freedom sometimes created the illusion that such modeling can be done on "automatic pilot," disregarding potential pitfalls. Regrettably, such pitfalls sometimes created "false positive" results (suggesting high $T_C$'s when no magnetic ordering should occur), spurring enthusiasm and optimism about high-temperature ferromagnetism even when not warranted. Indeed, it has often been the case that materials synthesized from computational recipes have not behaved in the way that was expected.

In the following we will (i) discuss some of the issues and potential pitfalls that can come with uncritical application of theoretical methodology, (ii) analyze the physics of the artifact, and (iii) suggest specific ways out of these dilemmas. The good news is that the insights gained from such critical evaluation can now be embodied (in different ways!) into practical first-principles calculations, offering a truly predictive design option. Furthermore, such calculations clarify what types of interactions are indeed critical and offer model-Hamiltonian approaches a safer, albeit *ex-post-facto*, method to construct more realistic Hamiltonians.

## Transition-metal impurities in semiconductors: Interesting physics, no ferromagnetism

The electronic properties of isolated transition-metal impurities in semiconductors were studied in great detail in the 1980s [3, 4], when the growth methods produced ultradilute samples (defined as containing $10^{15}$–$10^{17}$ impurities/cm$^3$) with virtually no impurity-impurity interaction. The high crystalline perfection of such samples enabled a clear experimental and theoretical understanding of the properties and chemical trends in the periodic table of isolated impurities, including their spin configuration, donor and acceptor transitions, and optical excitations. A number of intriguing physical phenomena were revealed due to coexisting localized 3*d* impurity states. One of these is the "self-regulating response" [5, 6], recently revisited by Raebiger *et al.*[7], which shows that adding electrons to the 3*d* derived impurity levels does not change the integrated electron density around the 3*d* impurity atom because the orbitals of the surrounding ligand atoms rehybridize to minimize the impact of impurity charging. A second phenomenon is the "vacuum pinning rule," whereby transition levels of the same impurity atom in different host semiconductors line up in approximately the same way, regardless of the host chosen [8, 9], thus permitting the deduction of band offsets between different pure materials from the knowledge of impurity level positions [10]. Third, another consequence is called "exchange-correlation negative *U*," [11] whereby the strongly varying exchange splitting of differently charged impurities can lead to effective electron-electron attraction. These systems all show interesting physics, however, such systems revealed no ferromagnetism, as the generally low solubility of dopants precluded collective interactions.

Collective magnetic ordering requires the individual magnetic impurity ions to be close enough to one another so as to interact. In practice, this implies large concentrations, often well above typical thermodynamic solubility limits. The low solubility limit of most 3*d* ions in semiconductors can, however, be overcome by using nonequilibrium growth techniques (such as very low-temperature molecular beam epitaxy). These techniques increased the impurity concentration by a few orders of magnitude to $10^{20}$ dopants/cm$^3$. This, however, comes

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53
© 2010 American Physical Society

at the price of severely lowering the structural quality of the samples. Such high concentrations of magnetic ions in semiconductors, such as GaAs:Mn [12, 13], leads to ferromagnetism with relatively high Curie temperatures of around 100 K—a temperature range that has opened the era of spintronics [14, 15].

The basic driving force of this form of ferromagnetism of 3*d* impurities in III-V semiconductors is now understood [16] as the energy stabilization ensuing from the interaction between partially occupied hybrid orbitals located on different impurity sites, $TM_1$ and $TM_2$, in the host crystal. Figure 1 (top) illustrates that if two such orbitals interact, there will be a gain in energy, i.e., a stabilization of the FM state (relative to the antiferromagnetic state [16]).

One might ask, how would such individual partially occupied hybrid orbitals come to exist in the first place? The answer to this is illustrated in Fig. 1 (bottom). The formation of a substitutional 3*d* impurity can be thought of as a two step process, involving first the removal of a host Ga atom, and second, the placement of a 3*d* impurity atom in its place. The combined impurity/host orbitals simply result from coupling between the host-crystal cation vacancy state ("dangling bonds" of t(p) symmetry) and the impurity orbital of the 3*d* atom filling this vacancy (having t(d) symmetry as well as nonbonding e(d) symmetry). The interaction between these states leads to a bonding state, and an antibonding state. The partial occupancy which ultimately drives ferromagnetism resides in the higher of these states. For Mn in GaAs the antibonding state, dominated by t(p), is called a "dangling bond hybrid" (DBH) state, and the bonding/nonbonding states, dominated by t(d) and e(d), are called "crystal field resonances" (CFR). As we move to the left from Mn in the periodic table, keeping the GaAs host, we approach the light 3*d* impurity elements such as V. For these, the atomic t(d) and e(d) are higher in energy and the roles of DBH and CFR are interchanged. The same interchange occurs in the case where Mn is placed in a host crystal with light anion elements, such as nitrogen in GaN, where the t(p) level is deeper and t(d) dominates the hole. It is the interaction between the partially occupied hybrid states, centered on different impurity sites in the crystal, which energetically stabilizes the ferromagnetic configuration.

