Atomic-orbital-based approximate self-interaction correction scheme for molecules and solids

C. D. Pemmaraju,[1] T. Archer,[1] D. Sánchez-Portal,[2] and S. Sanvito[1]

[1]*School of Physics, Trinity College, Dublin 2, Ireland*
[2]*Unidad de Fisica de Materiales, Centro Mixto CSIC-UPV/EHU and Donostia International Physics Center (DIPC), Paseo Manuel de Lardizabal 4, 20018 Donostia-San Sebastián, Spain*

We present an atomic-orbital-based approximate scheme for self-interaction correction (SIC) to the local-density approximation (LDA) of density-functional theory. The method, based on the idea of Filippetti and Spaldin [Phys. Rev. B **67**, 125109 (2003)], is implemented in a code using localized numerical atomic-orbital basis sets and is now suitable for both molecules and extended solids. After deriving the fundamental equations as a nonvariational approximation of the self-consistent SIC theory, we present results for a wide range of molecules and insulators. In particular, we investigate the effect of re-scaling the self-interaction correction and we establish a link with the existing atomiclike corrective scheme LDA+U. We find that when no re-scaling is applied, i.e., when we consider the entire atomic correction, the Kohn-Sham highest occupied molecular orbital (HOMO) eigenvalue is a rather good approximation to the experimental ionization potential for molecules. Similarly the HOMO eigenvalues of negatively charged molecules reproduce closely the molecular affinities. In contrast a re-scaling of about 50% is necessary to reproduce insulator band gaps in solids, which otherwise are largely overestimated. The method therefore represents a Kohn-Sham based single-particle theory and offers good prospects for applications where the actual position of the Kohn-Sham eigenvalues is important, such as quantum transport.

I. INTRODUCTION

Density-functional theory (DFT), in both its static[2] and time-dependent[3] forms, has become by far the most successful and widely used among all the electronic structure methods. The most obvious reason for this success is that it provides accurate predictions of numerous properties of atoms, inorganic molecules, biomolecules, nanostructures, and solids, thus serving different scientific communities.

In addition DFT has a solid theoretical foundation. The Hohenberg-Kohn theorem[2] establishes the existence of a unique energy functional $E[\rho]$ of the electron density ρ which alone is sufficient to determine the exact ground state of a N-electron system. Although the energy functional itself is not known, several of its general properties can be demonstrated rigorously. These are crucial for constructing increasingly more predictive approximations to the functional and for addressing the failures of such approximations.[4]

Finally, but no less important, the Kohn-Sham (KS) formulation of DFT (Ref. 5) establishes a one-to-one mapping of the intrinsically many-body problem onto a fictitious single-particle system and offers a convenient way for minimizing $E[\rho]$. The degree of complexity of the Kohn-Sham (KS) problem depends on the approximation chosen for the density functional. In the case of the local-density approximation (LDA),[5] the KS problem typically scales as N,[3] where the scaling is dominated by the diagonalization algorithm. However, clever choices with regard to basis sets and sophisticated numerical methods make order-N scaling a reality.[6,7]

The energy functional $E[\rho^\uparrow, \rho^\downarrow]$ (ρ^σ, $\sigma = \uparrow, \downarrow$ is the spin density, $\rho = \Sigma_\sigma \rho^\sigma$) can be written as

$$E[\rho^\uparrow, \rho^\downarrow] = T_S[\rho] + \int d^3\mathbf{r}\rho(\mathbf{r})v(\mathbf{r}) + U[\rho] + E_{xc}[\rho^\uparrow, \rho^\downarrow],$$

(1)

where T_S is the kinetic energy of the noninteracting system, $v(\mathbf{r})$ is the external potential, U is the Hartree electrostatic energy, and E_{xc} is the exchange and correlation (XC) energy. This last term is unknown and must be approximated. The construction of an approximated functional follows two strategies: empirical and "constraint satisfaction."

Empirical XC functionals usually violate some of the constraints imposed by exact DFT, and rely on parametrizations obtained by fitting representative data. One includes in this category XC functionals which borrow some functional dependence from other theories. This is, for instance, the case of the celebrated LDA+U scheme,[8,9] where the Hubbard-U energy takes the place of the LDA energy for certain "strongly correlated" atomic orbitals (typically d and f shells). The method, however, depends on the knowledge of the Coulomb and exchange parameters U and J, which vary from material to material, and can also be different for the same ion in different chemical environments.[10,11]

In contrast, the construction based on constraint satisfaction proceeds by developing increasingly more sophisticated functionals, which, nevertheless, satisfy most of the properties of exact DFT.[12] It was argued that this method constructs a "Jacob's ladder,"[13] where functionals are assigned to different rungs depending on the number of ingredients they include. Thus the LDA, which depends only on the spin densities, is on the first rung, the generalized gradient approximation (GGA),[14] which depends also on $\nabla\rho^\sigma$, is on the second rung, the so-called meta-GGA functionals,[15] which in addition to ρ^σ and $\nabla\rho^\sigma$ depend on either the Laplacian $\nabla^2\rho^\sigma$

or the orbital kinetic energy density, are on the third rung, and so on. The higher its position on the ladder the more accurate a functional becomes, but at the price of increasing computational overheads. Therefore it is worth investigating corrections to the functionals of the lower rungs, which preserve most of the fundamental properties of DFT and do not generate massive additional numerical overheads.

One of the fundamental problems intrinsic to the semilocal functionals of the first three rungs is the presence of self-interaction (SI).[16] This is the spurious interaction of an electron in a given KS orbital with the Hartree and XC potential generated by itself. Such an interaction is not present in the Hartree-Fock method, where the Coulomb self-interaction of occupied orbitals is canceled exactly by the nonlocal exchange. However, when using semilocal functionals such a cancellation is not complete and the rigorous condition for KS-DFT,

$$U[\rho_n^\sigma] + E_{xc}[\rho_n^\sigma, 0] = 0, \qquad (2)$$

for the orbital density $\rho_n^\sigma = |\psi_n^\sigma|^2$ of the fully occupied KS orbital ψ_n^σ is not satisfied. A direct consequence of the self-interaction in LDA/GGA is that the KS potential becomes too repulsive and exhibits an incorrect asymptotic behavior.[16]

This "schizophrenic" (self-interacting) nature of semilocal KS potentials generates a number of failures in describing elementary properties of atoms, molecules, and solids. Negatively charged ions (H^-, O^-, F^-) and molecules are predicted to be unstable by LDA,[17] and transition-metal oxides are described as small-gap Mott-Hubbard antiferromagnets (MnO, NiO),[18] or even as ferromagnetic metals (FeO, CoO),[18] instead of charge-transfer insulators. Moreover, the KS highest occupied molecular orbital (HOMO), the only KS eigenvalue that can be rigorously associated to a single-particle energy,[19–21] is usually nowhere near the actual ionization potential.[16]

Finally XC functionals affected by SI do not present a derivative discontinuity as a function of the occupation.[19,20] Semilocal functionals in fact continuously connect the orbital levels of systems of different integer occupation. This means, for instance, that when adding an extra electron to an open shell N-electron system the KS potential does not jump discontinuously by $I_N - A_N$ where I_N and A_N are, respectively, the ionization potential and the electron affinity for the N-electron system. This serious drawback is responsible for the incorrect dissociation of heteronuclear molecules into charged ions[22] and for the metallic conductance of insulating molecules.[23]

The problem of removing the SI from a semilocal potential was acknowledged a long time ago when Fermi and Amaldi proposed a first crude correction.[24] However, the modern theory of self-interaction corrections (SICs) in DFT is due to the original work of Perdew and Zunger from almost three decades ago.[16] Their idea consists in removing directly the self-Hartree and self-XC energy of all the occupied KS orbitals from the LDA XC functional (the same argument is valid for other semilocal functionals), thus defining the SIC functional as

$$E_{xc}^{SIC}[\rho^\uparrow, \rho^\downarrow] = E_{xc}^{LDA}[\rho^\uparrow, \rho^\downarrow] - \sum_{n\sigma}^{\text{occupied}} (U[\rho_n^\sigma] + E_{xc}^{LDA}[\rho_n^\sigma, 0]). \qquad (3)$$

Although apparently simple, the SIC method is more involved than standard KS DFT. The theory is still a density-functional one, i.e., it satisfies the Hohenberg-Kohn theorem, however, it does not fit into the Kohn-Sham scheme, since the one-particle potential is orbital dependent. This means that one cannot define a kinetic energy functional independently from the choice of E_{xc}.[16] Two immediate consequences are that the ψ_n^σ are not orthogonal, and that the orbital-dependent potential can break the symmetry of the system. This last aspect is particularly important for solids since one has to give up the Bloch representation.

In this paper we explore an approximate method for SIC to the LDA, which has the benefit of preserving the local nature of the LDA potential, and therefore maintains all of the system's symmetries. We have followed in the footsteps of Filippetti and Spaldin,[1] who extended the original idea of Vogel and co-workers[25–27] of considering only the atomic contributions to the SIC. We have implemented such a scheme into the localized atomic-orbital code SIESTA,[28] and applied it to a vast range of molecules and solids. In particular we have investigated in detail how the scheme performs as a single-particle theory, and how the SIC should be rescaled in different chemical environments.

II. REVIEW OF EXISTING METHODS

The direct subtraction proposed by Perdew and Zunger is the foundation of the modern SIC method. However, the minimization of the SIC functional (3) is not trivial, in particular for extended systems. The main problem is that E_{xc} itself depends on the KS orbitals. Thus it does not fit into the standard KS scheme and a more complicated minimization procedure is needed.

Following the minimization strategy proposed by Levy,[29] which prescribes to minimize the functional first with respect to the KS orbitals ψ_n^σ and then with respect to the occupation numbers p_n^σ, Perdew and Zunger derived a set of single-particle equations,

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff},n}^\sigma(\mathbf{r}) \right] \psi_n^\sigma = \epsilon_n^{\sigma,\text{SIC}} \psi_n^\sigma, \qquad (4)$$

where the effective single-particle potential $v_{\text{eff},n}^\sigma(\mathbf{r})$ is defined as

$$v_{\text{eff},n}^\sigma(\mathbf{r}) = v(\mathbf{r}) + u([\rho]; \mathbf{r}) + v_{xc}^{\sigma,\text{LDA}}([\rho^\uparrow, \rho^\downarrow]; \mathbf{r}) - u([\rho_n]; \mathbf{r})$$
$$- v_{xc}^{\sigma,\text{LDA}}([\rho_n^\uparrow, 0]; \mathbf{r}), \qquad (5)$$

and

$$u([\rho]; \mathbf{r}) = \int d^3\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \qquad (6)$$

$$v_{xc}^{\sigma,\text{LDA}}([\rho^\uparrow, \rho^\downarrow]; \mathbf{r}) = \frac{\delta}{\delta\rho^\sigma(\mathbf{r})} E_{xc}^{LDA}[\rho^\uparrow, \rho^\downarrow]. \qquad (7)$$

These are solved in the standard KS way for atoms, with good results in terms of quasiparticle energies.[16] In this particular case the KS orbitals ψ_n^σ show only small deviations from orthogonality, which is re-imposed with a standard Schmidt orthogonalization.

