Evolution of the impurity band in a weakly doped, highly compensated semiconductor

Ashley M. Cook and Mona Berciu

Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada, V6T 1Z1

(Received 26 March 2012; revised manuscript received 11 June 2012; published 18 June 2012)

We study the evolution of the impurity band in a weakly doped semiconductor as a function of the concentration of dopants, x. We present disorder-averaged results for the density of states of a doped simple cubic lattice and compare them with the predictions of the coherent potential approximation (CPA). For randomly distributed impurities the agreement is good, although CPA misses some qualitative features. We find that if electron-electron interactions can be ignored, as is the case in the highly compensated limit, the impurity band is still a clearly distinct feature in the spectrum even for dopant concentrations as large as $x \sim 0.10$.

DOI: 10.1103/PhysRevB.85.235130 PACS number(s): 71.23.−k, 71.55.−i, 02.60.Cb

I. INTRODUCTION

Ever since Anderson's seminal paper,[1] the effects of disorder have been studied in a wide variety of systems. If the disorder is introduced through the on-site potentials ϵ_i, as is most often the case, the model is defined by their distribution of probability. The two main classes of models use (i) a continuous (usually uniform) distribution with some desired width (in this case, known as Anderson disorder, the on-site potential varies randomly from site to site), or (ii) a bimodal distribution, where with probability x the potential is $\epsilon_i = U$, and with probability $1 - x$ the potential is $\epsilon_i = 0$ (up to an overall trivial energy shift).

This latter model is used as a simple way to study weakly doped semiconductors, $x \ll 1$ being the concentration of dopants. They are all assumed to be identical, hence their identical effect on the on-site energies in their vicinity. To also assume that the impurity potential is local (as opposed to spread over several sites), that the hopping integrals are not affected, and that the potential of a cluster of impurities is simply equal to the sum of individual single-impurity potentials, are all convenient additional approximations. These can be relaxed, if needed, but it is believed that this simple model should provide at least a qualitatively adequate description of the properties of a weakly doped semiconductor, such as the evolution of its density of states (DOS) with doping.

A complete model of the doped semiconductor must also state the concentration of charge carriers (electrons or holes, depending on the nature of the dopant). In the following, *we assume that the material is highly compensated*; i.e., it contains a second type of defect that binds most of the charge carriers at energies far removed from those of interest. As a result, the charge carrier concentration is much smaller than x and the electron-electron interactions can be safely ignored (we briefly discuss the consequences of this approximation below).

With these approximations, the problem can be studied in the single-electron limit. If $x \to 0$, one can assume that there is a single impurity in the system and the problem can be solved exactly. If the impurity potential is sufficiently attractive, an impurity state appears below the conduction band (for simplicity, from now on we assume donor doping) and is spatially localized in the vicinity of the impurity. As x increases, due to overlap between different such impurity states, an impurity band (IB) forms. Its width increases with x, and eventually one expects it to merge with the conductance band lying above it. Simultaneously, the states in this IB are expected to become more and more extended, and ultimately to regain their bandlike character.

While this phenomenology is universally accepted, little is known about the actual details, such as at what concentration x_m do the two bands merge, and whether there is some range of dopings above x_m where there is still a distinctive low-energy feature reminiscent of the IB and whose states have impurity-like nature, or this disappears as soon as the merging occurs, etc. Such questions are relevant for many issues of current interest—one particularly famous example is the dilute magnetic semiconductor $Ga_{1-x}Mn_xAs$, where after over a decade of studies there is still an ongoing debate as to whether the holes mediating the magnetism in this material occupy valence-band-like states or impurity-band-like states (dopings of interest here are below 10%, although things are complicated by various material issues).[2,3]

The lack of such answers is due to the difficulty in solving this problem numerically, even in the noninteracting approximation. If one is interested in low dopings $x \sim 0.01$, and if one assumes that a sufficiently large sample has at least a few hundred impurities, then one needs to deal with systems with at least 10^4 or more sites. Moreover, disorder averaging is required. Since disorder fluctuations in such models can be substantial, one may need to average many hundreds, if not thousands, of disorder configurations. In contrast, accurate numerical results can be obtained for Anderson-type disorder using smaller samples—and indeed, this latter problem has been studied numerically in great detail.[4]

Of course, significant efforts have been focused on proposing accurate analytical approximations for various disorder-averaged quantities. These range from the very simple-minded virtual crystal approximation (VCA) to the much more sophisticated coherent potential approximation (CPA), with the latter believed to be the most accurate "simple" approximation (more complicated methods, including various cluster generalizations of CPA, are available as well). These approximations have been tested mostly against one another, or against primarily one-dimensional numerical simulations. (Good introductions to these topics are given in Refs. 5,6. A recent review is Ref. 7.) The issue of their accuracy and range of validity is, therefore, not fully settled.

In this work, we present an extensive set of numerical results for three-dimensional lattices with $x < 10\%$, which allow

us to start answering some of these questions. Comparisons with CPA are also made. These numerical results are possible due to a recently proposed method of finding lattice Green's functions,[8] briefly reviewed below.

While here we focus only on the evolution of the DOS with the impurity concentration x and the impurity potential U, there are many other issues to be addressed, such as the nature of these electronic states (whether they are Anderson localized or extended) and the position of the mobility edge(s) in the spectrum. We briefly mention some preliminary results on these issues at the very end; however, a full analysis is postponed for future work.

The paper is organized as follows: Section II discusses the numerical solution. Results are reported in Sec. III, and Sec. IV contains our conclusions. Brief notes on our numerical implementation of the CPA self-consistency loop are given in the Appendix.

II. MODEL AND NUMERICAL SOLUTION

As already stated, our Hamiltonian is

$$\mathcal{H} = \mathcal{H}_0 + V = -t \sum_{\langle i,j \rangle} (c_i^\dagger c_j + \text{H.c.}) + U \sum_i p_i c_i^\dagger c_i. \quad (1)$$

Here, c_i is the electron annihilation operator at site i. For simplicity, we assume a simple cubic lattice; more complicated cases, such as FCC or BCC lattices, can be treated similarly.[8] The hopping is assumed to be nearest neighbor for simplicity, although generalizations to longer range hopping are also possible with the same method.[9] The on-site potential created by an impurity is $U < 0$, and $p_i = 1$ if there is an impurity at site i, and zero otherwise. Of course, $\langle p_i \rangle = x$, where $\langle ... \rangle$ indicates an average over all disorder configurations. Hamiltonian (1) ignores electron-electron interactions, which is reasonable in the highly compensated limit where the electron concentration is much smaller than x. The spin is also ignored since it is a trivial degree of freedom.