This simple mechanism implies simple rules [17] for impurity-host material combinations that might produce ferromagnetism: one obtains the electronic configuration by constructing an Aufbau ground state for *n*-*m* electrons (*n* is the number of TM valence electrons, and *m* is the valence of the cation site) occupying the relevant symmetry orbitals. For example, a ferromagnetism-promoting partial occupancy of DBH can be expected when Mn occupies a trivalent Ga site, but not the divalent Zn site. This mechanism was directly verified by first-principles calculations of the energy of impurities in a solid as a function of the occupancy of the DBH, and noting the stabilization (destabilization) of the ferromagnetism for partial (full) occupancy.

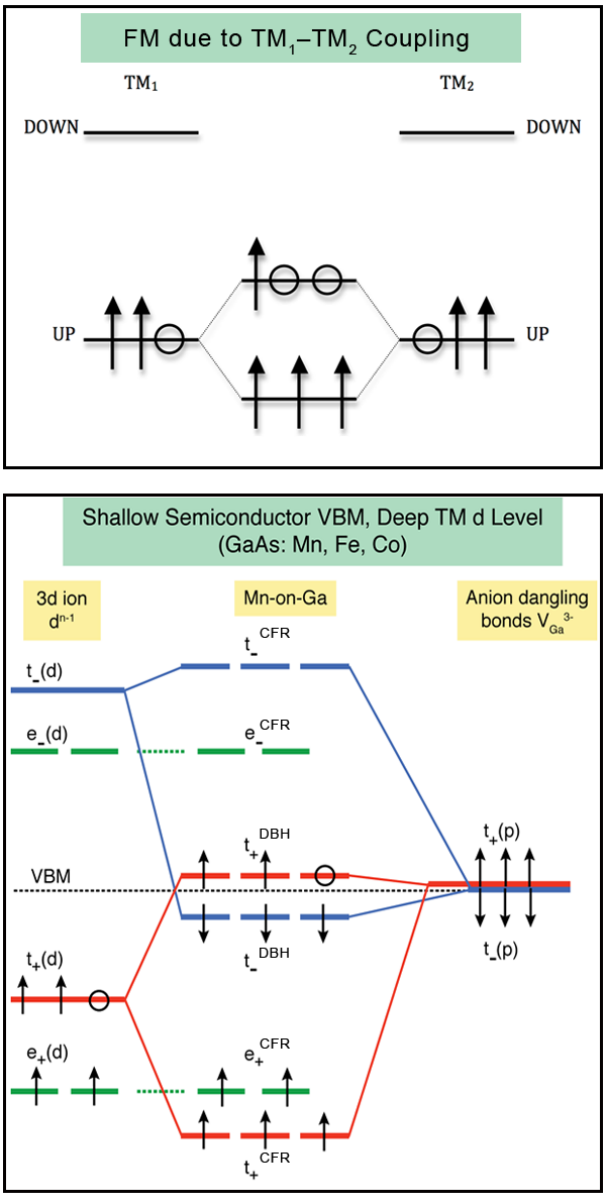


FIG. 1: (Top) Energy level diagram describing the stabilization of the ferromagnetic spin arrangement as a result of the interaction between two hybrid orbitals located on two impurities, here labeled $TM_1$ and $TM_2$, in a tetrahedral semiconductor. In the case of partial occupancy of the individual hybrid orbitals, the ferromagnetic configuration is stabilized by the preferential filling of the lower energy state. (Bottom) Energy level diagram describing how an individual hybrid orbital is formed from the coupling between the host cation vacancy orbitals t(p) and the 3*d* orbitals t(d)+e(d). See Ref. [16]. (Illustration: Carin Cain)

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53

© 2010 American Physical Society

## Even higher $T_C$ promised in wider-gap semiconductors

First-principles electronic structure theory of Mn doped into III-V nitrides, phosphides, and arsenides [18–20] have shown that the ferromagnetism-promoting hole (i.e., the orbital with partial occupancy, $t_+^{DBH}$ in Fig. 1) resides inside the host band gap. When the host anion changes from As to P and then to N, the band gap increases primarily by shifting the valence band maximum (VBM) to lower energies. Thus, as one moves from arsenides to phosphides and then to nitride semiconductors, the hole-carrying impurity level, being roughly pinned, appears farther and farther from the VBM, deeper in the band gap. This makes the hole-carrying impurity wave function progressively more localized, weakening its communication with other magnetic ions; thus we lose ferromagnetism. Thus the conclusion of such microscopic calculations [4, 16, 17] was that Mn will have lower $T_C$ as we go from a host arsenide to a host phospide and then to a nitride. While initial estimates of $T_C$ derived from directly associating the Weiss field with first-principles calculated magnetic coupling energies predicted optimistically higher $T_C$ in the wider gap (i.e., nitride) host crystals [21], it was quickly realized (by the same authors) that a proper statistical treatment of these exchange energies (via Monte Carlo, see Ref. [22]) predicted a much *reduced* $T_C$ as one moves to wider gap host crystals.