The problem of the nonorthogonality of the KS orbitals can be easily solved by imposing the orthogonality condition as a constraint to the energy functional, thus obtaining the following single-particle equation:

$$\left[-\frac{1}{2}\boldsymbol{\nabla}^2 + v_{\text{eff},n}^\sigma(\mathbf{r}) \right]\psi_n^\sigma = \sum_m \epsilon_{nm}^{\sigma,\text{SIC}}\psi_m^\sigma. \tag{8}$$

Even in this case where orthogonality is imposed, two major problems remain: the orbitals minimizing the energy functional are not KS-type and in general do not satisfy the system's symmetries.

If one insists in minimizing the energy functional in a KS fashion by constructing the orbitals according to the symmetries of the system, the theory will become size-inconsistent, or in other words it will be dependent on the particular representation employed. Thus one might arrive at a paradox, where in the self-interaction of N hydrogen atoms arranged on a regular lattice of large lattice spacing (in such a way that there is no interaction between the atoms) vanishes, since the SIC of a Bloch state vanishes for $N \to \infty$. However, the SIC for an individual H atom, when calculated using atomiclike orbitals, accounts for essentially all the ground-state energy error.[16] Therefore a size-consistent theory of SIC DFT must look for a scheme where a unitary transformation of the occupied orbitals, which minimizes the SIC energy, is performed. This idea is at the foundation of all modern implementations of SICs.

Significant progress towards the construction of a size-consistent SIC theory was made by Pederson, Heaton, and Lin, who introduced two sets of orbitals: localized orbitals ϕ_n^σ minimizing $E_{\text{xc}}^{\text{SIC}}$ and canonical (Kohn-Sham) delocalized orbitals ψ_n^σ.[30–32] The localized orbitals are used for defining the densities entering into the effective potential (5), while the canonical orbitals are used for extracting the Lagrangian multipliers $\epsilon_{nm}^{\sigma,\text{SIC}}$, which are then associated to the KS eigenvalues. The two sets are related by unitary transformation $\psi_n^\sigma = \sum_m \mathcal{M}_{nm}^\sigma \phi_m^\sigma$, and one has two possible strategies for minimizing the total energy.

The first consists in a direct minimization with respect to the localized orbitals ϕ_n^σ, i.e., in solving Eq. (8) when we replace ψ with ϕ and the orbital densities entering the definition of the one-particle potential (5) are simply $\rho_n^\sigma = |\phi_n^\sigma|^2$. In addition the following minimization condition must be satisfied:

$$\langle \phi_n^\sigma | v_n^{\text{SIC}} - v_m^{\text{SIC}} | \phi_m^\sigma \rangle = 0, \tag{9}$$

where $v_n^{\text{SIC}} = u([\rho_n]; \mathbf{r}) + v_{\text{xc}}^{\sigma,\text{LDA}}([\rho_n^\uparrow, 0]; \mathbf{r})$. An expression for the gradient of the SIC functional, which also constrains the orbitals to be orthogonal, has been derived[33] and applied to atoms and molecules with a mixture of successes and bad failures.[34–36]

The second strategy uses the canonical orbitals ψ and seeks the minimization of the SIC energy by varying both the orbitals ψ and the unitary transformation $\mathcal{M}$. The corresponding set of equations is

$$H_n^\sigma \psi_n^\sigma = (H_0^\sigma + \Delta v_n^{\text{SIC}})\psi_n^\sigma = \sum_m \epsilon_{nm}^{\sigma,\text{SIC}}\psi_m^\sigma, \tag{10}$$

$$\psi_n^\sigma = \sum_m \mathcal{M}_{nm}\phi_m^\sigma, \tag{11}$$

$$\Delta v_n^{\text{SIC}} = \sum_m \mathcal{M}_{nm}v_m^{\text{SIC}}\frac{\phi_m^\sigma}{\psi_n^\sigma}, \tag{12}$$

where H_0^σ is the standard LDA Hamiltonian (without SIC). Thus the SIC potential for the canonical orbitals appears as a weighted average of the SIC potential for the localized orbitals. The solutions of the set of equations (10) is somehow more appealing than that associated to the localized orbitals since the canonical orbitals can be constructed in a way that preserves the system's symmetries (for instance, translational invariance).

A convenient way for solving Eq. (10) is that of using the so-called "unified Hamiltonian" method.[30] This is defined as (we drop the spin index σ)

$$H_{\text{u}} = \sum_n^{\text{occup}} \hat{P}_n H_0 \hat{P}_n + \sum_n^{\text{occup}} (\hat{P}_n H_n \hat{Q} + \hat{Q}H_n\hat{P}_n) + \hat{Q}H_0\hat{Q}, \tag{13}$$

where $\hat{P}_n = |\psi_n^\sigma\rangle\langle\psi_n^\sigma|$ is the projector over the occupied orbital ψ_n^σ, and $\hat{Q}$ is the projector over the unoccupied ones $\hat{Q} = 1 - \sum_n^{\text{occup}}\hat{P}_n$. The crucial point is that the diagonal elements of the matrix $\epsilon_{nm}^{\sigma,\text{SIC}}$ and their corresponding orbitals ψ_n^σ are, respectively, eigenvalues and eigenvectors of H_{u}, from which the whole $\epsilon_{nm}^{\sigma,\text{SIC}}$ can be constructed. Finally, and perhaps most important, at the minimum of the SIC functional, the canonical orbitals diagonalize the matrix $\epsilon_{nm}^{\sigma,\text{SIC}}$, whose eigenvalues can now be interpreted as an analog of the Kohn-Sham eigenvalues.[32]

It is also interesting to note that an alternative way for obtaining orbital energies is that of constructing an effective SI-free local potential using the Krieger-Li-Iafrate method.[37] This has been recently explored by several groups.[38–40]

When applied to extended systems the SIC method demands considerable additional computational overheads over standard LDA. Thus for a long time it has not encountered the favor of the general solid-state community. In the case of solids the price to pay for not using canonical orbitals is enormous since the Bloch representation should be abandoned and in principle infinite cells should be considered. For this reason the second minimization scheme, in which the canonical orbitals are in a Bloch form, is more suitable. In this case for each $\mathbf{k}$-vector one can derive an equation identical to Eq. (10), where $\epsilon_{nm}^{\sigma,\text{SIC}} = \epsilon_{nm}^{\sigma,\text{SIC}}(\mathbf{k})$ and n is simply the band index.[41] The associated localized orbitals ϕ, for

instance, can be constructed as Wannier functions and the minimization scheme proceeds in a similar way to that done for molecules.

The problem here is that in practice, the cell needed to describe the localized states ϕ may be considerably larger than the primitive unit cell. This is not the case for ionic insulators,[41] where the localized orbitals are well approximated by atomic orbitals. Such a simplification is, however, not valid in general. For example supercells as large as 500 atoms have been considered for describing the localized d shells in transition-metal oxides.[42–44] Despite these difficulties the SIC scheme has been applied to a vast range of solid-state systems with systematic improvement over LDA. These include, in addition to transition-metal monoxides,[42,44,45] rare-earth materials,[46] diluted magnetic semiconductors,[47] Fe_3O_4,[48] heavy element compounds,[49] just to name a few.

In order to make the SIC method more scalable several approximations have been proposed. One possibility is that of incorporating part of the SIC into the definition of the pseudopotentials.[50] The idea consists in subtracting the atomic SI from the free atom, and then transferring the resulting electronic structure to the definition of a standard norm-conserving pseudopotential. This approximation is sustained by the fact that the transformation $\mathcal{M}$, which relates canonical and localized orbitals, does hardly mix core and valence states.[32] Thus the SIC contribution to the total energy can be separated into the contributions from the core and the valence and in first approximation the latter can be neglected.[51] The benefit of this method is that translational invariance is regained and the complicated procedure of minimizing $\mathcal{M}$ is replaced by a pseudopotential calculation.

A further improvement over the pseudopotential method was recently presented by Vogel and co-workers[25–27] and then extended by Filippetti and Spaldin.[1] The method still assumes separability between the core and the valence contributions to the SIC, but now the SIC for the valence electrons is approximated by an atomiclike contribution, instead of being neglected. This atomic SIC (ASIC) scheme is certainly a drastic approximation, since it implicitly assumes that the transformation $\mathcal{M}$ minimizing the SIC functional leads to atomiclike orbitals.

In the work of Vogel this additional contribution is not evaluated self-consistently for the solid, while the implementation of Filippetti assumes a linear dependance of the SIC over the orbital occupation. In spite of the approximations involved, the method has been applied successfully to a range of solids including II-VI semiconductors and nitrites,[1,25,26] transition metal monoxides,[1,52] silver halides,[27] noble metal oxides,[53] ferroelectric materials,[1,54,55] high-k materials[56] and diluted magnetic semiconductors.[57,58] Interestingly most of the systems addressed by the ASIC method are characterized by semicore d orbitals, for which an atomic correction looks appropriate, and a similar argument is probably valid for ionic compounds as recently demonstrated for the case of SiC.[59]

Here we further investigate the self-consistent ASIC method[1] by examining both finite and extended systems, and by critically considering whether a scaling factor, additional to the orbital occupation, is needed for reproducing the correct single-particle spectrum.

III. FORMALISM AND IMPLEMENTATION

In this section we derive the fundamental equations of the ASIC method, while looking closely at the main approximations involved in comparison to the fully self-consistent SIC approach. Our practical implementation is also described.

A. ASIC potential

The starting point of our analysis is the SIC Schrödinger-like equation (10) for the canonical orbitals. Let us assume, as from Ref. 51, that the rotation $\mathcal{M}$ transforming localized orbitals (to be determined) into canonical orbitals [see Eq. (11)] does not mix core and valence states. We also assume that core electrons are well localized into atomiclike wavefunctions and that they can be effectively described by a norm-conserving pseudopotential.