The needed quantities are the Green's functions:

$$G(i,j,\omega) = \langle 0 | c_i \hat{G}(\omega) c_j^\dagger | 0 \rangle, \quad (2)$$

where $\hat{G}(\omega) = [\omega + i\eta - \mathcal{H}]^{-1}$ is the resolvent, with $\eta \to 0_+$. For simplicity, we set $\hbar = 1$.

After these Green's functions are calculated as described below for a given disorder realization (i.e., a specified set of values $\{p_i\}$), we can find the quantity of interest, namely the disorder-averaged total density of states:

$$\rho(\omega) = -\frac{1}{\pi} \text{Im} \langle G(i,i,\omega) \rangle \quad (3)$$

(the disorder average makes this quantity independent of the chosen site i). Before describing the numerical method we use to calculate $G(i,j,\omega)$, let us briefly review the CPA. Within this approximation, the disorder-averaged value of the diagonal Green's function is

$$\langle G(i,i,\omega) \rangle = g_0(\omega - \sigma(\omega)), \quad (4)$$

where $\sigma(\omega)$ is a complex quantity that can be roughly thought of as the disorder-averaged self-energy, and is obtained from the self-consistency condition:[6]

$$\sigma(\omega) = xU + [U - \sigma(\omega)] \sigma(\omega) g_0(\omega - \sigma(\omega)), \quad (5)$$

where

$$g_0(\omega) = G_0(i,i,\omega) = \frac{1}{N} \sum_{\mathbf{k}} \frac{1}{\omega + i\eta - \epsilon_{\mathbf{k}}} \quad (6)$$

is the diagonal element of the Green's function G_0 of the clean system ($U = 0$ limit). The second equality expresses this as a sum over the Brillouin zone (BZ), which turns into an integral in the thermodynamic limit when the number of sites $N \to \infty$, over the $\mathbf{k}$-space propagator of the clean system, which depends on the free-electron dispersion. For a simple cubic lattice, $\epsilon_{\mathbf{k}} = -2t \sum_{\alpha=1}^{3} \cos(k_\alpha a)$, where a is the lattice constant. Details on how we solve Eq. (5) are in the Appendix.

The traditional approach to finding $G(i,j,\omega)$ for a given disorder realization is to numerically diagonalize the corresponding Hamiltonian to obtain its eigenvalues and eigenfunctions $\mathcal{H}|n\rangle = E_n|n\rangle$, and to use a Lehmann representation to build up the needed propagators:

$$G(i,j,\omega) = \sum_n \frac{\langle 0|c_i|n\rangle \langle n|c_j^\dagger|0\rangle}{\omega + i\eta - E_n}.$$

As discussed, this approach is time-consuming because of the large size of the systems that need to be diagonalized. This is made worse by the need to disorder average.

An improved approach is to use Dyson's identity $\hat{G}(\omega) = \hat{G}_0(\omega) + \hat{G}_0(\omega) V \hat{G}(\omega)$, to write

$$G(i,j,\omega) = G_0(i,j,\omega) + U \sum_l p_l G_0(i,l,\omega) G(l,j,\omega). \quad (7)$$

Note that the sum on the right-hand side has contributions only from the impurity sites l. As a result, this system of linear equations can be solved in two steps. First, for any values of j and ω of interest, one finds $G(l,j,\omega)$ from Eq. (7), where l runs over all impurity sites (because of the small x, the resulting linear system of equations has a rather small size). Once these values are known, Eq. (7) gives $G(i,j,\omega)$ for any other site i.

As a simple example, for a single impurity, say at site 0, Eq. (7) is $G(i,j,\omega) = G_0(i,j,\omega) + U G_0(i,0,\omega) G(0,j,\omega)$. To find $G(0,j,\omega) = G_0(0,j,\omega)/[1 - U G_0(0,0,\omega)]$ is now trivial, as it requires us to solve a linear system with a single equation, for the impurity site. All $G(i,j,\omega) = G_0(i,j,\omega) + U G_0(i,0,\omega) G_0(0,j,\omega)/[1 - U G_0(0,0,\omega)]$ are then known. In particular, we see that an impurity level appears at an energy E_I corresponding to the new pole: $U g_0(E_I) = 1$. This equation has a solution outside the continuum, i.e., with $E_I < -6t$, only if $U < -3.96t$.

The case with more impurities is an immediate generalization. While solving a linear system with xN unknowns is numerically much more efficient than diagonalizing a matrix of dimension N, especially for small x values, the hidden difficulty in this approach is the need to know many Green's functions $G_0(i,j,\omega)$ for the clean system. Expressing these as Fourier transforms over the Brillouin zone, in analogy to the second half of Eq. (6), is not very useful since the integrand is highly oscillating for large distances $|i - j|$, and moreover has a line cut for energies in the clean particle spectrum,

$\omega \in [-6t,6t]$. Thus, numerical integration to find $G_0(i,j,\omega)$ is inefficient.

An efficient way to calculate such Green's functions [which applies to $G(i,j,\omega)$ just as well as to $G_0(i,j,\omega)$] was proposed in Ref. 8. We review only its salient points, and refer the interested reader there for more details. It uses the identity $(\omega + i\eta - \mathcal{H})\hat{G}(\omega) = 1$ to calculate the desired Green's functions. For any diagonal matrix element, this gives $(\omega + i\eta - Up_i)G(i,i,\omega) = -t\sum_{i_1} G(i,i_1,\omega)$. Here, i_1 is the set of nearest-neighbor sites of i. It is actually more convenient to think of them as the sites at a Manhattan distance $M = 1$ from the site i (for the simple cubic lattice, the Manhattan distance between two sites $i = (i_x,i_y,i_z), j = (j_x,j_y,j_z)$ is defined as $M = |i_x - j_x| + |i_y - j_y| + |i_z - j_z|$). Because of the nearest-neighbor hopping, the equation of motion for any Green's function $G(i,i_M,\omega)$, where i_M is at a Manhattan distance M from the original site i, is linked only to Green's functions $G(i,i_{M\pm1},\omega)$. In other words, the infinite set of equations of motion can be grouped into simple recurrence relations over the Manhattan distance. This is the first key observation.

The second key observation is that $G(i,i_M,\omega) \to 0$ as $M \to \infty$. This is obvious for disordered samples at energies ω where electronic states are localized. However, it is also true for extended states, including $G_0(i,i_M,\omega)$ at energies ω in the band. The reason for this is the artificial "lifetime" $1/\eta$, which is finite in any numerical calculation since we cannot set $\eta = 0$. Because of it, the electron has an exponentially decaying probability to travel arbitrarily far from its original location i. [Remember that the real-time $G(i,i_M,t)$ is the amplitude of probability for the electron to move from i to i_M within time t. If this is vanishingly small for a suitably large value of M, so are its Fourier transforms $G(i,i_M,\omega)$.]