In parallel with such first-principles calculations, more traditional model-Hamiltonian approaches were developed [2, 23] that assume a scenario for the underlying interactions, and then solve the scenario mathematically. The classic $\mathbf{k} \cdot \mathbf{p}$ approach in semiconductor physics, as described in Ref. [24], expands the states of a system in terms of a small number of preselected bands of the host crystal at the Brillouin-zone center. In the field of dilute magnetic semiconductors, the use of such $\mathbf{k} \cdot \mathbf{p}$ concepts assumes that the hole that promotes magnetism can be described by just a few host crystal states, becoming essentially a "hostlike hole" [2, 23]). Such simple concepts differ substantially from the atomistic picture of Fig. 1 (bottom). The model Hamiltonian approach is then able to produce, analytically, the dependence of the ferromagnetism Curie temperature $T_C$ on basic material parameters. This lead to the highly influential bar-diagram published by Dietl *et al.*[1] (cited more than 3200 times) predicting the highest $T_C$ for binary semiconductors having the shortest bond length and thus the largest band gap. This design principle pointed towards wide-gap binary compounds containing a first-row element such as C, N, or O as the best candidates for high-$T_C$ ferromagnets. This prediction started a world-wide quest for high-$T_C$ ferromagnets in wide-gap metal oxides (ZnO, $HfO_2$), nitrides (GaN), and carbides based on dilute transition metal impurities. As exciting as such predictions were, they did not have the additional virtue of becoming true. Indeed, the assumed scenario that the ferromagnetism-promoting hole has the personality of the host valence band did not seem to be true for wide-gap insulators doped with transition metals such as Mn. The simple lesson from microscopic models that the Mn acceptor (hole-carrying) level becomes progressively *more* localized (i.e., contributing to lower $T_C$) as one moves up the group V elements from arsenic to nitrogen was lost in the global enthusiasm to seek high $T_C$ in the widest gap host systems.

## The role of theory as a guide: Pitfalls and fixes

Magnetic ions were not the only dilute species in wide-gap insulators that were implicated with high-$T_C$ ferromagnetism. Indeed, simple structural defects such as cation vacancies [25–28] or carbon or nitrogen impurities [29–31] also became prime candidates for point sources of ferromagnetism behavior in wide-gap insulators. The basic thought here is that a proper combination of defect/impurity and a host crystal will lead to a hole-carrying (partially occupied) level, centered about this defect/impurity site, and that a ferromagnetic spin arrangement will be stabilized when different such centers interact [Fig. 1 (top)]. The search for the particular combination of defect/impurity and host crystal was, in many instances, guided by theory. Indeed, an impressive sign of the maturity of electronic structure theory based on first-principles was the central role that it had occupied in guiding the international quest for ferromagnetism in dilute wide-gap insulators [32–35].

Unfortunately, a number of overlooked factors contributed to confusion. We next discuss a number of ways in which such calculations can produce unfounded optimism, such as a too high-$T_C$ ferromagnetism. Fortunately, recognition of such factors has lead to proposing a fix. Efforts worldwide to improve these fixes promise to create a robust guide for new design principles for such materials.

### Problem 1: False occupancy

A well-known problem of the local density approximation (LDA) and its relatives like the generalized gradient approximation (GGA) is the underestimation of the band gap of nonmetals; an underestimation often manifested by a conduction band minimum that is set too low. This can become a problem in the context of magnetism if the defect/impurity center contemplated will be resonating within the (artificially too low) conduction band of the host when LDA is used, whereas in reality this level should be in the band gap (Fig. 2). It turns out that this is not a problem for Mn in III-V semiconductors, where the conduction band minimum

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53

© 2010 American Physical Society

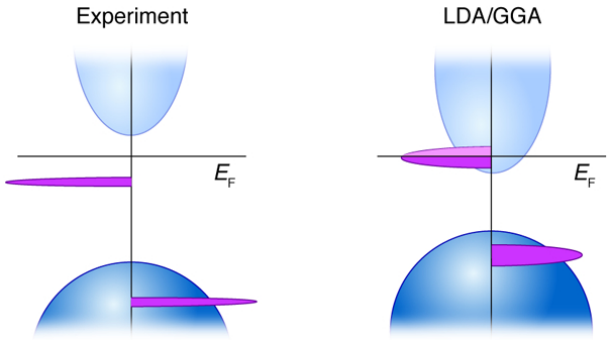


FIG. 2: Ferromagnetism requires that there is an imbalance in the occupancy of spin-up (pink, right-pointing region) and spin-down (left-pointing) orbitals. Experimentally, both the spin-up and spin-down orbitals are filled (left panel) so there is no ferromagnetism. However, LDA calculations (right panel) predict that the conduction band is set so low that it "swallows" the spin-down band, causing spurious charge spilling from the transition metal orbitals into conduction states and a false positive prediction for ferromagnetism. (Illustration: Carin Cain)