Let us now assume that $\mathcal{M}$ is known and so are the localized orbitals ϕ_m^σ. In this case the canonical orbitals diagonalize $\epsilon_{nm}^{\sigma,\mathrm{SIC}}$ and Eq. (10) simply reduces to

$$(H_0^\sigma + \Delta v_n^{\mathrm{SIC}})\psi_n^\sigma = \epsilon_n^{\sigma,\mathrm{SIC}}\psi_n^\sigma, \qquad (14)$$

with $\Delta v_n^{\mathrm{SIC}}$ defined in Eq. (12). The Hamiltonian $H_0^\sigma + \Delta v_n^{\mathrm{SIC}}$ can be then rewritten in a convenient form as

$$H_0^\sigma + \Delta v_n^{\mathrm{SIC}} = H_0^\sigma + \sum_m v_m^{\mathrm{SIC}}\hat{P}_m^\phi, \qquad (15)$$

where v_m^{SIC} is the self-interaction potential for the localized orbital ϕ_m^σ, and $\hat{P}_m^\phi$ is the projector over the same state,

$$\hat{P}_m^\phi \psi_n^\sigma(\mathbf{r}) = \phi_m^\sigma(\mathbf{r}) \int d^3\mathbf{r}' \psi_n^\sigma(\mathbf{r}') \phi_m^{\sigma\dagger}(\mathbf{r}') = \phi_m^\sigma(\mathbf{r})\langle\phi_m^\sigma|\psi_n^\sigma\rangle. \qquad (16)$$

Two main approximations are then taken in the ASIC approach.[1,25] First the localized states ϕ_m^σ are assumed to be atomiclike orbitals Φ_m^σ (ASIC orbitals) and the SIC potential is approximated as

$$\sum_m v_m^{\mathrm{SIC}}(\mathbf{r})\hat{P}_m^\phi \rightarrow \alpha \sum_m \tilde{v}_m^{\sigma\mathrm{SIC}}(\mathbf{r})\hat{P}_m^\Phi, \qquad (17)$$

with $\tilde{v}_m^{\sigma\mathrm{SIC}}(\mathbf{r}) = u([\rho_m];\mathbf{r}) + v_{\mathrm{xc}}^{\sigma,\mathrm{LDA}}([\rho_m^\uparrow,0];\mathbf{r})$ and $\rho_m^\sigma = |\Phi_m^\sigma|^2$, $\hat{P}_m^\Phi$ is the projector of Eq. (16) obtained by replacing the ϕ's with the ASIC orbitals Φ, and α is a scaling factor. Note that the orbitals Φ_m are not explicitly spin dependent and one simply has $\Phi_m^\sigma = \Phi_m p_m^\sigma$ with p_m^σ the orbital occupation ($p_m^\sigma = 0, 1$). The factor α is an empirical factor, which accounts for the particular choice of ASIC orbitals. This first approximation is expected to be accurate for systems retaining an atomiclike charge density as in the case of small molecules. It is also formally exact in the one-electron limit (for $\alpha = 1$). In the case of extended solids the situation is less transparent, since in general the functions minimizing $E_{\mathrm{xc}}^{\mathrm{SIC}}$ are Wannier-like functions.[60]

The second approximation taken in the ASIC method is that of replacing the nonlocal projector $\hat{P}_m^\Phi$ with its expectation value. The idea is that the SIC potential for the canonical orbitals $\Delta v_n^{\mathrm{SIC}}$ is formally a weighted average of the SIC

potential for the localized orbitals v_m^{SIC}. For the exact SIC method the weighting factor is the nonlocal projector $\mathcal{M}_{nm}\frac{\phi_m^\sigma}{\psi_n^\sigma}$. This means that the SIC potential for a given canonical orbital ψ_n is maximized in those regions where the overlap with some of the localized orbitals ϕ_m is maximum. In the ASIC method such nonlocal projector is replaced more conveniently by a scalar. In the original proposal by Vogel and co-workers[25–27] this was simply set to 1. Here we consider the orbital occupation p_m^σ of the given ASIC orbital Φ_m, i.e., we replace $\hat{P}_m^\Phi$ with its expectation value

$$\hat{P}_m^\Phi \rightarrow \langle \hat{P}_m^\Phi \rangle = p_m^\sigma = \sum_n f_n^\sigma \langle \psi_n^\sigma | \hat{P}_m^\Phi | \psi_n^\sigma \rangle, \qquad (18)$$

where f_n^σ is the occupation number of the Kohn-Sham orbital ψ_n^σ. The final form of ASIC potential is then

$$v_{\mathrm{ASIC}}^\sigma(\mathbf{r}) = \alpha \sum_m \tilde{v}_m^{\sigma\mathrm{SIC}}(\mathbf{r}) p_m^\sigma. \qquad (19)$$

Let us now comment on the empirical scaling factor α. In Ref. 1 α was set to $1/2$ in order to capture eigenvalue relaxation. This choice, however, violates the one-electron limit of the SIC potential, which is correctly reproduced for $\alpha=1$. We can then interpret α as a measure of the deviation of the ASIC potential from the exact SIC potential. Ultimately α reflects the deviation of the actual ASIC projectors $|\Phi\rangle\langle\Phi|$ from the localized orbitals defining the SI corrected ground state. One then expects α to be close to 1 for systems with an atomiclike charge density, and to vanish for metals, whose valence charge density resembles that of a uniform electron gas.[61] Intermediate values are then expected for situations different from these two extremes, and we will show that a values around $1/2$ describe well a vast class of mid-to-wide-gap insulators.

B. Implementation

The final form of the SIC potential to subtract from the LSDA (local spin-density approximation) one [Eq. (19)] is that of a linear combination of nonlocal pseudopotential-like terms. These are uniquely defined by the choice of exchange and correlation potential used and by ASIC orbitals Φ_m. The practical way of constructing such potentials, i.e., the way of importing the atomic SIC to the solid state, depends on the specific numerical implementation used for the DFT algorithm. At present plane-wave and Gaussian orbital implementations are available,[1,25–27] while here we present our scheme based on the pseudoatomic orbital (PAO) (Ref. 65) code SIESTA.[28]

We start by solving the atomic all-electron SIC-LSDA equation for all the species involved in the solid-state calculation. Here we apply the original Perdew-Zunger (PZ-SIC) formalism[16] and we neglect the small nonorthogonality between the Kohn-Sham orbitals. Thus we obtain a set of SI corrected atomic orbitals Φ_m, which exactly solve the atomic SIC-LSDA problem. The atomic orbitals Φ_m describing the valence electrons are then used to define the ASIC potentials $\tilde{v}_m^{\sigma\mathrm{SIC}}$,

$$\tilde{v}_m^{\sigma\mathrm{SIC}}(\mathbf{r}) = u([\rho_m];\mathbf{r}) + v_{\mathrm{xc}}^{\sigma,\mathrm{LDA}}([\rho_m^\uparrow,0];\mathbf{r}) \qquad (20)$$

with $\rho_m^\sigma = |\Phi m|^2$.

At the same time a standard LSDA calculation for the same atoms is used to construct the pseudopotentials describing the core electrons. These are standard norm-conserving scalar relativistic Troullier-Martins pseudopotentials[62] with nonlinear core corrections.[63] Thus we usually neglect the SIC over the core states, when constructing the pseudopotentials. This is justified by the fact that the eigenvalues for the SIC-LSDA pseudoatom, i.e., for the free atom where the effects of core electrons are replaced by LSDA pseudopotentials but SIC is applied to the valence electrons, are in excellent agreement with those obtained by all-electron SIC-LSDA calculations.[25]

The final step is that of recasting the ASIC potentials $\tilde{v}_m^{\mathrm{SIC}}(\mathbf{r})$, which have a $-2/r$ asymptotic behavior, in a suitable nonlocal form. This is obtained with the standard Kleinman-Bylander[64] scheme and the final ASIC potential [Eq. (19)] is written as

$$v_{\mathrm{ASIC}}^\sigma = \sum_m \frac{|\gamma_m^\sigma\rangle\langle\gamma_m^\sigma|}{C_m^\sigma}, \qquad (21)$$

where the ASIC projectors are given by

$$\gamma_m^\sigma(\mathbf{r}) = \alpha p_m^\sigma \tilde{v}_m^{\sigma\mathrm{SIC}}(\mathbf{r}) \Phi_m'(\mathbf{r}). \qquad (22)$$

and the normalization factors are

$$C_m^\sigma = \alpha p_m^\sigma \langle \Phi_m' | \tilde{v}_m^{\sigma\mathrm{SIC}} | \Phi_m' \rangle. \qquad (23)$$

The orbital functions Φ_m' are atomiclike functions with a finite range, which ensure that the ASIC projectors γ_m do not extend beyond that range. These are constructed in the same way as the SIESTA basis set orbitals, i.e., as solutions of the pseudoatomic problem with an additional confining potential at the cutoff radius r_{cutoff}.[65] The choice of the appropriate cutoff radius for the SIC projectors should take into account the two following requirements. On the one hand it should be sufficiently large to capture most of the SIC corrections. A good criterion[1] is that the SIC-LSDA contribution to the orbital energy of the free atom,

$$\delta\varepsilon_{\mathrm{SIC}m}^\sigma = \langle \Phi_m' | \tilde{v}_m^{\sigma\mathrm{SIC}} | \Phi_m' \rangle, \qquad (24)$$

is reproduced within some tolerance. On the other hand, the cutoff should be reasonably short so as not to change the connectivity of the matrix elements of the PAO Hamiltonian. In other words, we need to ensure that orbitals otherwise considered as disconnected in evaluating the various parts of the Hamiltonian matrix are not considered connected for the v_{ASIC}^σ matrix elements alone.

As a practical rule we set the cutoff radius for a particular orbital of a given atom to be either equal to the longest among the cutoff radii of the PAO basis set for that particular atom (typically the first ζ of the lowest angular momentum), or, if shorter, the radius at which $\delta\varepsilon_{\mathrm{SIC}m}^\sigma < 0.1$ mRy. Typically, when reasonable cutoff radii (6–9 bohr) are used, we find that the atomic SIC-LSDA eigenvalues are reproduced to within 1–5 mRy for the most extended shells and to within 0.1 mRy for more confined shells. Thus $\delta\varepsilon_{\mathrm{SIC}m}^\sigma$ are

rather well converged already for cutoff radii defined by a PAO energies shifts[28] of around 20 mRy, although usually smaller PAO energy shifts are necessary for highly converged total energy calculations.

Finally the matrix elements of the SIC potential are calculated as usual over the SIESTA basis set. Additional basis functions χ_m are constructed from the confined localized atomic orbitals described before using the split-norm scheme.[65–67] The density matrix ρ^σ is represented over such basis $\rho^\sigma_{\mu\nu}$ and the orbital populations are calculated as

$$p_m^\sigma = \sum_{\mu\nu} S_{m\mu} \rho_{\mu\nu}^\sigma S_{\nu m}, \qquad (25)$$

where $S_{m\mu}$ is a matrix element of the overlap matrix. Note that in principle the orbital population should be constructed for the atomic SIC orbital Φ_m. However, we notice that p_m^σ is rather insensitive to the specific choice of orbital, once this has a reasonable radial range. For practical numerical reasons in the present implementation, we always use the orbital populations projected onto the basis set subspace consisting of the most extended first-ζ orbitals of the atomic species involved. The matrix elements of the SIC potential are simply

$$(v_{ASIC}^\sigma)_{\mu\nu} = \sum_m \frac{\langle \chi_\mu | \gamma_m^\sigma \rangle \langle \gamma_m^\sigma | \chi_\nu \rangle}{C_m^\sigma}, \qquad (26)$$

and the ASIC-KS equation takes the final form

$$\left[-\frac{1}{2}\boldsymbol{\nabla}^2 + v_{PP} + u + v_{xc}^{\sigma,LSDA} - v_{ASIC}^\sigma \right] \psi_n^\sigma = \epsilon_n^{\sigma,SIC} \psi_n^\sigma \quad (27)$$

with v_{PP} the pseudopotential.