As a result, the infinite set of recurrence equations can be made finite, by setting all $G(i,i_M,\omega) = 0$ for $M \geqslant M_c$, where the cutoff M_c needs to be adjusted so that $G(i,i,\omega)$ is insensitive to its further increase. (Details about our choices of η and M_c are provided when we discuss our results. We also note that we use a more suitable way to truncate these equations, which is detailed in Ref. 8.) The truncated set of equations can be solved in terms of continued fractions of matrices, as discussed in Refs. 8,9. Alternatively, it can also be treated as one very large but very sparse set of linear equations, and solved with specialized packages such as PARDISO.[10,11] Of course, one could also use this approach to store all needed values of $G_0(i,j,\omega)$, and then solve Eq. (7) for many impurity configurations. All these approaches are significantly more efficient than brute-force diagonalization and allow us to find $G(i,i,\omega)$ for many disorder realizations in a reasonable amount of time, using off-the-shelf desktops.

III. RESULTS

We begin with strongly attractive donors: $U/t = -6$; in this case the deep impurity level is well below the conduction band and it should be easy to follow the evolution of the impurity band with increasing x. Figure 1 shows results for values of x ranging from 0.01 to 0.11. The top panels show the disorder-averaged DOS $\rho(\omega)$, defined by Eq. (3). We have also calculated the disorder-averaged DOS at impurity sites $\rho_i(\omega)$, defined similarly to Eq. (3) but keeping only sites i which have an impurity; i.e., $p_i = 1$ in each contributing disorder realization. Similarly, we also calculated a disorder-averaged DOS at nonimpurity (host) sites $\rho_h(\omega)$, by averaging only over disorder realizations with $p_i = 0$. As expected,

$$\rho(\omega) = x\rho_i(\omega) + (1 - x)\rho_h(\omega).$$

For convenience, the top and bottom panels also show the DOS of the clean system, $\rho_0(\omega) = -\frac{1}{\pi}\mathrm{Im}g_0(\omega)$ (dashed line). The CPA approximation for $\rho(\omega)$ is shown in the upper panels (full line).

In terms of technical details, for this value of $\eta = 0.05t$, we find results already converged, within the error bars due to disorder averaging, for a rather modest cutoff $M_c = 20$. Since there are $1 + \frac{2}{3}M(2M^2 + 3M + 4)$ sites within a Manhattan distance M of site i, this cutoff implies that samples containing over 11 000 sites (centered around the site i of interest) have been sampled. The results in Fig. 1 are for 1500 disorder realizations, each producing its own $G(i,i,\omega)$ value. This already gives reasonably small error bars. [Of course, they are larger for $\rho_i(\omega)$ because only a fraction x of the configurations contribute to this average. However, the values of $\rho_i(\omega)$ are also significantly larger, so this is not a problem.]

As anticipated, for $x = 0.01$ we see a narrow impurity band centered around the energy of the isolated impurity level [in this case, $E_I \approx -7.1t$; see Fig. 2(a)], and well separated from the conduction band lying above it. The overwhelming contribution to the IB states comes from impurity sites, as expected at such small x: $\rho_i(\omega) \gg \rho_h(\omega)$ in the IB. There is a small contribution to IB from $\rho_h(\omega)$ as well, due to the fact that even though impurity levels are strongly localized for such deep levels, they do spread over a few sites. The contribution to the IB seen in the $\rho_h(\omega)$ is from host sites which happen to be nearest neighbor to an impurity site and therefore have a finite impurity LDOS. One may argue whether these sites should be grouped into the "host" or the "impurity" category; however, we will continue with our original definition. Also, as expected, only host sites contribute to the DOS in the conduction band; impurity sites have vanishing DOS here.

As x increases, the IB DOS broadens considerably. Its maximum height also increases fast with x for small x, but seems to saturate for $x > 0.05$. This is most clearly seen by looking at $\rho_h(\omega)$, although one must remember the factor of x when considering the contribution from impurity states to the total DOS.

The results in Fig. 1 also reveal a major surprise: Even though the IB broadens considerably so that for $x = 0.05$ its upper edge is already above $\omega = -6t$, where the lower band edge of the continuum is originally located, in reality the IB is not merged with the conduction band even at $x = 0.11$. The reason for this is that the lower band edge of the continuum moves monotonically to higher values with increasing x, and this proceeds almost as fast as the IB broadens, so the two features remain distinct. This "migration" to higher energies of the continuum is very clearly seen from the $\rho_h(\omega)$ plots. Ignoring small contributions at IB energies (again, from host sites located in the immediate vicinity of impurity sites), the main contribution is in the conduction band. Comparison with the DOS $\rho_0(\omega)$ of the clean system (dashed line) shows how

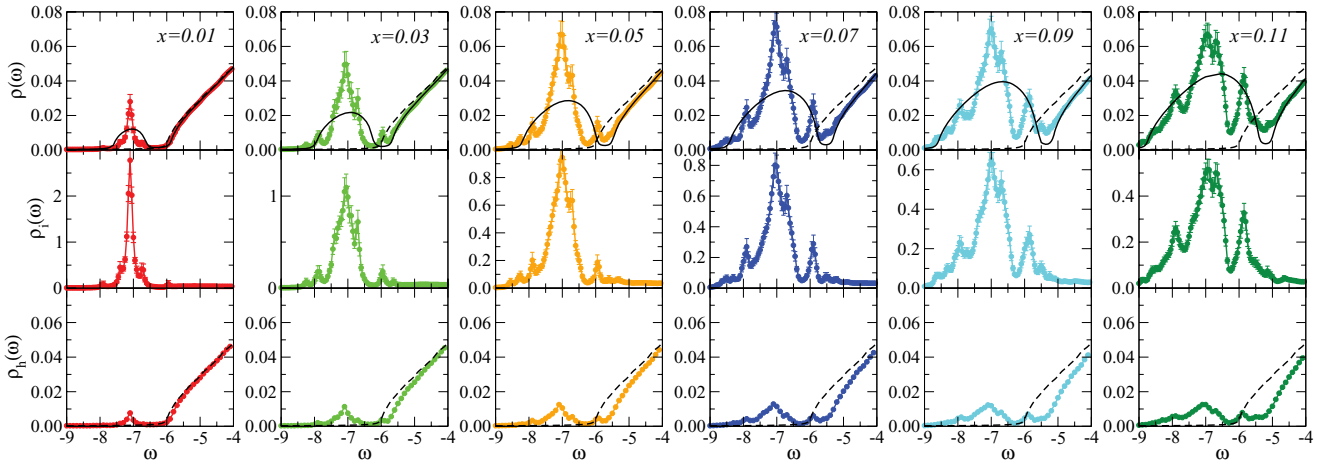

FIG. 1. (Color online) Disorder-averaged density of states $\rho(\omega)$ (upper panels), disorder-averaged density of states at impurity sites $\rho_i(\omega)$ (middle panels), and disorder-averaged density of states at host sites $\rho_h(\omega)$ (bottom panels) for various concentrations $x = 0.01$–0.11, for strongly attractive donors with $U = -6, t = 1$. The dashed lines show the DOS in the clean system. Full lines are the CPA predictions for $\rho(\omega)$. Where not shown, error bars are smaller than the size of the symbols. Other parameters are $\eta = 0.05$, $M_c = 20$ and each average is over 1500 disorder realizations. See text for more details.

the continuum is "eroded" and how its band edge moves to higher energies with increasing x.