is high enough that the hole-carrying impurity orbital is inside the band gap, even in LDA [36]. However, it is often a problem in oxides, when the substantial LDA underestimation of the conduction band minimum energy can spuriously place the impurity orbital inside the host crystal conduction band (for example, Mn in ZnO [32, 33, 37]). The electrons will then drop into the conduction band minimum, creating a hostlike delocalized orbital that promises long-range interaction with other centers, to the benefit of (spurious) high-$T_C$ ferromagnetism. This type of LDA-error was demonstrated also in ZnO:Cr [31]. Using GGA or GGA+$U$ (which seeks to correct problems by adding a repulsive potential to GGA) instead of LDA does not solve the problem, as it shifts the unoccupied 3$d$ impurity level to yet higher energies, nor does the "scissors operator" correction, which shifts the host conduction band minimum to higher energies *after* the calculation is done. Such false occupancy errors have led to a substantial number of publications predicting optimistically long-range ferromagnetism. Clearly, one needs to maintain correct host band edge positions *before* interaction with the defect/impurity electrons is treated, and then see if ferromagnetism will survive a correct positioning of the levels.

Once this principle is understood, a number of strategies can be applied to overcome the false occupancy problem. One such example is our earlier work in which we designed empirical nonlocal external potentials (NLEP) [38], in the spirit of the method of Christensen [39], where external potentials are placed on host atoms (and interstitial sites) and the potential parameters are adjusted to reproduce the correct host band structure. Applying such band-gap corrections to self-consistent impurity calculations allows one to place the impurity levels in the correct manifold of host states. For ZnO:Cr, the impurity electrons were now in a (Jahn-Teller split) level within the band gap, having no partial occupancy, hence no ferromagnetism [38]. Other approaches to avoid false occupancy may involve using, at the outset, band-gap corrected methods, such as hybrid functionals [40] or self-interaction corrections (SIC) [41, 42] instead of LDA and its relatives.

However, all is not lost: one can turn false occupancy around. One important outcome of the understanding of the significance of the impurity versus host level alignment is the prediction of electron-induced ferromagnetism in oxides. As noted above, ferromagnetism in dilute III-V materials was universally due to holes. However, certain impurity-host combinations might have a legitimate impurity level in resonance with the host conduction band; if that level is not too high inside the conduction band, it might be partially occupied by additional intentional donor doping, creating the desired partial occupancy and hence ferromagnetism. This idea led to a search for 3$d$ impurities in ZnO and $In_2O_3$ [38, 41, 43–46] that have levels in resonance with the lowest part of the host conduction band. It is important here to note that the relevant quantity describing the energy of the TM-impurity "level" is the quasiparticle energy, that is, the electron addition energy into the unoccupied defect state. Thus the empty $t_2$ level of ZnO:Co lies inside the conduction band [38, 41], whereas (incorrect) estimates based on typical $d$-$d$ transition energies of tetrahedral Co placed the $t_2$ state at a much lower energy inside the gap [45]. Since, however, such $d$-$d$ transitions correspond to charge-conserving excitations, they are not suited to determine quasiparticle energies [46], indeed, the Co$t_2$ "level" is inside the ZnO conduction band.

One of the few cases estimated to have a 3$d$ level inside, but not too high within, the conduction band was Cr in ZnO [38] and $In_2O_3$ [43]. However, according to the calculation, ferromagnetic coupling requires a rather large number of carriers—of the order of $10^{21}$ cm$^{-3}$. This may be difficult to attain in structurally high-quality oxides, which may explain why ferromagnetism in dilute oxides is generally associated with highly defective samples and never occurs in single crystals.

The good news is that once understood, false occupancy in LDA can be cured, so the trap of false positive ferromagnetism due to this effect can now be put behind us.

### Problem 2: The burden of proof

What does one need to prove theoretically to establish that a given defect or impurity in a wide-gap host crystal can lead to collective ferromagnetism? Early papers [18, 21, 25, 29, 32, 35] focused mostly on the fact that certain point defects in solids carried local moments,

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53
© 2010 American Physical Society

thereby opening the way to ferromagnetism. Indeed, the neutral Si vacancy in bulk Si, the As vacancy in GaAs, the Zn vacancy in ZnO, and the Ca vacancy in CaO all have orbital configurations with local moments, and thus should all qualify as candidates for producing ferromagnetism. Yet to establish collective ferromagnetism, one needs to show more.

First, one would like to establish the magnitude and range of defect-defect magnetic interaction energies. This can be done via direct total energy calculations of various spin configurations [40, 47–49], or approximated via the linear response formula [50, 51]. For example, the magnetic interaction between two Ca vacancies in CaO [47] extends only to 4 to 5 neighbors, after which it's basically zero.