C. Total energy

The energy corresponding to the SIC-LSDA functional is given by[16]

$$E^{SIC}[\rho^\uparrow, \rho^\downarrow] = E^{LSDA}[\rho^\uparrow, \rho^\downarrow] - \sum_{n\sigma}^{occ.} (U[\rho_n^\sigma] + E_{xc}^{LSDA}[\rho_n^\sigma, 0]), \qquad (28)$$

where

$$U[\rho_n^\sigma] = \int d^3\mathbf{r} \frac{1}{2} \rho_n(\mathbf{r}) u([\rho_n];\mathbf{r}), \qquad (29)$$

$$E_{xc}^{LSDA}[\rho_n^\sigma, 0]) = \int d^3\mathbf{r} \rho_n(\mathbf{r}) \mathcal{E}_{xc}^{LSDA}([\rho_n];\mathbf{r}), \qquad (30)$$

with $\mathcal{E}_{xc}^{LSDA}$ the LSDA exchange and correlation energy density. The orbital densities entering in the SI term are those associated to the local orbitals ϕ. As we have already mentioned, this functional needs to be minimized with respect to the ϕ's, which are an implicit function of the total spin density ρ^σ. In the ASIC approximation these orbitals are not minimized, but taken as atomic functions. This means that in the present form the theory is not variational, in the sense that there is no functional related to the KS equation (27) by

a variational principle. With this in mind we adopt the expression of Eq. (28) as a suitable energy, where the orbital densities are those given by the ASIC orbitals,

$$\rho_m^\sigma(\mathbf{r}) = p_m^\sigma |\Phi_m|^2. \qquad (31)$$

In our implementation the LSDA KS energy E^{LSDA} is directly available as calculated in the SIESTA code[28] and thus only the second term of Eq. (28) needs to be calculated. This is easily done by calculating both U and E_{xc}^{LSDA} on an atomic radial grid for each atomic orbital in the system.

D. ASIC and LDA+U

We now compare our ASIC method with another atomic-like correction to LSDA, namely the LDA+U method.[8,9] In LDA+U one replaces the LSDA exchange and correlation energy associated to the "correlated" orbitals (d or f shells), with the Hubbard-U energy. Thus the functional becomes

$$E^{LDA+U}[\rho(\mathbf{r})] = E^{LSDA}[\rho(\mathbf{r})] + E^U[\{p_m^\sigma\}] - E^{DC}[\{p_m^\sigma\}], \qquad (32)$$

where the Hubbard energy E^U and the double counting term E^{DC} depend on the orbital populations p_m^σ of the correlated orbitals. Several forms for the LDA+U functional have been proposed to date. A particularly simple and transparent one,[10,68] which is also rotationally invariant, redefines the U parameter as an effective parameter $U_{eff} = U - J$ and the functional takes the form

$$E^U - E^{DC} = \frac{U_{eff}}{2} \sum_I \sum_{m\sigma} \left[p_{mm}^{I\sigma} - \sum_n p_{mn}^{I\sigma} p_{nm}^{I\sigma} \right], \qquad (33)$$

wherein we separate out the index for the atomic position I from the magnetic quantum number m, and introduce the off-diagonal populations $p_{mn}^{I\sigma} = \Sigma_\alpha f_\alpha^\sigma \langle \psi_\alpha^\sigma | \hat{P}_{mn}^{I\Phi} | \psi_\alpha^\sigma \rangle$ with $\hat{P}_{mn}^{I\Phi} = |\Phi_m^I\rangle\langle\Phi_n^I|$. Note that the LDA+$U$ functional is SI free for those orbitals that are corrected.

Although a rotationally invariant form of the ASIC potential can be easily derived, we assume here for simplicity that the system under consideration is rotationally invariant, or alternatively that we have carried out a rotation, which diagonalizes the $p_{mn}^{I\sigma}$ matrix. In this case the energy becomes simply

$$E^U - E^{DC} = \frac{U_{eff}}{2} \sum_{Im\sigma} p_m^{I\sigma} [1 - p_m^{I\sigma}], \qquad (34)$$

with $p_m^{I\sigma} = p_{mm}^{I\sigma}$. It is then easy to compute the KS potential

$$v^{LDA+U} = v^{LSDA} + \sum_{Im\sigma} U_{eff} \left[\frac{1}{2} - p_m^{I\sigma} \right] \hat{P}_m^{I\Phi}, \qquad (35)$$

and the orbital energy

$$\epsilon_m^{I\sigma} = \frac{\partial E}{\partial p_m^{I\sigma}} = \epsilon_m^{I\sigma LSDA} + U_{eff} \left(\frac{1}{2} - p_m^{I\sigma} \right). \qquad (36)$$

These need to be compared with the ASIC potential [Eq. (19)] and orbital energy

$$\epsilon_m^{I\sigma} = \epsilon_m^{I\sigma\text{LSDA}} - \alpha p_m^{I\sigma}\langle\Phi_m^{I\prime}|\tilde{v}_{Im}^{\sigma\text{SIC}}|\Phi_m^{I\prime}\rangle, \qquad (37)$$

where the last term follows from $\frac{\partial E}{\partial p_m^{I\sigma}} = C_m^{I\sigma}$ and from Eq. (28). The main difference between the ASIC and LDA+U methods is in the way in which unoccupied states are handled. In fact, while LDA+U corrects unoccupied states and pushes the orbital energies upwards by $\sim U_{\text{eff}}/2$, ASIC operates only on occupied states, that are shifted towards lower energies by $C_m^{I\sigma}$. This reflects the fact that the SIC is defined only for occupied KS orbitals. An important consequence is that the opening of band gaps in the electronic structures, one of the main features of both the LDA+U and ASIC schemes, is then driven by two different mechanisms. On the one hand in LDA+U, gaps open up since occupied and unoccupied states are corrected in opposite directions leading to a gap of $\sim U_{\text{eff}}$. On the other hand ASIC is active only over occupied states and gaps open only if occupied and unoccupied bands have large differences in their projected atomic orbital content. Thus one should not expect any corrections for covalent materials where conduction and valence bands are simply bonding and antibonding states formed by the same atomic orbitals. This is, for instance, the case of Si and Ge. In contrast ASIC will be extremely effective for more ionic situations, where the orbital contents of conduction and valence bands are different.

Finally, by comparing the corrections to the orbital energy of a fully occupied state, one finds

$$U = 2\alpha\langle\Phi_m^{I\prime}|\tilde{v}_{Im}^{\sigma\text{SIC}}|\Phi_m^{I\prime}\rangle, \qquad (38)$$

which establishes an empirical relation between the Hubbard energy and the ASIC correction. Since U is sensitive to the chemical environment due to screening,[10] while all the other quantities are uniquely defined by an atomic calculation, we can re-interpret the parameter α as empirically describing the screening from the chemical environment within the ASIC scheme.

IV. RESULTS: EXTENDED SYSTEMS

The test calculations that we present in this work are for two classes of materials: extended and finite. First we investigate how our implementation performs in the solid state. In particular we discuss the role of the parameter α in determining the band structure of several semiconductors, considering both the KS band gap and the position of bands associated with tightly bound electrons.

A. Estimate of α for semiconductors

The quasiparticle band gap E_g in a semiconductor is defined as the difference between its ionization potential I and electron affinity A. These can be rigorously calculated from DFT as the HOMO energy, respectively, of the neutral and negatively charged systems. This actual gap cannot be directly compared with the KS band gap E_g^{KS}, defined as the difference between the orbital energy of the HOMO and lowest occupied molecular orbital (LUMO) states of the neutral system. In fact, the presence of a derivative discontinuity in the DFT energy as a function of the electron occupation
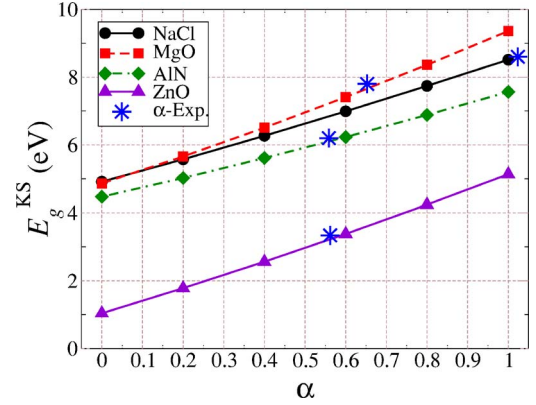

FIG. 1. (Color online) Calculated band gap for NaCl, MgO, AlN, and ZnO as a function of the parameter α. $\alpha=0$ corresponds to normal LSDA and $\alpha=1$ to using the atomic SIC at full strength. The lattice parameters used for the calculations are either the equilibrium LSDA or the experimental when available.

establishes the following rigorous relation:[20,69]

$$E_g = E_g^{\text{KS}} + \Delta_{\text{xc}}, \qquad (39)$$

with

$$\Delta_{\text{xc}} = \lim_{\omega\to 0^+}\left\{\left.\frac{\delta E_{\text{xc}}[n]}{\delta n}\right|_{N+\omega} - \left.\frac{\delta E_{\text{xc}}[n]}{\delta n}\right|_{N-\omega}\right\}_{n_N}. \qquad (40)$$

This is valid even for the exact XC potential, and therefore in principle one has to give up KS band structures as a tool for evaluating semiconductor band gaps. The size of Δ_{xc} is, however, not known for real extended systems and the question of whether most of the error in determining E_g from E_g^{KS} is due to the approximation in the XC potential or due to the intrinsic Δ_{xc} is a matter of debate.

In general, SI-free potentials bind more than LSDA and one expects larger gaps. Surprisingly, functionals based on exact exchange, provide KS gaps rather close to the experimental values.[70,71] The reason for such a good agreement is not fully understood, but it is believed that the exact KS gaps should be smaller than the actual ones.

With this in mind, we adopt a heuristic approach and we use the KS band gap as a quality indicator for interpreting the parameter α and for providing its numerical value for different classes of solids. Here we investigate the dependence of E_g^{KS} over α and we determine the value for α yielding the experimental band gap. Assuming that Δ_{xc} does not vary considerably across the materials investigated, this will allow us to relate α to the degree of localization in a semiconductor and to extract the value useful for ASIC to be an accurate single-particle theory.

In Fig. 1 we present the band gap of four representative semiconductors as a function of α together with the value needed to reproduce the experimental band gap. LSDA corresponds to $\alpha=0$ and while $\alpha=1$ accounts for the full ASIC. In general E_g^{KS} increases as α increases, as a result of the stronger SIC. The $E_g^{\text{KS}}(\alpha)$ curve is almost linear with a slope, which appears to be material specific.

TABLE I. Experimental E_g^{exp} and KS E_g^{KS} band gap (in eV) for a number of semiconductors. E_g^{KS} are calculated with both LSDA and ASIC ($\alpha=1$). In the last column we report the value of $\alpha=\alpha^*$ needed for $E_g^{exp}=E_g^{KS}$. The lattice parameters used for the calculations are either the equilibrium LSDA or the experimental when available (in Å). RS=rocksalt, WZ=wurtzite, ZB=zinc blende. The value for the experimental gaps are from the literature.

Solid	Structure	E_g^{exp}	$E_g^{KS\text{-}LSDA}$	$E_g^{KS\text{-}ASIC}$	α
LiCl	RS ($a=5.13$)	9.4[a]	6.23	9.76	0.89
NaCl	RS ($a=5.63$)	8.6[b]	4.91	8.51	1.02
KCl	RS ($a=6.24$)	8.5[c]	4.90	8.51	0.99
MgO	RS ($a=4.19$)	7.8[d]	4.86	9.36	0.65
CaO	RS ($a=4.74$)	7.08[d]	4.93	9.28	0.49
SrO	RS ($a=5.03$)	5.89[e]	4.20	7.80	0.47
AlN	WZ ($a=3.11$, $c=4.98$)	6.20[f]	4.47	7.56	0.56
GaN	WZ ($a=3.16$, $c=5.13$)	3.39[g]	2.21	5.03	0.44
lnN	WZ ($a=3.54$, $c=5.70$)	0.7[h]	0.09	2.09	0.45
ZnO	WZ ($a=3.23$, $c=5.19$)	3.43[i]	0.85	5.13	0.57
ZnS	ZB ($a=5.40$)	3.78[i]	2.47	4.90	0.53
ZnSe	ZB ($a=5.63$)	2.82[i]	1.77	3.53	0.58

[a]Reference 72.
[b]Reference 73.
[c]Reference 74.
[d]Reference 75.
[e]Reference 76.
[f]Reference 77.
[g]Reference 78.
[h]Reference 79.
[i]Reference 80.