In itself, this "erosion" of the conduction band is not unexpected; after all, the states in the IB are pulled out of the original conduction band states. Put another way, the DOS is normalized, $\int_{-\infty}^{\infty} d\omega\rho_0(\omega) = \int_{-\infty}^{\infty} d\omega\rho(\omega) = 1$, irrespective of the values of x and U. The low-energy DOS inside the IB must therefore come at the expense of missing DOS from higher energies. This fact has been used, for example, to explain magnetic circular dichroism in weakly doped (Ga,Mn)As.[3,12] The surprise is that the bottom of the conduction band is fully depleted almost at the same speed at which the IB spreads, so that the two features are still separated at $x = 0.11$.

This separation is clearly seen in the CPA results. These reproduce very well the higher energy DOS inside the band [coming primarily from $\rho_h(\omega)$] and its evolution with x. For the IB, on the other hand, CPA predicts a contribution with roughly the correct width and overall spectral weight, but the CPA DOS is a smooth broad peak whereas the exact results have lots of structure.

The peaks and valleys appearing in the IB DOS are not random "structure" that could be blamed on lack of sufficient disorder averaging. One can clearly see that the peaks appear at roughly the same energies in all the plots, and simply become more prominent and broader with increasing x. This suggests that they must be intrinsic features of the model. Their origin is easily understood as being related to the electronic structure of small impurity clusters. This is demonstrated in Fig. 2 where we show the LDOS measured at an impurity site if that impurity is isolated (dark full line), or part of a cluster of two nn (light full line), two nnn (dark dashed line), or three nn (light dashed line) impurities. For clusters of two nearby impurities, the degeneracy between their impurity levels is lifted and we see two new levels, corresponding to bonding and antibonding states. The split between the two levels decreases as the distance between the impurities increases, as expected. In particular, for a cluster of 2 nn impurities we see peaks around $-8t$ and $-6t$, which explain the appearance of prominent peaks in the finite x DOS at these energies as coming from such clusters. The peaks associated with clusters of 2 nnn impurities (especially the one at higher energies) are also seen at smaller x but they merge into the broader central peak as x increases and clusters with various relative distances are sampled. Peaks below $-8t$ cannot come from clusters of 2 impurities; instead they are associated with clusters of 3 or more impurities. Some detective work can uncover the origin of all these peaks.

Obviously, one-site CPA cannot describe such structure associated with clusters of impurities; this explains the smoothness of the CPA DOS in the impurity band. Cluster generalizations of CPA should, presumably, be able to remedy

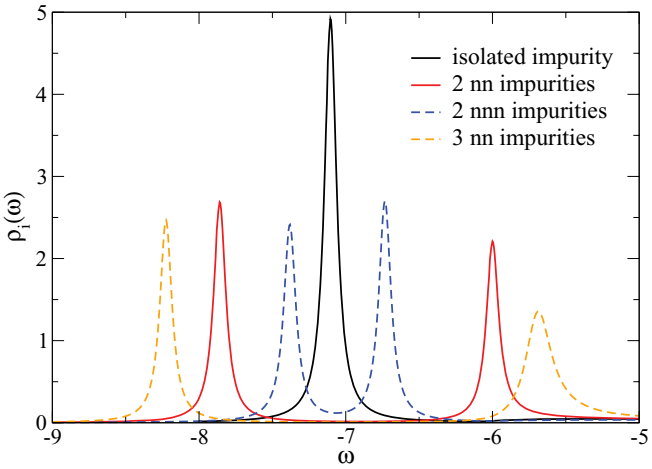

FIG. 2. (Color online) Local density of states for samples with a single impurity, with clusters of two nearest-neighbor (nn) or two next-nearest-neighbor (nnn) impurities, or a cluster of 3 nn impurities. In all cases the LDOS is measured at one of the impurity sites. Parameters are $U = -6, t = 1, \eta = 0.05$.

this situation. Apart from this, however, CPA captures quite nicely the evolution of the IB and neatly illustrates how it slowly grows and approaches the (remnants) of the conduction band, and that the two are still not merged at $x = 0.11$. We can generate results for larger x to see when the merging occurs, but for $x > 0.10$ we are no longer in the weak-doping regime and it is more and more questionable whether our simple model is appropriate to describe such systems, which are very likely to have large impurity clusters.

The absence of merging even for $x \sim 0.10$ could be simply due to the fact that we considered a case with very deep impurity levels. One would expect that for shallower impurity levels, where the IB forms much closer to the conduction band, this merging would occur at lower x. We test this expectation by generating data similar to that of Fig. 1 but for $U = -5t$ and $-4.5t$ (as discussed above, a bound impurity level only appears for $U < -3.96t$). We also decrease $\eta = 0.025$ to better resolve features, and accordingly increase $M_c = 25$ (over 22 000 sites are now included in the calculation).

The corresponding results are shown in Figs. 3 and 4. Overall, we see similar behavior with that of Fig. 1; however, there are some notable differences. Consider first the $x = 0.01$ results. As expected, as U becomes weaker and the impurity energy E_I moves closer to the band edge, so does the resulting IB. For the $U = -5$ case, one can still argue (with some help from the CPA results) that the IB and the conduction band are separated for $x = 0.01$. For the shallower level, for $U = -4.5$, they seem to already be merged. Another difference is the substantial contribution of $\rho_h(\omega)$ to the IB DOS, particularly in Fig. 4. This is due to the fact that wave functions of the shallower levels are much more spread out (the characteristic length scale which governs the exponential decay of these bound wave functions diverges as the binding energy vanishes). As such, many more "host" sites in the vicinity of an impurity have a finite LDOS at the impurity energy, and their signature is much more visible in $\rho_h(\omega)$ for all x, roughly mirroring (on a reduced scale) $\rho_i(\omega)$ at these energies.