Second, given this interaction range, one would like to know what is the minimum concentration $X_p$ of such defects in the lattice that assures percolation, that is, an uninterrupted, wall-to-wall chain of moments. It turns out that it's not all that easy to find in the literature the percolation threshold for $N$th neighbor interactions in various lattice types (many results are available though for $N = 1, 2$ on standard Bravais lattices). The appendix of Ref. [52] provides percolation ladders ($X_p$ vs $N$) for a few lattice types, from which one can learn, for example, that fourth neighbor interaction in fcc-type, CsCl-type, diamond-type, and anatase-type lattices require a minimum concentration of $X_p = 4.9, 8.0, 7.6$, and 7.1% defects, respectively. This is a very large concentration of defects to demand, certainly much larger than what most crystalline materials will tolerate without mechanically collapsing.

Third, one wonders what will be the maximum concentration of given defects that can be granted thermodynamically, under the most favorable growth conditions. There is indeed a rather well-established method to calculate the formation energies of various point-defects in a lattice as a function of the Fermi energy and chemical potential [53–55] and then use such results to calculate the equilibrium concentration as a function of temperature for a given (e.g., most favorable) chemical potential. When applied to Ca vacancy in CaO [47], this gives only $X = 0.003\%$ even at high temperatures close to melting. This then implies that to achieve percolation (4.9%) one needs a growth method that will defeat equilibrium by a factor of $4.9/0.003 \sim 1600$. It is unlikely that this is achievable throughout a single crystal sample, except at special locations such as grain boundaries, surfaces, or dislocation lines. Perhaps these are some of the "secret ingredients" that have tacitly led to the observation of ferromagnetism?

We conclude that the burden of proof for potential ferromagnetism-promoting defects should include not only the establishment of the local spin-configuration (partially occupied orbitals with magnetic moment), but also steps 1–3 above. The good news is that once these requirements are understood, putting them into practice is doable, and will certainly significantly narrow down the number of viable candidate cases that should be examined seriously in the laboratory.

### Problem 3: Failure to localize

Symmetry breaking of defect orbitals is known in simple oxides. A Mg vacancy in MgO has six crystallographically equivalent oxygen neighbors, yet the two holes created by a charge-neutral Mg vacancy are localized mostly on just two of the six oxygen atoms [56]. Significantly, LDA approaches will fail to localize the holes—thus overestimating the interaction between holes associated with different cation vacancies—and will predict spurious ferromagnetism. The failure of LDA to localize stems fundamentally from the convex, rather than linear, dependence of its total energy on occupation numbers [57–59] and is a manifestation of the need for self-interaction correction. Since failure to localize may artificially enhance intersite wave function overlap, it is important to estimate its impact on defect-defect ferromagnetism. What is required is a formalism that corrects the concave $E(N)$ tendency of Hartree-Fock (favoring localization) and the convex $E(N)$ tendencies in LDA (favoring delocalization), replacing it by a linear dependence. This requirement can be cast as a generalized Koopmans' condition, i.e., the electron addition energy be equal to the orbital energy, $E(N+1) - E(N) = e_i(N)$ [60]. Such cancellation of nonlinearity (CONL) may be achieved approximately by mixing Hartree-Fock with LDA [61], or by self-interaction corrections to DFT [62, 63], or more explicitly, by adding to DFT an on-site potential with a parameter obtained by requiring the generalized Koopmans' condition [60], therefore enforcing the CONL by construction. Using such a CONL approach leads to self-trapped holes in both rutile and anatase $TiO_2$ [64], and leads to much deeper acceptor levels and more localized hole orbits for acceptors in oxides and nitrides, e.g., N:ZnO [65], Cu:ZnO, Mg:$In_2O_3$, In:$SnO_2$ [60], and Be:GaN [66]. When applied to metal vacancies implicated previously via LDA calculations in ferromagnetism [31, 67], Chan *et al.*[68] found that the cancellation of the nonlinearity led to a much reduced $V_{Zn} - V_{Zn}$ interaction in ZnO, from 32 meV obtained in LDA [31] for a pair separated by 8.4 Å, to less than 1 meV. This reduction of magnetic interaction results from the strong localization of the $V_{Zn}$ spin-density, as shown in Fig. 3. After CONL, the $V_{Zn}$ does not even possess a robust magnetic moment any more, because the singlet state ($S = 0$) is almost degenerate with the triplet state ($S = 1$) [68]. Clearly, the hopes for ferromagnetism due to cation vacancies based on the standard DFT [28, 31, 47, 67] evaporate when the spurious delocalization is corrected.

The good news is that once the role of CONL on magnetism is understood, this can be readily cured either empirically or by complete generalized Koopmans' methods.