For the most ionic compound, NaCl, the experimental gap is reproduced almost exactly by $\alpha=1$, i.e., by the full ASIC. This is somehow expected since the charge density of solid NaCl is rather close to a superposition of the Na^+ and Cl^- ionic charge densities. In this case of strongly localized charge densities the ASIC approximation is rather accurate yielding results substantially identical to those obtained with full self-consistent PZ-SIC.[41] Indeed, earlier calculations for LiCl (Ref. 41) demonstrate that the SIC band structure is rather insensitive of the localized orbitals ϕ once these have an atomiclike form.

For the other compounds the localized orbitals ϕ's are not necessarily atomiclike functions and deviations from $\alpha=1$ are expected. Interestingly we find that, for all the materials investigated, a value of around 1/2 reproduces the experimental band gap rather accurately. As an illustration, in Table I we compare the experimental band gap E_g^{exp} to the calculated E_g^{KS} for ASIC ($\alpha=1$) and LDA, for several semiconductors ranging from ionic salts to wide-gap II-VI and III-V semiconductors. We also report the value of $\alpha=\alpha^*$ needed for $E_g^{exp}=E_g^{KS}$.

Clearly for all the strongly ionic compounds (LiCl, NaCl, and KCl) the full ASIC correction $\alpha=1$ reproduces quite accurately the experimental gap and agrees with previous self-consistent SIC calculations.[81] For all the other compounds a value of around 1/2 is always adequate, confirming the initial choice of Filippetti and Spaldin. For these materials we do not find any particular regularity. In general α is large when the experimental gap is large, however, there is

no direct connection between α and the ionicity or covalency of a compound. In fact, the improvement of the band gap is not simply due to a rigid shift of the valence band, but usually corresponds to a general improvement of the whole quasiparticle spectrum. Examples for ZnO and GaN will be presented in the next section.

As a further proof of this point in Table II we present the valence bandwidth for the semiconductors investigated as calculated from LSDA ΔE_v^{LSDA} and ASIC for both $\alpha=1$ ($\Delta E_v^{ASIC_1}$) and $\alpha=\alpha^*$ ($\Delta E_v^{ASIC_{\alpha^*}}$). We also report the experimental values ΔE_v^{exp} whenever available, although a direct comparison with experiments is difficult, since these values are rather imprecise and sometimes not known. The general feature is that ASIC produces only minor corrections over LSDA, and that these corrections do not follow a generic trend. Thus while for the nitrites ASIC always increases the bandwidth, it does just the opposite for KCl, SrO, and CaO.

B. Wide-gap semiconductors: ZnO and GaN

Having established $\alpha=1/2$ as an appropriate value for II-VI and III-V semiconductors, we now look at the whole band structure (not just the fundamental gap) for a few test cases. Here we consider ZnO and GaN for which photoemission data disagree quite remarkably from LSDA calculations. In Fig. 2 we compare the band structure of wurtzite ZnO obtained, respectively, from LSDA and our ASIC.

In ZnO, the valence-band top (VBT) is predominantly oxygen $2p$ in character and the conduction-band minimum

TABLE II. Valence-band experimental bandwidth ΔE_v^{exp} compared with those obtained from ASIC ($\alpha=1$) $\Delta E_v^{\text{ASIC}_1}$, LSDA ($\Delta E_v^{\text{LSDA}}$), and ASIC with the optimal $\alpha=\alpha^*$ from Table I $\Delta E_v^{\text{ASIC}_{\alpha^*}}$ for a number of semiconductors. The lattice parameters used for the calculations are either the equilibrium LSDA or the experimental when available (in Å). The experimental values are from the literature (last column).

Solid	ΔE_v^{exp}	$\Delta E_v^{\text{ASIC}_1}$	ΔE_v^{LSDA}	$\Delta E_v^{\text{ASIC}_{\alpha^*}}$	Reference
LiCl	4–5	3.52	3.06	3.51	81
NaCl	1.7–4.5	2.11	2.06	2.11	81
KCl	2.3–4.3	1.09	1.21	1.09	81
MgO	3.3–6.7	5.16	4.83	5.06	82 and 83
CaO	0.9	2.72	2.89	2.82	83
SrO		2.21	2.53	2.39	
AlN	6.0	7.44	6.27	6.92	84
GaN	7.4	8.42	7.33	7.85	85
InN	6.0	6.66	6.01	6.34	86
ZnO	~5	5.66	4.77	5.54	87
ZnS	5.5	6.49	5.57	6.05	88
ZnSe	5.6	7.14	5.35	6.38	88

(CBM) is essentially zinc 4s. With a value of ~0.5 for the scaling parameter α, the ASIC band gap closely matches the experimental gap of $E_g=3.43$ eV, whereas the LDA band gap is very small (~0.85 eV). Some part of the LDA band gap error in ZnO can be traced to an underestimation of the position of the semicore Zn 3d states. The LDA binding energy for the Zn 3d states is ~5.5 eV while photoemission results place them at around ~7.8 eV. ASIC, however, rectifies the problem and is in very good agreement with experiment. This results furthermore in the removal of the spurious Zn$_{3d}$-O$_{2p}$ band mixing seen in LDA. An additional feature is that the bandwidth of the valence band increases considerably as an effect of the downshift of the d manifold. Its worth mentioning that the positions of the Zn 3d levels obtained from ASIC in the case of ZnS, ZnSe, and ZnTe also agree remarkably well with experiment.

The wide-gap III-V semiconductor GaN presents similar phenomenology to that of ZnO. Figure 3 compares the band structure for wurtzite GaN obtained from LSDA and ASIC. When compared to x-ray photoemission spectra,[89] the LSDA band structure of GaN has several shortcomings. First, the band gap between N 2p bands (VBT) and Ga 4s bands (CBM) is underestimated at around 2.2 eV against the experimental value of 3.4 eV. Second, the 3d states of Ga are too shallow in LSDA, leading to a spurious $3d-2s$ hybridization. As a result the Ga 3d states overlap with and split the N 2s bands. ASIC rectifies the picture on both counts by improving the band gap and lowering the position of the Ga 3d bands with respect to the N 2s bands. These results corroborate those of Refs. 1, 25, and 26 where in ZnO and GaN have previously been discussed in a pseudopotential based SIC context.

C. Transition-metal oxide: MnO

Transition-metal oxides like MnO and NiO are characterized by partially filled 3d orbitals and an associated local
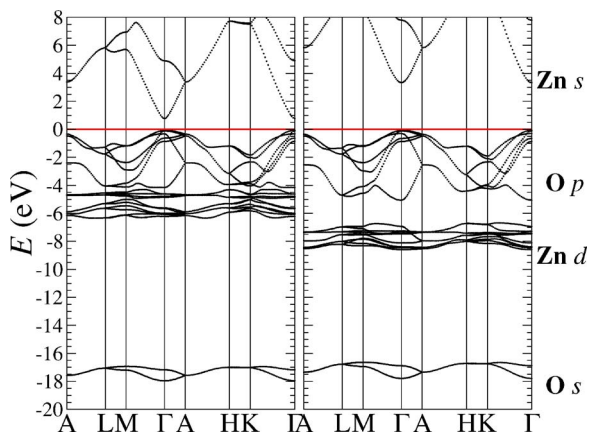

FIG. 2. (Color online) Calculated band structure of wurtzite ZnO obtained from LSDA (left) and ASIC (right). Owing to the ionic character of ZnO each group of bands can be clearly labeled according to a single, dominant orbital character as shown. The VBT is aligned at 0 eV.
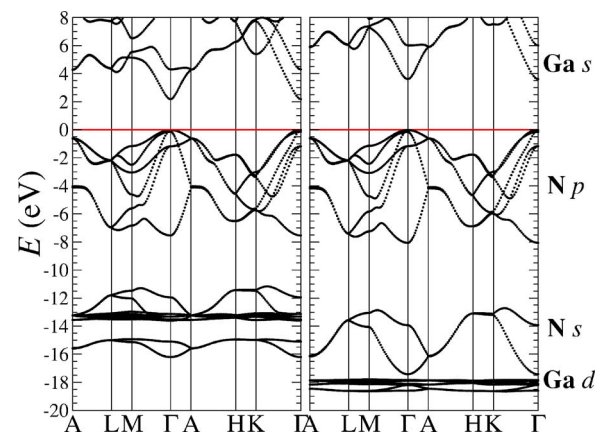

FIG. 3. (Color online) Calculated band structure of wurtzite GaN obtained from LDA (left) and ASIC (right). The primary orbital character of the bands is indicated. The VBT is aligned at 0 eV.
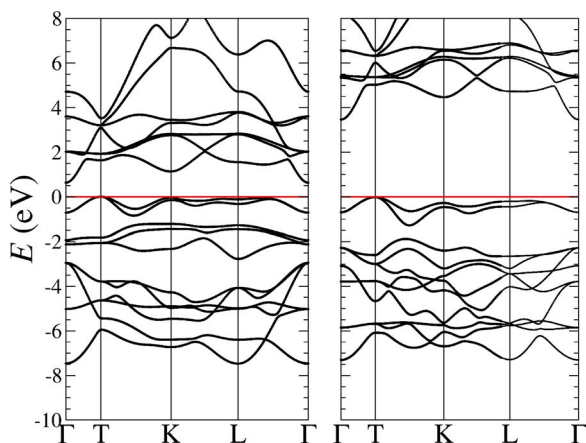
FIG. 4. (Color online) Calculated band structure of antiferromagnetic MnO obtained from LSDA (left) and ASIC (right). In our calculation we obtain an LSDA band gap of ~0.65 eV whereas the ASIC band gap is much improved at ~3.5 eV. The VBT is aligned at 0 eV.

magnetic structure. In particular the Mn^{2+} ions in MnO are magnetic with a half filled $3d$ shell. In the ground state, MnO is an A type antiferromagnetic insulator in the intermediate charge-transfer Mott-Hubbard regime with a band gap in the region of 3.8–4.2 eV. The VBT is expected to be of mixed Mn $3d$-O $2p$ character and the CBM pure Mn $3d$ in character. However, the LSDA description of MnO is flawed in several aspects, most notably in describing MnO as a narrow gap (E_g=0.92 eV) Mott-Hubbard insulator with both the VBT and CBM composed of purely of Mn $3d$ states. This is due to the severe underestimation of d electron binding energies in LSDA. The calculated antiferromagnetic band structures of MnO from LSDA and ASIC (α=1/2) are shown in Fig. 4.