With increasing x, in both cases the IB broadens and exhibits the characteristic peak patterns discussed for the $U = -6t$ case, associated with various small clusters. Apart from these, the agreement with CPA remains very acceptable. Based on these results, we can conclude that, for $U = -5t$, the IB merges with the continuum for $x \sim 0.02$, whereas for $U = -4.5t$ case, the IB is already merged with the conduction band at $x = 0.01$. However, after the merging occurs, for all $x < 0.10$ we can still easily identify a feature in the DOS which is clearly related to the IB and due to contributions primarily from the impurity sites and their immediate neighbors. Given the evolution trends in our results, we expect this to continue to be true at even higher x. There is no evidence that the overall DOS is becoming smooth and featureless at higher x; instead this rather distinct IB-related low-energy feature grows roughly linearly with x. However, it is probably better to have more realistic models to study larger x values, outside the weakly doped regime.

Taken at face value, these results suggest that in any weakly doped semiconductor ($x < 0.10$) that is reasonably described by this simple model, the occupied low-energy states are impurity-band-like, whether an impurity band is explicitly separated from the conduction band or the two are merged. One important caveat to keep in mind is that this is valid for highly compensated samples, where electron-electron interactions can be ignored due to the small number of electrons available to populate these states. For weakly or even uncompensated systems, where the number of available electrons becomes comparable to the number of impurities, this approximation fails. In this limit one cannot ignore the screening that an electron trapped in the vicinity of one impurity provides for that impurity's potential, insofar as all other electrons are concerned.

How to properly treat both disorder and electron-electron interactions is, of course, a major challenge. If treating interactions within the Hartree-Fock approximation is reasonable, these techniques based on calculating single-electron Green's functions could be used to find the spectra for any mean-field potential profile, and then the resulting DOS to calculate the new expectation values for the mean-field potential, to complete the self-consistency loop. If interactions are strong, then Hartree-Fock fails. (In most real materials an impurity does not bind two electrons, or if it does, this state has a very weak binding energy. Electron-electron repulsion is, therefore,

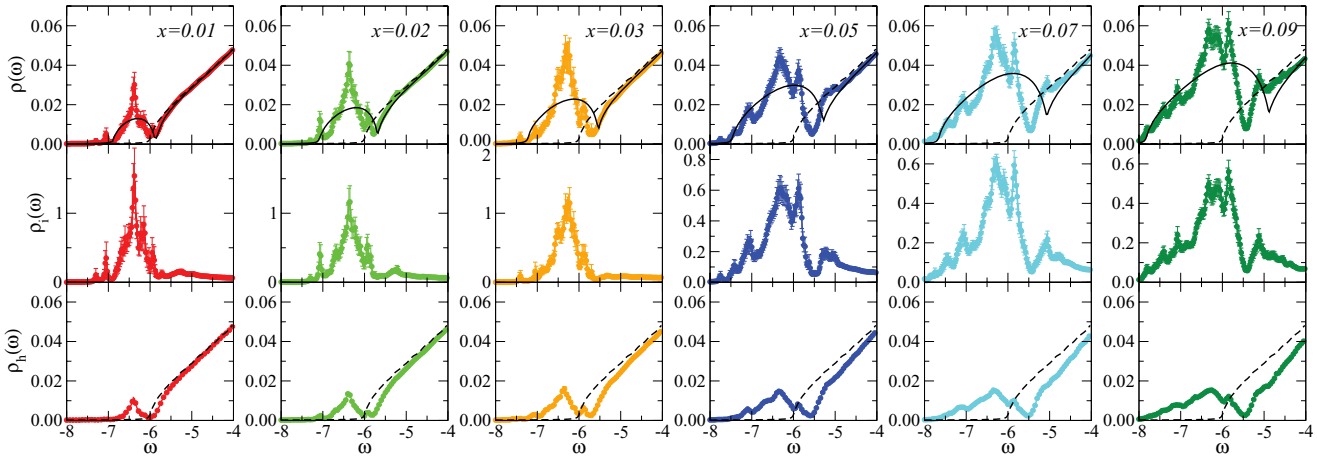

FIG. 3. (Color online) Same as in Fig. 1, for $t = 1, U = -5, \eta = 0.025$, $M_c = 25$ and averages over 2000 disorder realizations.

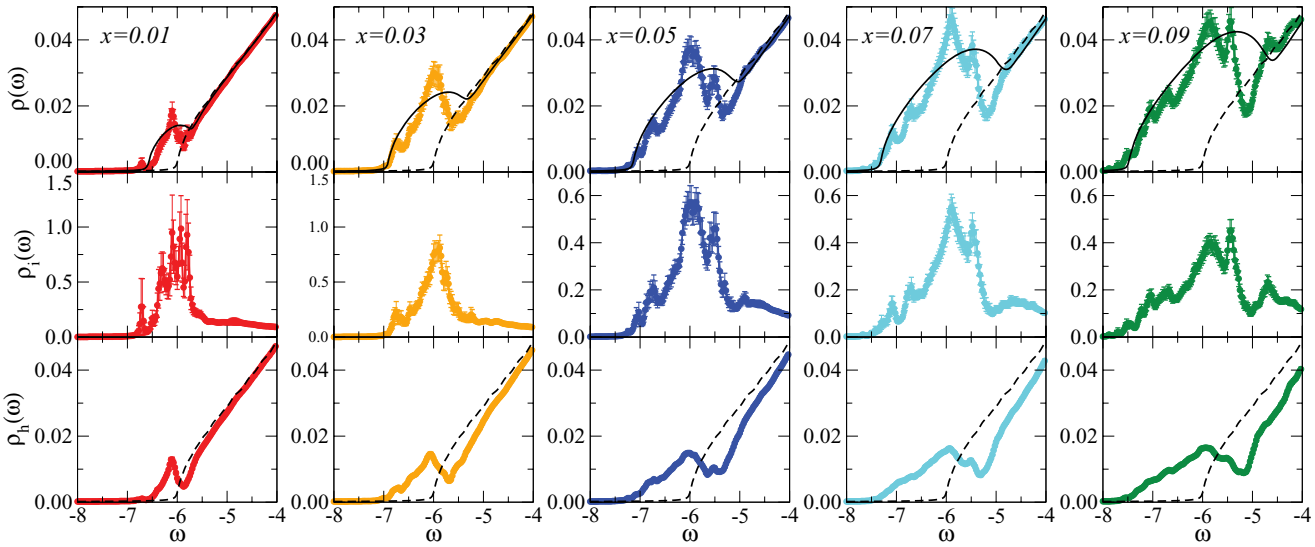

FIG. 4. (Color online) Same as in Fig. 1, for $t = 1, U = -4.5, \eta = 0.025$, $M_c = 25$ and averages over 2000 disorder realizations.

on a scale comparable or larger than the binding energy of a single electron.) Exact results can be obtained in this case using various numerical methods such as quantum Monte Carlo (for examples, see Refs. 13,14), however usually for very small systems. In any event, proper inclusion of the screening effects, which is necessary if the sample is not highly compensated, may significantly change the results.