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53

© 2010 American Physical Society

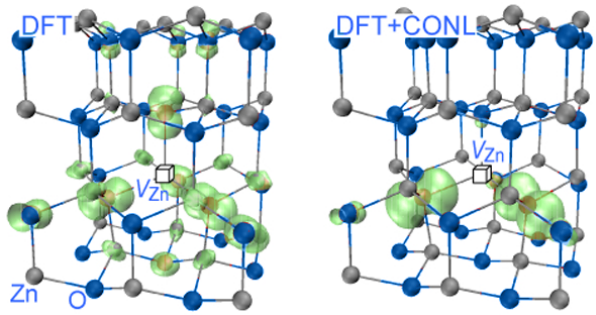


FIG. 3: Calculations of the spin-density (green) of a Zn vacancy ($V$Zn) in ZnO in the triplet state ($S = 1$) in standard density-functional theory (DFT) (left) and after a correction (CONL) (right) show how uncorrected DFT predicts too delocalized spins. Consequently, the defect-defect wave function overlap and the resulting ferromagnetic coupling are strongly overestimated [68]. (Illustration: Carin Cain)

## If it's not one thing, it's another

Guiding experiments towards promising defects or impurities that can promote carrier-induced ferromagnetism in certain wide-gap materials has proven to be difficult. Even if you correctly formulate the underlying electronic structure problem so that false-occupancy is avoided, the LDA failure to localize holes can get you. And if these are fixed, you still need to establish the defect concentration that satisfies percolation, yet is tolerated mechanically by the lattice. And you still need to consider the microstructure, i.e., the microscopic distribution of defects, since ideal single oxide crystals with ideal impurities do not lead to ferromagnetism. Indeed, if it's not one thing that leads to false FM prediction, it's another.

Finally, one should ask whether it is appropriate to describe the collective magnetism of a sample in terms of interacting single isolated point defects? Statistics tells us that the probability of two or more defects on one of the sublattices in a zinc blende, wurtzite, or rock salt crystal occupying neighboring sites is $> 50\%$ for defect concentrations $> 5\%$ ($\sim 10^{20}/\mathrm{cm}^3$) [69]. Once a defect pair has another defect next to it (e.g., with a 50% probability), it has a lowered point group symmetry and consequently a different electronic spectrum, and is likely to exhibit completely different magnetic interactions compared to its isolated relatives [48, 70]. This has the detrimental effect that magnetic interactions no longer can be mapped on Dirac-van Vleck-Heisenberg type models ($E = \Sigma J_{ij}\, \mathbf{s}_i \cdot \mathbf{s}_j$) [71]. As expected, this problem can be circumvented by introducing additional correction terms or explicit configuration dependences to the model, as was done in Refs. [49, 51], which, however, mandates the explicit calculation of various additional clustered/microstructured configurations and their energetics. Moreover, the flaws in the underlying spin models seem to be of a fundamental nature, since, as it turns out, magnetic coupling energies, even of the simplest hydrogen molecules, have no mapping thereupon [72]. This failure of Heisenberg's theory [73] can further be traced back onto the Heitler-London theory [74], of which it is a generalization, and which has been abandoned on several occasions [75–77].

## Outlook

**Bad is Good?** No, this is not an attempt at moral revisionism. But consider this: The functionality of optoelectronic semiconductor devices such as lasers, light-emitting diodes, or photovoltaic solar cells is predicated on the structural quality of the active material. "Good materials," i.e., those leading to efficient functionality, invariably need to have good crystalinity, purity, and very few defects. Yet if one were to use such near-perfect bulk samples of *oxides*, no ferromagnetism of the type discussed here would be present. Considered from this viewpoint, structurally imperfect and compositionally nonstoichiometric oxide materials (which would be considered "bad materials" in the semiconductor device culture) are "good materials" in this field of dilute magnetic oxides, as they are the only ones that yield, so far, the desired functionality.

## The future of theory guide

Although much of the article has pointed out the various pitfalls encountered in running LDA on automatic pilot, the conclusions are more optimistic: Electronic structure calculations are a powerful tool to describe magnetic ground states, even in complicated clustered or nanostructured configurations. The ensuing results clarify which types of interactions are indeed critical, offering model-Hamiltonian approaches a safer, albeit *ex-post-facto* method to construct safe, simpler Hamiltonians. Furthermore, the results may be used as input for statistical models constructed to avoid the obvious pitfalls in earlier Heisenberg-like attempts.

## Acknowledgment

This work was funded by the U.S. Department of Energy, Office of Science, Basic Energy Science, Materials Science and Energy Division, under Contract No. DE-AC36-08GO28308 to NREL. HR is supported by a Grant-in-Aid for Young Scientists (A) grant (No. 21686003) from the Japan Society for the Promotion of Science.