Note that these are for the rhombohedral unit cell with four atoms per cell. The two Mn ions are antiferromagnetically aligned and the oxygen ions are nonmagnetic. This results in a layered ferromagnetic order of the (111) planes, which in turn are antiferromagnetically coupled to each other. Also in this case, ASIC is a considerable improvement over LSDA. The size of the fundamental gap now resembles the experimental one and the VBT recovers some p character. We point the reader once again to Ref. [1] wherein transition-metal oxides have been discussed in much more detail.

V. RESULTS: MOLECULES

A. Ionization potentials

In view of the fact that the ASIC method gives improved eigenvalue spectra for several solid-state systems, it is worth taking a cautious look at how it performs with molecules. This is particularly important for assessing whether the ASIC scheme can be adapted to work in DFT electron transport schemes based on the KS spectra.[23,90] In exact KS DFT only the highest occupied orbital eigenvalue (ϵ^{HOMO}) has a rigorous physical interpretation and corresponds to the negative of the first ionization potential.[19,20] More generally, for a N electron system, the following equations hold in exact KS-DFT:

$$\epsilon^{HOMO}(M) = -I_N \text{ for } (N-1 < M < N), \quad (41)$$

$$\epsilon^{HOMO}(M) = -A_N \text{ for } (N < M < N+1), \quad (42)$$

where $-I_N$ and $-A_N$ are the ionization potential (IP) and the electron affinity (EA), respectively. Therefore we start our analysis by looking at these quantities as calculated by ASIC. Also in this case we investigate different values of α. How-

TABLE III. Experimental ionization potential (IP) compared to calculated HOMO eigenvalues for neutral molecules. Columns 3 and 4 present the results from ASIC with, respectively, α=1/2 and α=1. The experimental data are taken from Ref. [91].

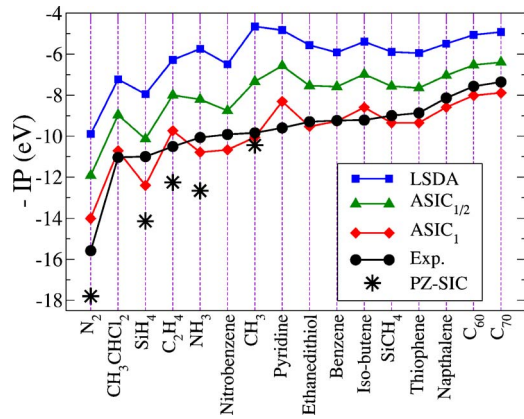| Molecule	ϵ^{HOMO} (eV)			−IP (eV)
	LSDA	ASIC$_{1/2}$	ASIC$_1$	Experiment
CH_3	−4.65	−7.34	−10.06	−9.84
NH_3	−5.74	−8.21	−10.79	−10.07
SiH_4	−7.95	−10.14	−12.41	−11.00
C_2H_4	−6.28	−8.00	−9.74	−10.51
$SiCH_4$	−5.89	−7.57	−9.35	−9.00
CH_3CHCl_2	−7.23	−8.97	−10.72	−11.04
C_4H_4S	−5.95	−7.65	−9.35	−8.87
$C_2H_6S_2$	−5.56	−7.54	−9.53	−9.30
Pyridine	−4.83	−6.57	−8.31	−9.60
Benzene	−5.92	−7.59	−9.28	−9.24
Isobutene	−5.39	−6.98	−8.6	−9.22
Nitrobenzene	−6.49	−8.76	−10.67	−9.92
Naphthalene	−5.49	−7.04	−8.59	−8.14
C_{60}	−5.06	−6.53	−8.02	−7.57
C_{70}	−4.92	−6.40	−7.89	−7.36

FIG. 5. (Color online) Experimental negative ionization potential IP compared to the calculated HOMO eigenvalues for molecules. The experimental data are from Ref. 91, while the star symbol represents full PZ-SIC calculations from Ref. 35.

ever, here we limit ourselves only to $\alpha = 1$ (ASIC$_1$) and $\alpha = 1/2$ (ASIC$_{1/2}$).

In Table III and Fig. 5 we compare the experimental negative IP for several molecules with the corresponding (ϵ^{HOMO}) obtained using LSDA and ASIC. It is clear that LSDA largely underestimates the removal energies in all the cases and that the values obtained from ASIC$_{1/2}$ are also consistently lower than the experimental value. However, as made evident by the figure the agreement between ASIC$_1$ and experiments is surprisingly good. In fact, the mean deviation $\delta(X)$ (X=LSDA, ASIC$_{1/2}$, ASIC$_1$) from experiment,

$$\delta(X) = \sqrt{\frac{\sum_{i=1}^{N} [\epsilon_X^{HOMO,i} + IP_{Expt}^i]^2}{N}},$$

is 3.56 eV for LSDA, 1.69 eV for ASIC$_{1/2}$, and only 0.58 eV for ASIC$_1$ (N runs over the molecules of Table III). It is

worth noting that we have excellent agreement over the whole range of molecules investigated going from N$_2$ to large fullerenes C$_{60}$ and C$_{70}$.

For comparison in Fig. 5 we have also included results obtained with a full self-consistent PZ-SIC approach.[35] Surprisingly our atomic approximation seems to produce a better agreement with experiments than the self-consistent scheme, which generally overcorrects the energy levels. This is a rather general feature of the PZ-SIC scheme and it is generally acknowledged that some re-scaling procedure is needed.[92,93]

B. Electron affinities

In Hartree Fock theory where Koopmans' theorem holds,[94] the lowest unoccupied molecular orbital (LUMO) energy (ϵ^{LUMO}), corresponds to the vertical EA of the N electron system, if one neglects electronic relaxation. No such interpretation exists for (ϵ^{LUMO}) in DFT and so the EA is not directly accessible from the ground-state spectrum of the N electron system. However, as Eq. (42) indicates, the EA is in principle accessible from the ground-state spectrum of the $N+1-f$ ($0 < f < 1$) electron system and asserts in particular that it must be relaxation free through noninteger occupation. Unfortunately, the LSDA/GGA approximate functionals usually perform rather poorly in this regard as the $N+1$ electron state is unbound with a positive eigenvalue. So one resorts instead to extracting electron affinities from more accurate total-energy differences,[95] or by extrapolating them from LSDA calculations for the N electron system.[96] This failing of approximate functionals has been traced in most part to the SI error and so SIC schemes are expected to be more successful in describing the $N+1$ electron spectrum.

In Table IV we compare HOMO energies (denoted as ϵ_{N+1}^{HOMO}) of several singly negatively charged molecules with the experimental electron affinities. We also report the LUMO energies for the corresponding neutral species
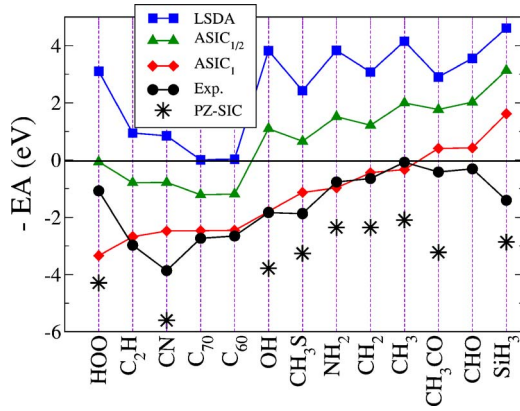
TABLE IV. Calculated HOMO eigenvalues for singly negatively charged molecules compared to experimental negative electron affinities ($-$EA). Columns 6, 7, and 8 present the LUMO eigenvalues for the corresponding neutral species.

Molecule	ϵ_{N+1}^{HOMO} (eV)				ϵ_N^{LUMO} (eV)		
	LSDA	ASIC$_{1/2}$	ASIC$_1$	Expt. $-$EA (eV)	LSDA	ASIC$_{1/2}$	ASIC$_1$
CN$^-$	0.84	−0.79	−2.48	−3.86	−8.13	−9.03	−9.42
C$_2$H$^-$	0.94	−0.80	−2.68	−2.97	−6.91	−7.38	−7.48
CH$_3$S$^-$	2.42	0.65	−1.14	−1.87	−5.20	−5.31	−5.34
OH$^-$	3.82	1.09	−1.80	−1.83	−0.16	−0.43	−0.69
SiH$_3^-$	4.61	3.13	1.61	−1.41	−2.66	−3.30	−4.07
HOO$^-$	3.10	−0.07	−3.34	−1.08	−5.30	−6.14	−6.40
NH$_2^-$	3.83	1.51	−0.98	−0.77	−5.27	−4.80	−4.39
CH$_2^-$	3.07	1.21	−0.45	−0.65	−3.80	−3.84	−3.91
CH$_3$CO$^-$	2.90	1.76	0.40	−0.42	−2.94	−3.88	−4.85
CHO$^-$	3.55	2.02	0.42	−0.31	−3.30	−4.40	−5.51
CH$_3^-$	4.15	1.99	−0.34	−0.08	−2.73	−2.59	−2.47
C$_{60}^-$	0.03	−1.19	−2.45	−2.65	−3.44	−4.66	−5.90
C$_{70}^-$	0.00	−1.22	−2.47	−2.73	−3.17	−4.41	−5.66

FIG. 6. (Color online) Experimental negative electron affinities ($-$EA) compared to calculated HOMO eigenvalues of negative radicals.

(denoted as ϵ_N^{LUMO}). LSDA relaxed geometries for the neutral molecule are used for both the neutral and charged cases. We find that various $\epsilon_{N+1}^{\text{HOMO}}$ obtained from ASIC$_1$ once again are in reasonably good agreement with corresponding experimental electron affinities while LSDA and ASIC$_{1/2}$ continue to be poor even in this regard. In this case $\delta(X)$ stands at 4.1, 2.31, and 0.83 eV for LSDA, ASIC$_{1/2}$, and ASIC$_1$, respectively. Notice that $\epsilon_{N+1}^{\text{HOMO}}$ from LSDA is positive in most cases as the states are unbound.

In Fig. 6 we present our data together with $\epsilon_{N+1}^{\text{HOMO}}$ as calculated using the PZ-SIC.[35] Again ASIC$_1$ performs better than PZ-SIC, that also for the EA systematically overcorrects.

C. Vertical excitations

Having shown that ASIC offers a good description of both IP and EA for a broad range of molecules, we turn our attention to the remaining vertical ionization potentials. As mentioned before, KS-DFT lacks of Koopmans theorem, and therefore the KS energies are not expected to be close to the negative of the vertical ionization potentials. However, at least for atoms, the introduction of SIC brings a remarkable cancellation between the negative relaxation energy and the positive non-Koopmans corrections.[16] For this reason the SIC KS eigenvalues are a good approximation to the relaxed excitation energies. As an example, in Table V we present the orbital energies calculated with ASIC$_1$ and ASIC$_{1/2}$ for the N$_2$ molecule. These are compared with experimental data[97] and orbital energies obtained, respectively, with

Hartree-Fock (HF), self-consistent SIC, and SIC where molecular orbitals are used instead of localized orbitals (D-SIC).[31]

Remarkably ASIC$_1$ seems to offer good agreement over the whole spectrum, improving considerably over LSDA and in some cases even over SIC and HF results. This improvement is not just quantitative, but also qualitative. For instance, while rectifying the LSDA spectrum of the N$_2$ molecule, ASIC$_1$ preserves the correct order between $3\sigma_g$ and $1\pi_u$ orbitals, which are erroneously inverted by both SIC and HF. So why does ASIC perform better than the other methods with regards to removal energies? In LSDA, electron relaxation typically cancels only half of the non-Koopmans contributions, resulting in energies that are too shallow.[16] In contrast HF lacks energy relaxation and the orbital energies are too deep. The reason why ASIC$_1$ performs better than self-consistent SICs is less clear. As a general consideration, also for the case of vertical ionization energies self-consistent SICs seem to overcorrect the actual values. Thus the SIC potential appears too deep, and the averaging procedure behind the ASIC approximation is likely to make it more shallow.