A second caveat is linked to the fact that all these results assume no correlations whatsoever between the location of impurities; i.e., all disorder realizations are equally likely. The other extreme limit would be to place the impurities on a fully ordered superlattice inside the host semiconductor. This is the noninteracting, lattice analog of Mott's problem of a lattice of hydrogen atoms as a simplified model to study the metal-insulator transition in a doped system.[15,16] For commensurate values of x, this is easily done. In particular, we focus on concentrations $x = 1/n^3$ for n an integer, where we can order the impurities on cubic superlattices of constant na. For $n \geqslant 3$, this gives $x \leqslant 1/3^3 = 0.037$, in the weakly doped regime of interest.

In Fig. 5 we show the evolution of the impurity DOS $\rho_i(\omega)$ with the cutoff M_c, for such a superlattice. We see a large IB, separated through a significant gap from the next feature in the spectrum, even though in the fully disordered case, the IB is already merged with the conduction band for these parameters.

Since the impurities are now perfectly ordered, the eigenfunctions must be Bloch states, however in the 27 times smaller Brillouin zone associated with this large supercell. Because of the considerable folding of the BZ, we expect the original band to split into many subbands. Indeed, this is what we see (only the lower 2 such subbands are shown in Fig. 5). Because all eigenstates are now extended, larger values of M_c may be needed before convergence is reached. Indeed, this is demonstrated by Fig. 5. The IB is the feature most sensitive to the value of M_c; the other subbands are already well converged even for $M_c \sim 25$. This is not surprising since M_c practically defines the size of the system, and therefore controls the finite-size-like fluctuations of the IB DOS. We see that the width of the IB is well reproduced even for $M_c = 25$, and that, as M_c increases, the DOS closer to the band edges converges faster than that near the center of the band. This behavior is quite typical for this method.[8] While for $M_c = 50$–60 results at the center of the IB are still not fully converged, they are sufficiently representative that we can stop at such values of the cutoff.

Figure 6 shows the evolution of the local DOS at any impurity site, $\rho_i(\omega)$, with U and x. The left panels are for a fixed $x = 1/3^3$ and varying U. As U becomes more negative, the IB moves to lower energies and becomes narrower. This is expected. The IB is centered roughly at the single-impurity energy E_I, which moves down as U becomes more negative. For an ordered superlattice, the IB bandwidth is proportional to the effective hopping between nn impurity levels. This varies roughly like $\exp(-R/a_B)$,[17] where R is the distance between neighboring impurities (here kept constant) and a_B is the analog of the Bohr radius for the isolated impurity wave

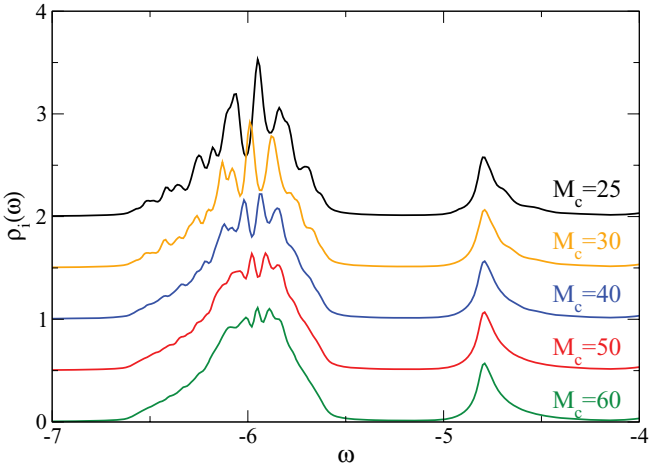

FIG. 5. (Color online) Local density of states at an impurity site $\rho_i(\omega)$ for an ordered cubic superlattice of impurities, for various cutoffs M_c. Parameters are $x = 1/27, U = -4.5, t = 1, \eta = 0.025$. Curves are shifted to ease the comparison.

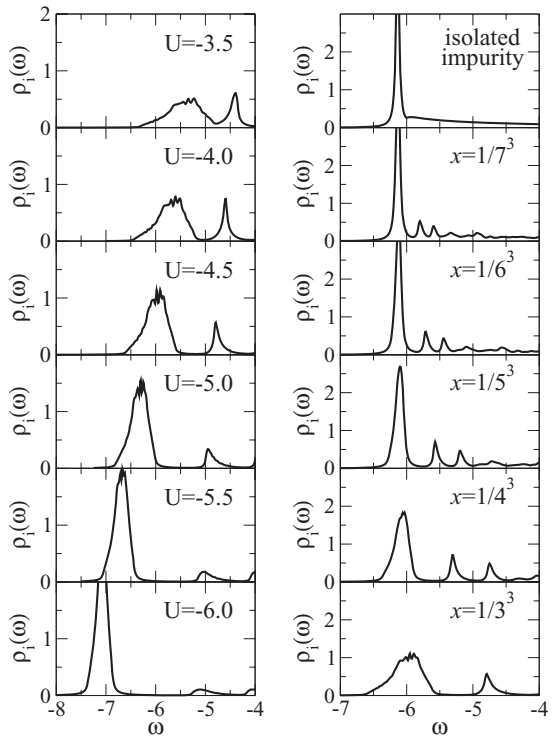

FIG. 6. Local density of states at an impurity site, $\rho_i(\omega)$, for an ordered cubic superlattice of impurities. Left panels: $x = 1/3^3$ and various U. Right panels: $U = -4.5$ and various x. Other parameters are $t = 1, \eta = 0.025, M_c = 50$.

function. Since a_B decreases as U becomes more negative and the impurity wave function is more localized, the narrowing of the IB is expected. The upper subbands also move towards lower energies as U becomes more negative, however much more slowly. In the lower two panels, one can see the onset of the third subband.

The interesting observation here is that the IB is still distinct from the upper subbands (i.e., no merging has yet occurred) even for $U = -3.5t$. At first sight this is rather surprising since impurity potentials with $U > -3.96t$ are too weak to bound a single-impurity level, so the existence of an IB at these values is not *a priori* expected. In the superlattice framework, however, the gap between the IB and the next subband depends only on the hybridization between states at the folded BZ surface, controlled by U. There is no reason for this to change discontinuously with U, and indeed the evolution of $\rho_i(\omega)$ is smooth through the $U_c = -3.96$ value.