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53

© 2010 American Physical Society

## References

[1] M. A. Ruderman and C. Kittel, *Phys. Rev.* **96**, 99 (1954); T. Kasuya, *Prog. Theor. Phys.* **16**, 45 (1956); K. Yosida, *Phys. Rev.* **106**, 893 (1957).
[2] T. Dietl *et al.*, *Science* **287**, 1019 (2000).
[3] V. F. Masterov, *Fiz. Tekh. Poluprovodn.* **18**, 3 (1984).
[4] A. Zunger in *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic Press, Orlando, 1986), Vol. 39, pp. 275-464.
[5] F. Haldane and P.W. Anderson, *Phys. Rev. B* **13**, 2553 (1976).
[6] A. Zunger and U. Lindefelt, *Solid State Commun.* **45**, 343 (1983).
[7] H. Raebiger, S. Lany, and A. Zunger, *Nature* **453**, 763 (2008).
[8] L. A. Ledebo and B. K. Ridley, *J. Phys.* **15**, L961 (1982).
[9] M. J. Caldas, A. Fazzio, and A. Zunger, *Appl. Phys. Lett.* **45**, 671 (1984).
[10] A. Zunger, *Phys. Rev. Lett.* **54**, 849 (1985).
[11] H. Katayama-Yoshida and A. Zunger, *Phys. Rev. Lett.* **55**, 1618 (1985).
[12] H. Munekata *et al.*, *Phys. Rev. Lett.* **63**, 1849 (1989).
[13] H. Ohno *et al.*, *Appl. Phys. Lett.* **69**, 363 (1996).
[14] A. Bonanni and T. Dietl, *Chem. Soc. Rev.* **39**, 528 (2010).
[15] K. Kikoin, *J. Magn. Magn. Mater.* **321**, 702 (2009).
[16] P. Mahadevan *et al.*, *Phys. Rev. Lett.* **93**, 177201 (2004); *Phys. Rev. B* **69**, 115211 (2004).
[17] Y. J. Zhao *et al.*, *J. Appl. Phys.* **98**, 113901 (2005).
[18] K. Sato and H. Katayama-Yoshida, *Jpn. J. Appl. Phys.* **40**, 485 (2001).
[19] L. Sandratskii and P. Bruno, *Phys. Rev. B* **67**, 214402 (2003).
[20] T. Schulthess *et al.*, *Nature Mat.* **4**, 838 (2005).
[21] K. Sato *et al.*, *Europhys. Lett.* **61**, 403 (2003).
[22] K. Sato and H. Katayama-Yoshida, *Phys. Rev. B* **70**, 201202 (2004).
[23] T. Jungwirth *et al.*, *Rev. Mod. Phys.* **78**, 809 (2006).
[24] P. Yu and M. Cardona, *Fundamentals of Semiconductors: Physics and Materials Properties* (Springer, New York, 2005).
[25] I. S. Elfimov, S. Yunoki, and G.A. Sawatzky, *Phys. Rev. Lett.* **89**, 216403 (2002).
[26] M. Venkatesan, C. B. Fitzgerald, and J. M. D. Coey, *Nature* **430**, 630 (2004).
[27] N. H. Hong *et al.*, *Phys. Rev. B* **73**, 132404 (2006).
[28] P. Dev, Y. Xue, and P. H. Zhang, *Phys. Rev. Lett.* **100**, 117204 (2008).
[29] K. Kenmochi *et al.*, *Jpn. J. Appl. Phys.* **43**, L934 (2004).
[30] B. Gu *et al.*, *Phys. Rev. B* **79**, 024407 (2009).
[31] H. Peng *et al.*, *Phys. Rev. Lett.* **102**, 017201 (2009).
[32] K. Sato and H. Katayama-Yoshida, *Jpn. J. Appl. Phys.* **39**, L555 (2000).
[33] O. Eriksson *et al.*, *J. Appl. Phys.* **101**, 09H114 (2007).
[34] H. Pan *et al.*, *Phys. Rev. B* **77**, 125211 (2008).
[35] K. Griffin Roberts *et al.*, *Phys. Rev. B* **78**, 014409 (2008).
[36] A. Filippetti *et al.*, *Chem. Phys.* **309**, 59 (2005).
[37] D. Iusan et al., Phys. Status Solidi A 204 53 (2007).
[38] S. Lany, H. Raebiger, and A. Zunger, *Phys. Rev. B* **77**, 241201(R) (2008).
[39] N. E. Christensen, *Phys. Rev. B* **30**, 5753 (1984).
[40] C. H. Patterson, *Phys. Rev. B* **74**, 144432 (2006).
[41] C. D. Pemmaraju, T. Archer, D. Sánchez-Portal, and S. Sanvito, *Phys. Rev. B* **75**, 045101 (2007).
[42] H. Kizaki *et al.*, *Appl. Phys. Express* **2**, 053004 (2009).
[43] H. Raebiger, S. Lany, and A. Zunger, *Phys. Rev. Lett.* **101**, 027203 (2008).
[44] H. Raebiger, S. Lany, and A. Zunger, *Phys. Rev. B* **79**, 165202 (2009).
[45] A. Walsh, J.L.F. Da Silva, and S.H. Wei, *Phys. Rev. Lett.* **100**, 256401 (2008).
[46] S. Sanvito and C.Das Pemmaraju, *Phys. Rev. Lett.* **102**, 159701 (2009).
[47] J. Osorio-Guillén, S. Lany, S. V. Barabash, and A. Zunger, *Phys. Rev. Lett.* **96**, 107203 (2006).
[48] H. Raebiger *et al.*, *Phys. Rev. B* **72**, 014465 (2005).
[49] T. Hynninen *et al.*, *J Phys. Condens. Matter* **18**, 1561 (2006).
[50] A. I. Liechtenstein *et al.*, *J. Magn. Magn. Mater.* **67**, 65 (1987).
[51] A. Franceschetti *et al.*, *Phys. Rev. Lett.* **97**, 047202 (2006).
[52] J. Osorio-Guillén, S. Lany, S. V. Barabash, and A. Zunger, *Phys. Rev. B* **75**, 184421 (2007).
[53] C. G. van de Walle and J. Neugebauer, *J. Appl. Phys.* **95**, 3851 (2004).
[54] C. Persson, Y. J. Zhao, S. Lany, and A. Zunger, *Phys. Rev. B* **72**, 035211 (2005).
[55] S. Lany and A. Zunger, *Phys. Rev. B* **78**, 235104 (2008).
[56] A. M. Stoneham, *J. Phys. Condens. Matter* **22**, 074211 (2010).
[57] J. P. Perdew *et al.*, *Phys. Rev. Lett.* **49**, 1691 (1982).
[58] J. P. Perdew *et al.*, *Phys. Rev. A* **76**, 040501 (2007).
[59] P. Mori-Sánchez, A. J. Cohen, and W. Yang, *Phys. Rev. Lett.* **100**, 146401 (2008).
[60] S. Lany and A. Zunger, *Phys. Rev. B* **80**, 085202 (2009).
[61] A. L. Shluger *et al.*, *Modelling Simul. Mater. Sci. Eng.* **17**, 084004 (2009).
[62] M. d'Avezac, M. Calandra, and F. Mauri, *Phys. Rev. B* **71**, 205210 (2005).
[63] A. Droghetti, C. D. Pemmaraju, and S. Sanvito, *Phys. Rev. B* **78**, 140404 (2008).
[64] S. Lany and A. Zunger (unpublished).
[65] S. Lany and A. Zunger, *Phys. Rev. B* **81**, 205209 (2010).
[66] S. Lany and A. Zunger, *Appl. Phys. Lett.* **96**, 142114 (2010).
[67] Q. Wang *et al.*, *Phys. Rev. B* **77**, 205411 (2008).
[68] J. A. Chan, S. Lany, and A. Zunger, *Phys. Rev. Lett.* **103**, 016404 (2009).
[69] R. Behringer, *J. Chem. Phys.* **29**, 537 (1958).
[70] H. Raebiger *et al.*, *J Phys. Condens. Matter* **16**, L457 (2004).
[71] M. van Schilfgaarde and O. Mryasov, *Phys. Rev. B* **63**, 233205 (2001).
[72] I. Ciofini *et al.*, *Chem. Phys.* **309**, 133 (2005).
[73] W. Heisenberg, *Z. Phys.* **49**, 619 (1928).
[74] W. Heitler and F. London, *Z. Phys.* **44**, 455 (1927).
[75] L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, Ithaca, 1960).
[76] J. C. Slater, *Rev. Mod. Phys.* **25**, 199 (1953).
[77] K. Hongo *et al.*, *Mat. Trans.* **48**, 662 (2007).