As a further test we calculated the orbital energies for a few other molecules and compared them both with LSDA and experiment.[98] These are presented in Table VI. Again the ASIC$_1$ results compare rather well with experiment, and we can conclude that the ASIC method offers a rather efficient and inexpensive theory for single-particle vertical excitations.

D. HOMO-LUMO gap and discontinuity of the exchange and correlation potential

We are now in a position to discuss the HOMO-LUMO gap in ASICs. As already mentioned, even for the exact XC functional, the KS gap $E_g^{\text{KS}} = \epsilon^{\text{LUMO}} - \epsilon^{\text{HOMO}}$ does not account for the actual quasiparticle gap $E_g = I_N - A_N$. This in turn is the sum of E_g^{KS} and the discontinuity of the exchange and correlation potential Δ_{xc}. Equivalently,

$$\Delta_{\text{xc}} = \lim_{f \to 0^+} \epsilon_{N+f}^{\text{HOMO}} - \epsilon_N^{\text{LUMO}}, \qquad (43)$$

i.e., Δ_{xc} is the discontinuity in the eigenvalue of the LUMO state at N. Therefore in order to extract the actual gap from the KS gap, provided that the spectrum is reasonably well described at integer electron numbers N, what remains is to model the derivative discontinuity at N and ensure that

TABLE V. Orbital energies of N$_2$ calculated with various methods. The results for Hartree-Fock and SIC are from Ref. 31. Experimental results are from Ref. 97.

Orbital	HF	SIC	D-SIC	ASIC$_1$	ASIC$_{1/2}$	LSDA	Expt.
$2\sigma_g$	-41.49	-38.86	-37.85	-38.29	-33.22	-28.16	
$2\sigma_u$	-21.09	-20.27	-16.44	-18.42	-15.64	-12.93	-18.75
$3\sigma_g$	-17.17	-17.39	-13.88	-14.01	-11.70	-9.90	-15.58
$1\pi_u$	-16.98	-16.33	-16.68	-15.97	-13.74	-11.54	-16.93

TABLE VI. Orbital energies for CO, HF, and H$_2$O calculated with LSDA and ASIC$_1$. The experimental results are from Ref. 98 and references therein.

Molecule	Orbital	LSDA	ASIC$_1$	Expt.
CO	5σ	−8.74	−12.85	−14.01
	1π	−11.54	−16.64	−16.91
	4σ	−13.97	−19.36	−19.72
HF	1π	−9.83	−16.96	−16.19
	3σ	−13.61	−19.68	−19.90
H$_2$O	$1b_1$	−7.32	−13.38	−12.62
	$3a_1$	−9.32	−14.66	−14.74
	$1b_2$	−13.33	−18.03	−18.55

$\epsilon_{N\pm f}^{HOMO}$ is relaxation free for $(0 < f < 1)$. Local and semilocal (LSDA/GGA) XC functionals lack such a discontinuity, while self-interaction corrections are able to restore it, at least in part. For instance the PZ-SIC scheme is successful in this regard.

In Fig. 7 we illustrate the ionization curve for the ethylene (C$_2$H$_4$) molecule as the occupation of the HOMO state is varied from 0 to 1 in going from the ionized C$_2$H$_4^+$ to the netural C$_2$H$_4$ configuration. It is seen that among the three schemes presented, only the PZ-SIC scheme approximately models the behaviour required by Eq. (41). The ASIC HOMO eigenvalue roughly agrees with the PZ-SIC eigenvalue at integer occupation but behaves linearly through non-integer values. Thus we find that the derivative discontinuity for the molecule is smoothed out in ASIC, which still connects continuously different integer occupations. This is one of the limitations of the atomic representation employed in ASICs.

In view of the foregoing discussion, the actual size of the HOMO-LUMO gap in ASICs becomes significant with a di-

rect bearing on the physics described. Ideally, we want $\epsilon^{LUMO}(N)$ (LUMO for the N-electron system) to be as close to $\epsilon^{HOMO}(N+1)$ so that the range of eigenvalue relaxation through fractional occupation numbers $M \in (N, N+1)$ is minimized. Looking at columns 6, 7, and 8 in Table IV, however, we see that for almost all the molecules, this is hardly the case. The agreement between $\epsilon^{LUMO}(N)$ and $-EA$ from experiment $[\simeq \epsilon^{HOMO}(N+1)]$ for ASIC$_1$ is quite poor implying a considerable energy range spanning fractional particle number. We still expect this energy range to be smaller for ASIC$_1$ than LSDA. It is also apparent from the Table IV that $\epsilon_{ASIC}^{LUMO}(N)$ usually differs from $\epsilon_{LSDA}^{LUMO}(N)$ and in fact by considerable magnitudes in some cases. Thus the ASIC$_1$ "correction" to the empty LUMO state does not vanish in contrast to the PZ-SIC scheme where, by definition, the empty eigenstates are SIC free.

Since the SIC operator v_{ASIC}^σ is constructed in an atomic-orbital representation, the correction to any KS eigenstate ψ_n^σ either filled or empty,

TABLE VII. HOMO-LUMO gap obtained from ASIC compared to the LSDA value. The values marked with $*$ correspond to unbound LUMO levels.

Molecule	$\epsilon^{LUMO} - \epsilon^{HOMO}$ (eV)		
	LSDA	ASIC$_{1/2}$	ASIC$_1$
CH$_3$	1.92	4.75	7.59
NH$_3$	7.1*	9.29*	11.61*
SiH$_4$	8.44*	9.68	10.94
C$_2$H$_4$	5.81	6.59	7.38
SiCH$_4$	6.19	7.07	8.06
CH$_3$CHCl$_2$	5.79	6.84	7.88
C$_4$H$_4$S	4.46	5.13	5.8
C$_2$H$_6$S$_2$	4.44	6.02	7.6
Pyridine	3.85	4.56	5.26
Benzene	5.22	5.9	6.59
Isobutene	4.88	5.56	6.26
Nitrobenzene	3.25	4.03	4.42
Naphthalene	3.36	3.83	4.29
C$_{60}$	1.62	1.87	2.12
C$_{70}$	1.75	1.99	2.23

$$\delta E_{\text{ASIC}}^{n\sigma} = \langle \psi_n^\sigma | v_{\text{ASIC}}^\sigma | \psi_n^\sigma \rangle, \tag{44}$$

is not necessarily zero unless ψ_n^σ only projects onto empty atomic orbitals. Also this correction to the LUMO with respect to the LSDA is negative in most cases, exceptions being NH_2 and CH_3 where it is desirably positive. Thus the fundamental HOMO-LUMO gap in ASICs is a combination of both the HOMO and LUMO corrections. Table VII shows how this combination works out in $ASIC_{1/2}$ and $ASIC_1$ when compared to LSDA. The molecular test set is the same as that in Table III.

We see in almost all cases the ASIC gap is systematically larger than the LSDA one. This is expected because the correction to the HOMO is usually much stronger than that to the LUMO. In general, ASIC is expected to work well for systems where the occupied and unoccupied KS eigenstates of the extended system have markedly different atomic orbital signatures being derived predominantly from filled and empty atomic orbitals, respectively. In such a case, the ASIC correction to the empty states would be nullified in being scaled by near-zero atomic-orbital populations. In some cases, provided phase factors combine suitably, the correction to the empty states can even be positive with respect to the same in LSDA.

E. Final remarks

Before we conclude, we discuss some general properties of the ASIC method which are relevant to any orbital dependent SIC implementation and also some possible pitfalls. As with other SIC schemes, ASIC is not invariant under unitary transformations of the orbitals used in constructing the SIC potential. Thus the ASIC correction is likely to change as the atomic orbitals used for projecting onto the KS eigenstates of the system are rotated or transformed otherwise. Unlike the Perdew-Zunger method, however, there can be no variational principle over all possible unitary transformations of the atomic orbitals because in the general case they do not represent the Hamiltonian of the system under consideration. This also implies that if the scheme is used with a system that is already well described by LSDA, the "correction" additional to the LSDA result does not necessarily vanish.

Simple metals and narrow gap systems are likely candidates for this scenario.

Furthermore, it is pertinent to mention that ASIC becomes ineffective if not counterproductive for materials with homonuclear bonding, in which valence and conduction bands have the same atomic orbital character. In this situation the ASIC potential will shift the bands in an almost identical way, without producing any quantitative changes, such as the opening up of the KS gap. Note that this is a pitfall of the ASIC approximation, which distinguishes occupied from empty states only through their projected atomic orbital occupation, but not of the SIC in general. Typical cases are those of Si and Ge. The KS gap in Si goes from 0.48 eV in LSDA to only 0.57 eV for $ASIC_{1/2}$, while Ge is a metal in both cases. In addition the LSDA calculated valence bandwidths of 12.2 eV for Si and 12.8 eV for Ge, in good agreement with experiments, are erroneously broadened to 14.3 and 14.8 eV, respectively.

VI. CONCLUSIONS

In conclusion, we have implemented the ASIC scheme proposed by Filipetti and Spaldin within the pseudopotential and localized orbital framework of the SIESTA code. We have then investigated a broad range of semiconductors and molecules, with the aim of providing a reasonable estimate for the scaling parameter α. We found that $\alpha=1$, which accounts for the full atomic SI, describes surprisingly well ionic semiconductors and molecules. In particular for molecules, both the IP and the EA can be obtained with good accuracy from the HOMO KS eigenvalues, respectively, for the neutral and singly charged molecule. This makes the ASIC scheme particularly suited for applications such as quantum transport, where the position of the HOMO level determines most of the $I-V$ curve.

In contrast III-V and II-VI semiconductors are better described by $\alpha=1/2$, which corrects the atomic SI for screening. This makes $ASIC_{1/2}$ an interesting effective band theory for semiconductors. The relation of the present scheme with the fully self-consistent SIC methods has been emphasized, and so has been that with $LDA+U$.
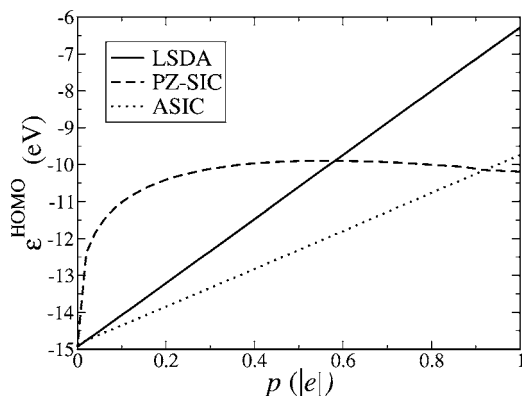
FIG. 7. Ionization curve for the ethylene (C_2H_4) molecule as the occupation of the HOMO state is varied from 0 to 1 in going from the ionized $C_2H_4^+$ to the netural C_2H_4 state.