The right panels show $\rho_i(\omega)$ for a fixed $U = -4.5$ and varying superlattice constants. The top panel is for an isolated impurity (equivalent to a superlattice $x = 1/n^3$ with $n \to \infty$). It shows the impurity level at an energy E_I just below the continuum band edge at $\omega = -6$. The two features are not fully separated because of the finite value of η, although the onset of the continuum is quite clearly visible. As n decreases, the IB stays roughly at the same energy E_I, as expected, and becomes broader, because the effective nn hopping increases as Ra decreases (here a_B is kept constant). We can also see the evolution of the higher subbands. For a given n, the BZ is folded down n^3 times and one expects up to n^3 subbands to replace the original band (there can be fewer subbands since full gaps do not necessarily open up at each crossing of the folded BZ surface). This expected increase in the number of subbands with increasing n is apparent. For $n = 6,7$, only a few lower subbands can be resolved; the upper ones merge into a continuum (again, one must also remember the finite broadening η). As $n \to \infty$, the number of subbands diverges but the gaps between them become extremely small so that all of them, except the IB, merge into the expected conduction band.

These results show that whether the IB is distinct from or is merged with the conduction band depends not only on U and x, but also on the degree of disorder of the impurities. If the impurities are perfectly ordered, the existence of an IB separated through a large gap from the higher features is much more likely than in a fully disordered case.

Of course, we can also consider intermediate levels of disorder, which should interpolate between these two extreme cases. One way to achieve this is to allow each impurity to be distributed with some probability around its superlattice location. For instance, assume that it is equally likely for any impurity to be at its superlattice site or any of its nn locations; such a situation allows for some degree of disorder, while still maintaining a rather uniform distribution of the impurities, with roughly one per superlattice unit cell. Then, one can systematically increase the region where the impurity is allowed to be, thus increasing the amount of disorder in the system.

Results for such intermediate disorder configurations are shown in Fig. 7, for cutoffs of 1 and 2, respectively, in the Manhattan distance at which an impurity can be located with respect to its superlattice sites (all allowed sites are equally probable). When compared with the perfect superlattice case (full line), we see the IB first broaden and then narrow somewhat as the disorder increases, although additional lower energy peaks associated with clusters appear in the latter case. This type of behavior has been observed for other values

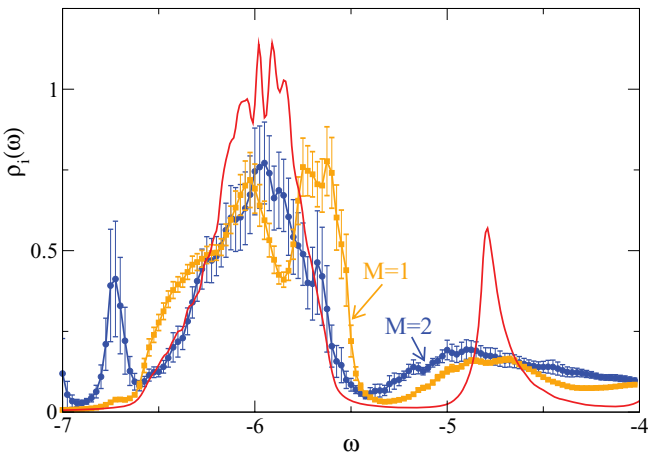

FIG. 7. (Color online) Local density of states at an impurity site, $\rho_i(\omega)$, when impurities are equally likely to be within a Manhattan distance 1 (light squares) or 2 (dark circles) of their superlattice locations. For these points, $M_c = 40$, and we averaged over 500 disorder realizations. For comparison, the full line shows the superlattice LDOS. Parameters are $t = 1, U = -4.5, \eta = 0.025, x = 1/3^3$.

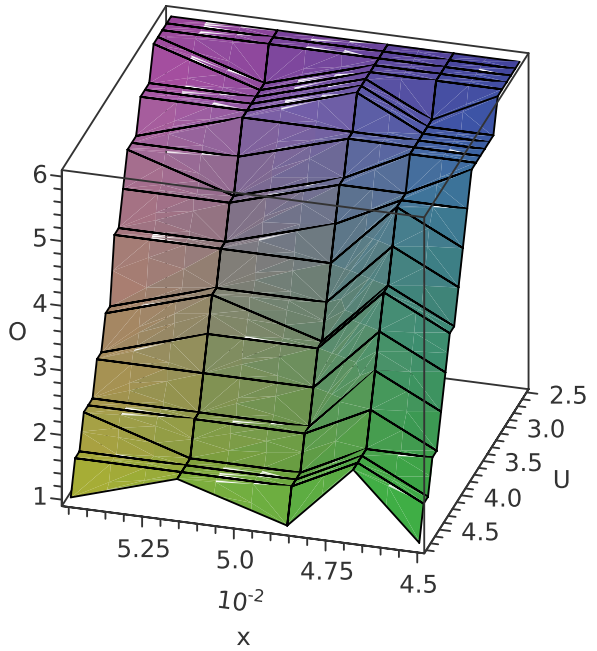

FIG. 8. (Color online) Estimated cutoff O for the minimum distance allowed between two impurities for which the IB merges with the continuum, vs x and $|U|$.

of U as well (not shown). The disorder has a larger effect on the higher energy features. The higher subband becomes very broad as soon as disorder is allowed, and its lower edge approaches the IB with increasing disorder. We estimate that for $M = 2$, the IB is already very close to merging with the continuum located above it.

Another, somewhat related way to vary the amount of disorder is to choose random positions for impurities but subject to the constraint that the Manhattan distance between any two impurities is equal or larger than a cutoff O. If $O = 1$, this allows any totally random disorder distribution, but as O increases the impurities are spread more homogeneously, however without an implied underlying superlattice structure. We have studied many averaged DOS for configurations with such disorder and estimated the value of O where the merging between the IB and the conduction band occurs. The results are shown in Fig. 8, as a function of x and $|U|$. Since whether merging has occurred or not is, to some extent, a subjective call, this data should be taken as pointing to the qualitative behavior with degree of disorder, and not so much as a definite quantitative criterion for merging.

The results in Fig. 8 show that for large $|U|$, where the IB forms well below the conduction band, one needs a large amount of disorder with $O \to 1$ before the considerable gap fills up. By contrast, as U decreases and the IB starts closer to (or even within) the conduction band, even for fairly homogeneous disorder configurations with a large minimum distance O between any two impurities, the merging has already occurred.

IV. CONCLUSIONS

To summarize, we have investigated here the formation of the impurity band in a weakly doped semiconductor, assuming a highly compensated system where electron-electron interactions can be neglected. This set of large-scale numerical simulations for a lattice model is made possible by an approach of dealing with lattice Green's functions. We find that if we consider completely random disorder configurations, the concentration x where the IB merges with the continuum above it depends on U: The more negative U, the deeper the impurity levels, the higher x must be before merging occurs. CPA describes quite well such situations, except for some structure associated with small clusters. For fixed x and U, the degree of disorder of the impurities controls whether the IB is a distinct feature or not. Generally, configurations with more homogeneously distributed impurities tend to have an IB separated from the conduction band. This can be understood in the extreme limit of an ordered superlattice, where one generically expects a gap to open between consecutive subbands because of the BZ folding.