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53

© 2010 American Physical Society

## About the Authors

### Alex Zunger

Alex Zunger's research field is condensed matter theory of real materials. He is the recipient of the 2010 Tomassoni Physics Prize and Science Medal of Scola Physica Romana, the 2001 John Bardeen Award of The Material Society, the Rahman Award of the American Physical Society, the 2009 Gutenberg Award (Germany) and the 2000 Cornell Presidential Medal. Israeli raised and educated, he received his Ph.D. from Tel-Aviv University, Israel. He did his postdoctoral research at Northwestern University (with Art Freeman). He then received an IBM Fellowship, which he spent at the University of California, Berkeley (with Marvin Cohen). Dr. Zunger established the Solid State Theory group at the National Renewable Energy Laboratory (NREL), Golden, Colorado, a position he still holds today. He is an NREL Institute Research Fellow, and Director of the Energy Frontier Research Center on Inverse Design.

### Stephan Lany

Stephan Lany received his Ph.D. in physics from the Universität des Saarlandes, Germany, in 2002. He is currently a senior scientist at the National Renewable Energy Laboratory. In his recent research activities, he improved the methodology for accurate prediction of the properties of defects and impurities in semiconductors and insulators from electronic structure theory. He applied such methods to magnetic semiconductors and photovoltaic materials, and used the predicted defect energies for thermodynamic modeling of defects, off-stoichiometry, and doping.

### Hannes Raebiger

Hannes Raebiger obtained his doctorate at the Helsinki University of Technology in 2006. In 2005 he was awarded the J. W. Corbett prize at the 23rd International Conference on Defects in Semiconductors for his studies elucidating the magnetic interactions in dilute magnetic semiconductors. After obtaining his doctorate, he joined Alex Zunger's Solid State Theory group at the National Renewable Energy Laboratory as a post-doc. In 2008 he moved to Japan, becoming the first foreign faculty member at the Yokohama National University Graduate School of Engineering. He now heads his independent research group, focusing on the quantum theory of molecules and solids.

DOI: 10.1103/Physics.3.53
URL: http://link.aps.org/doi/10.1103/Physics.3.53

© 2010 American Physical Society