[1] A. Filippetti and N. A. Spaldin, Phys. Rev. B **67**, 125109 (2003).

[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[3] E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).

[4] J. P. Perdew, A. Ruzsinsky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, J. Chem. Phys. **123**, 062201 (2005).

[5] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[6] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[7] P. Ordejón, Comput. Mater. Sci. **12**, 157 (1998).

[8] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).

[9] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys.: Condens. Matter **9**, 767 (1997).

[10] M. Cococcioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005).

[11] M. Wierzbowska, D. Sánchez-Portal, and S. Sanvito, Phys. Rev. B **70**, 235209 (2004).

[12] J. P. Perdew, Phys. Rev. Lett. **55**, 1665 (1985).

[13] J. P. Perdew and K. Schmidt, in *Density Functional Theory and Its Applications to Materials*, edited by V. Van Doren, C. Van Alsenoy, and P. Geerlings (AIP, Melville, NY, 2001).

[14] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[15] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).

[16] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[17] H. B. Shore, J. H. Rose, and E. Zaremba, Phys. Rev. B **15**, 2858 (1977).

[18] K. Terakura, T. Oguchi, A. R. Williams, and J. Kübler, Phys. Rev. B **30**, 4734 (1984).

[19] J. F. Janak, Phys. Rev. B **18**, 7165 (1978).

[20] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Jr., Phys. Rev. Lett. **49**, 1691 (1982); J. P. Perdew and M. Levy, *ibid.* **51**, 1884 (1983).

[21] J. P. Perdew and M. Levy, Phys. Rev. B **56**, 16021 (1997).

[22] J. P. Perdew, in *Density Functional Methods in Physics*, edited by R. M. Dreizler and J. da Providencia (Plenum, New York, 1985), p. 265.

[23] C. Toher, A. Filippetti, S. Sanvito, and K. Burke, Phys. Rev. Lett. **95**, 146402 (2005).

[24] E. Fermi and E. Amaldi, Accad. Ital. Rome **6**, 119 (1934).

[25] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **54**, 5495 (1996).

[26] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **55**, 12836 (1996).

[27] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **58**, 3865 (1998).

[28] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sanchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).

[29] M. Levy, Proc. Natl. Acad. Sci. U.S.A. **76**, 6062 (1979).

[30] J. G. Harrison, R. A. Heaton, and C. C. Lin, J. Phys. B **16**, 2079 (1983).

[31] M. R. Pederson, R. A. Heaton, and C. C. Lin, J. Chem. Phys. **80**, 1972 (1984).

[32] M. R. Pederson, R. A. Heaton, and C. C. Lin, J. Chem. Phys. **82**, 2688 (1985).

[33] S. Goedecker and C. J. Umrigar, Phys. Rev. A **55**, 1765 (1997).

[34] O. A. Vydrov and G. E. Scuseria, J. Chem. Phys. **121**, 8187 (2004).

[35] O. A. Vydrov and G. E. Scuseria, J. Chem. Phys. **122**, 184107 (2005).

[36] S. Kümmel and J. P. Perdew, Mol. Phys. **101**, 1363 (2003).

[37] J. B. Krieger, Y. Li, and G. J. Iafrate, Phys. Rev. A **45**, 101 (1992).

[38] J. Garza, J. A. Nichols, and D. A. Dixon, J. Chem. Phys. **112**, 7880 (2000).

[39] S. Patchkovskii, J. A. Nichols, and D. A. Dixon, J. Chem. Phys. **114**, 639 (2001).

[40] Y. H. Kim, M. Stadele, and R. M. Martin, Phys. Rev. A **60**, 3633 (1999).

[41] R. A. Heaton, J. G. Harrison, and C. C. Lin, Phys. Rev. B **28**, 5992 (1983).

[42] A. Svane and O. Gunnarsson, Phys. Rev. Lett. **65**, 1148 (1990).

[43] A. Svane, Phys. Rev. Lett. **68**, 1900 (1992).

[44] M. Arai and T. Fujiwara, Phys. Rev. B **51**, 1477 (1995).

[45] Z. Szotek, W. M. Temmerman, and H. Winter, Phys. Rev. B **47**, 4029 (1993).

[46] W. M. Temmerman, Z. Szotek, and H. Winter, Phys. Rev. B **47**, 1184 (1993); Z. Szotek, W. M. Temmerman, and H. Winter, Phys. Rev. Lett. **72**, 1244 (1994); A. Svane, *ibid.* **72**, 1248 (1994); Phys. Rev. B **53**, 4275 (1996); A. Svane, V. Kanchana, G. Vaitheeswaran, G. Santi, W. M. Temmerman, Z. Szotek, P. Strange, and L. Petit, *ibid.* **71**, 045119 (2005).

[47] L. Petit, T. C. Schulthess, A. Svane, Z. Szotek, W. M. Temmerman, and A. Janotti, Phys. Rev. B **73**, 045107 (2006).

[48] Z. Szotek, W. M. Temmerman, A. Svane, L. Petit, G. M. Stocks, and H. Winter, Phys. Rev. B **68**, 054415 (2003).

[49] L. Petit, A. Svane, W. M. Temmerman, and Z. Szotek, Phys. Rev. Lett. **88**, 216403 (2002); L. Petit, A. Svane, W. M. Temmerman, and Z. Szotek, Phys. Rev. B **63**, 165107 (2001).

[50] A. Zunger, Phys. Rev. B **22**, 649 (1980).

[51] M. M. Rieger and P. Vogl, Phys. Rev. B **52**, 16567 (1995).

[52] A. Filippetti and V. Fiorentini, Phys. Rev. Lett. **95**, 086405 (2005).

[53] A. Filippetti and V. Fiorentini, Phys. Rev. B **73**, 035128 (2005).

[54] A. Filippetti and N. A. Spaldin, Phys. Rev. B **68**, 045111 (2003).

[55] B. B. Van Aken, T. T. M. A. Palstra, A. Filippetti, and N. A. Spaldin, Nat. Mater. **3**, 164 (2003).

[56] P. Delugas, V. Fiorentini, and A. Filippetti, Phys. Rev. B **71**, 134302 (2005).

[57] A. Filippetti, N. A. Spaldin, and S. Sanvito, Chem. Phys. **309**, 59 (2005).

[58] A. Filippetti, N. A. Spaldin, and S. Sanvito, J. Magn. Magn. Mater. **290**, 1391 (2005).

[59] B. Baumeier, P. Krüger, and J. Pollmann, Phys. Rev. B **73**, 195205 (2006).

[60] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[61] M. R. Norman, Phys. Rev. B **28**, 3585 (1983).

[62] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[63] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[64] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[65] O. F. Sankey and D. J. Niklewski, Phys. Rev. B **40**, 3979 (1989).

[66] E. Artacho, D. Sanchez-Portal, P. Ordejón, A. García, and J. M. Soler, Phys. Status Solidi B **215**, 809 (1999).

[67] J. Junquera, O. Paz, D. Sanchez-Portal, and E. Artacho, Phys. Rev. B **64**, 235111 (2001).

[68] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).

[69] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983); Phys. Rev. B **32**, 3883 (1985).

[70] W. G. Aulbur, M. Städele, and A. Görling, Phys. Rev. B **62**, 7121 (2000).

[71] R. J. Magyar, A. Fleszar, and E. K. U. Gross, Phys. Rev. B **69**, 045111 (2004).

[72] G. Baldini and B. Bosacchi, Phys. Status Solidi **38**, 325 (1970).

[73] S. Nakai and T. Sagawa, J. Phys. Soc. Jpn. **26**, 1427 (1969).

[74] J. C. Phillips, Phys. Rev. **136**, A1705 (1964).

[75] R. C. Whited, C. J. Flaten, and W. C. Walker, Solid State Commun. **13**, 1903 (1973).

[76] Y. Kaneko and T. Koda, J. Cryst. Growth **86**, 72 (1988).

[77] H. Yamashita, K. Fukui, S. Misawa, and S. Yoshida, J. Appl. Phys. **50**, 896 (1979).

[78] V. Bougrov, M. E. Levinshtein, S. L. Rumyantsev, and A. Zubrilov, in *Properties of Advanced Semiconductor Materials GaN, AlN, InN, BN, SiC, SiGe*, edited by M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur (John Wiley & Sons, Inc., New York, 2001), pp. 1–30.

[79] B. R. Nag, Phys. Status Solidi B **237**, 2R1 (2003), and references therein.

[80] M. Oshikiri and F. Aryasetiawan, Phys. Rev. B **60**, 10754 (1999), and references therein.

[81] S. C. Erwin and C. C. Lin, J. Phys. C **21**, 4285 (1988), and references therein.

[82] M. L. Bortz, R. H. French, D. J. Jones, R. V. Kasowski, and F. S. Ohuchi, Phys. Scr. **41**, 537 (1990).

[83] E. A. Mikajloa, H. E. Dorsett, and M. J. Fordb, J. Chem. Phys. **120**, 10799 (2004).

[84] W. R. L. Lambrecht and B. Segall, in *Properties of Group II Nitrides* Ref. 1, p. 125. See this compilation, as well, for a comprehensive list of references up to 1994.

[85] S. A. Ding, S. R. Barman, K. Horn, and V. L. Alpherovich, in *Proceedings of ICPS-23*, edited by M. Scheffler and R. Zimmermann (World Scientific, Singapore, 1996), p. 525.

[86] L. F. J. Piper, T. D. Veal, P. H. Jefferson, C. F. McConville, F. Fuchs, J. Furthmüller, F. Bechstedt, H. Lu, and W. J. Schaff, Phys. Rev. B **72**, 245319 (2005).

[87] K. Ozawa, K. Sawada, Y. Shirotori, K. Edamoto, and M. Nakatake, Phys. Rev. B **68**, 125417 (2003).

[88] L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, and D. A. Shirley, Phys. Rev. B **9**, 600 (1974).

[89] W. R. L. Lambrecht, B. Segall, S. Strite, G. Martin, A. Agarwal, H. Morkoc, and A. Rockett, Phys. Rev. B **50**, 14155 (1994).

[90] A. R. Rocha, V. M. Garcia-Suarez, S. Bailey, C. Lambert, J. Ferrer, and S. Sanvito, Phys. Rev. B **73**, 085414 (2006); Nat. Mater. **4**, 335 (2005).

[91] NIST Computational Chemistry Comparison and Benchmark DataBase, http://srdata. nist. gov/cccbdb/default. htm

[92] O. A. Vydrov, G. E. Scuseria, J. P. Perdew, A. Ruzsinszky, and G. I. Csonka, J. Chem. Phys. **124**, 094108 (2006).

[93] O. A. Vydrov and G. E. Scuseria, J. Chem. Phys. **124**, 191101 (2006).

[94] T. Koopmans, Physica (Amsterdam) **1**, 104 (1933).

[95] L. Hedin and A. Johansson, J. Phys. B **2**, 1336 (1969).

[96] A. Filippetti, Phys. Rev. A **57**, 914 (1998).

[97] P. Baltzer, M. Larsson, L. Karlsson, B. Wannberg, and M. Carlsson Göthe, Phys. Rev. A **46**, 5545 (1992).

[98] D. P. Chong, O. V. Gritsenko, and E. J. Baerends, J. Chem. Phys. **116**, 1760 (2002).