One of the surprises revealed by this study is that for this model and with these approximations, a low-energy feature reminiscent of the IB is clearly visible for any weakly doping concentration $x < 0.10$, even if the merging has occurred at a much lower x value. The corresponding electronic states have primarily an impurity-like nature, and therefore one expects the behavior in such materials to be very much dominated by impurity-type physics, even if a fully separated IB no longer exists.

As already discussed, we completely ignore screening processes; this should be a reasonable approximation in highly compensated samples. However, for weakly compensated and uncompensated samples such processes cannot be ignored and the behavior of the system may be strongly affected by them. Our results may serve as a starting point to understand the precise role of such screening, by comparing them against results where screening processes are taken into consideration. Our results also serve as a benchmark for various approximations dealing with disorder, going beyond the CPA.

Finally, let us comment on the nature (localized or extended) of the states in the IB. In principle, our data can be used to investigate this, since we can easily make histograms of the LDOS values $\rho(i,\omega) = -\frac{1}{\pi}\mathrm{Im}G(i,i,\omega)$ (here we have only shown the corresponding averages and standard deviations). As generally expected, we find Gaussian-type distributions at energies high into the conduction band, typical of extended states. By contrast, at energies within the IB and for highly disordered samples, the distributions tend towards the log-normal distribution typical of localized states. However, a complete analysis seems to require size-dependent results (for a clear discussion of these issues, see Ref. 18 and references therein). While our cutoff M_c plays, to some extent, the role of fixing the system size, the link between the two is not so direct and more work is needed to settle this issue.

ACKNOWLEDGMENTS

We thank R. N. Bhatt for suggesting this problem, and H. Ebrahimnejad and H. Feshke for useful discussions. This work was supported by NSERC and CIfAR.

APPENDIX: NUMERICAL SOLUTION FOR CPA

CPA requires a self-consistent solution to Eq. (5), and the usual way to obtain it is by iterations. It is, however, not *a priori* obvious what is the best way to write Eq. (5) so that the iterations reach convergence most efficiently. We found three different formulations which work well in different energy ranges.

For energies below the IB, we find that if we start with the guess $\sigma(\omega) = xU$ and use it on the right-hand side of Eq. (5) to obtain the next iteration, etc., the process converges smoothly and reasonably fast to an acceptable self-consistent solution. Technically, we defined self-consistency to be reached when the absolute value of the difference between consecutive values of $\sigma(\omega)$ is below 0.01η (η is the small energy scale in this problem).

However, using this approach for higher energies either leads to an unphysical self-consistent solution (for example, one which gives unphysical negative DOS), or does not converge in a reasonable interval of time. We found that for energies within the IB, convergence to a physical self-consistent solution is fast if we rewrite Eq. (5) as

$$\sigma(\omega) = \frac{xU - [\sigma(\omega)]^2 g_0(\omega - \sigma(\omega))}{1 - U g_0(\omega - \sigma(\omega))}$$

and start with the guess $\sigma(\omega) = xU/[1 - U g_0(\omega - xU)]$.

Finally, for energies above the IB, fast convergence to a physical self-consistent solution was achieved if we rewrote Eq. (5) as

$$\sigma(\omega) = \frac{-1 + U g_0(\omega-\sigma) - \sqrt{[1 - U g_0(\omega-\sigma)]^2 + 4xU g_0(\omega-\sigma)}}{2 g_0(\omega-\sigma)}$$

[this comes from thinking of Eq. (5) as a quadratic equation in σ—ignoring the $g_0(\omega - \sigma)$ complication—and picking the root which vanishes at these energies when $x \to 0$]. The initial guess here was $\sigma = 0$.

The different solutions overlap in the common intervals where two of them work and combining all three of them results in a smooth function, giving us some confidence in this approach. The reasonable agreement with the exact results also suggests that we have likely found the correct CPA solutions.

In any event, one can certainly find self-consistent CPA solutions which are unphysical. Moreover, it is not *a priori* obvious that there is a unique self-consistent physical solution, although we only found one such solution in the cases we investigated. Thus, some care is needed when working with this approximation.

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).

[2] T. Dietl, Nat. Mater. **9**, 965 (2010).

[3] M. Dobrowolska, K. Tivakornsasithorn, X. Liu, J. K. Furdyna, M. Berciu, K. M. Yu, and W. Walukiewicz, Nat. Mater. **11**, 444 (2012).

[4] See, for example, P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985) or B. Kramer and A. MacKinnon, Rep. Prog. Phys. **56**, 1469 (1993).

[5] E. N. Economou, *Green's Functions in Quantum Physics*, 3rd edition (Springer-Verlag, New York, 2006).

[6] A. Gonis, *Green Functions for Ordered and Disordered Systems* (Elsevier, Amsterdam, 1992).

[7] D. A. Rowlands, Rep. Prog. Phys. **72**, 086501 (2009).

[8] M. Berciu and A. M. Cook, Europhys. Lett. **92**, 40003 (2010).

[9] M. Moeller, A. Mukherjee, C. P. J. Adolphs, D. J. J. Marchand, and M. Berciu, J. Phys. A: Math. Theor. **45**, 115206 (2012).

[10] O. Schenk and K. Gärtner, Journal of Future Generation Computer Systems **20**, 475 (2004).

[11] O. Schenk and K. Gärtner, Electron. Trans. Numer. Anal. **23**, 158 (2006).

[12] M. Berciu, R. Chakarvorty, Y. Y. Zhou, M. T. Alam, K. Traudt, R. Jakiela, A. Barcz, T. Wojtowicz, X. Liu, J. K. Furdyna, and M. Dobrowolska, Phys. Rev. Lett. **102**, 247202 (2009).

[13] S. Chiesa, P. B. Chakraborty, W. E. Pickett, and R. T. Scalettar, Phys. Rev. Lett. **101**, 086401 (2008).

[14] H.-Y. Chen, R. Wortis, and W. A. Atkinson, Phys. Rev. B **84**, 045113 (2011).

[15] N. F. Mott, Proc. Phys. Soc. London, Sect. A **62**, 416 (1949); Can. J. Phys. **34**, 1356 (1956); Philos. Mag. **6**, 287 (1964).

[16] J. H. Rose, H. B. Shore, and L. M. Sander, Phys. Rev. B **21**, 3037 (1980).

[17] R. N. Bhatt, Phys. Rev. B **24**, 3630 (1981); **26**, 1082 (1982).

[18] G. Schubert, J. Schleede, K. Byczuk, H. Fehske, and D. Vollhardt, Phys. Rev. B **81**, 155106 (2010).