Resolution of the exponent puzzle for the Anderson transition in doped semiconductors

Eduardo G. Carnio,* Nicholas D. M. Hine, and Rudolf A. Römer

Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

(Received 4 October 2017; revised manuscript received 1 December 2018; published 6 February 2019)

The Anderson metal-insulator transition (MIT) is central to our understanding of the quantum mechanical nature of disordered materials. Despite extensive efforts by theory and experiment, there is still no agreement on the value of the critical exponent ν describing the universality of the transition—the so-called “exponent puzzle.” In this Rapid Communication, going beyond the standard Anderson model, we employ ab initio methods to study the MIT in a realistic model of a doped semiconductor. We use linear-scaling density functional theory to simulate prototypes of sulfur-doped silicon (Si:S). From these we build larger tight-binding models close to the critical concentration of the MIT. When the dopant concentration is increased, an impurity band forms and eventually delocalizes. We characterize the MIT via multifractal finite-size scaling, obtaining the phase diagram and estimates of ν. Our results suggest an explanation of the long-standing exponent puzzle, which we link to the hybridization of conduction and impurity bands.

DOI: 10.1103/PhysRevB.99.081201

The Anderson metal-insulator transition (MIT) is the paradigmatic quantum phase transition, resulting from spatial localization of the electronic wave function due to increasing disorder [1]. As for any such transition, universal critical exponents capture its underlying fundamental symmetries. This universality allows one to disregard microscopic detail and the Anderson MIT is expected to share a single set of exponents. The last decade has witnessed many ground-breaking experiments designed to observe Anderson localization directly: with light [2–11], photonic crystals [9,12], ultrasound [13,14], matter waves [15], Bose-Einstein condensates [16], and ultracold matter [17,18]. The mobility edge [19], separating extended from localized states, was only measured for the first time in 2015 [20]. The hallmark of these experiments is the tunability of the experimental parameters and the ability to study systems where many-body interactions are absent or can be neglected. Under such controlled conditions, the observed exponential wave-function decay, the existence of mobility edges, and the critical properties of the transition [21,22] are in excellent agreement with the noninteracting Anderson model [1]. Furthermore, scaling at the transition [23] leads to high-precision estimates of the universal critical exponent ν from transport simulations [ν = 1.57(1.55, 1.59) [24]] and wave-function statistics [ν = 1.590(1.579, 1.602) [25]].

Anderson’s original challenge was to describe localization in doped semiconductors. For these ubiquitous materials, the existence of the MIT was confirmed indirectly by measuring the scaling of the conductance σ ∼ (n − nc)ν when increasing the dopant concentration n beyond its critical value nc. However, a puzzling discrepancy remains: A careful analysis by Itoh et al. [26] highlights that the value of ν can change significantly with the control of dopant concentration around the transition point, the homogeneity of the doping, and the purity of the sample. Following Stupp et al. [27], they suggest that the intrinsic behavior of an uncompensated semiconductor gives ν ≈ 0.5 [28], while any degree of compensation results in ν ≈ 1 [29]. Evidently, these values disagree with the aforementioned theoretical and experimental studies. The inability to characterize the Anderson transition in terms of a single, universal value for ν is known as the “exponent puzzle” [27,30].

Most theoretical models that have been applied to this problem lack the ability to capture the full complexity of a semiconductor. The Anderson model, for example, ignores the detail of the crystal lattice and the electronic structure, and also simplifies the physics by ignoring many-body interactions and interactions between the dopant and host material. These factors are known to change the universal behavior [31,32] and the value of ν, as shown in studies on correlated disorder [33–35] and hydrogenic impurities in an effective medium, where ν ≈ 1.3 [36,37]. Here, we propose a fundamental shift from studying localization using highly simplified tight-binding Anderson models, to atomistically correct ab initio simulations [38,39] of a doped semiconductor. We illustrate the power of our approach for sulfur-doped silicon, Si:S, where the MIT occurs for concentrations in the range 1.8–4.3 × 10^20 cm^−3 [40]. We model the donor distribution in Si:S by randomly placing the impurities in the lattice [41]. While we concentrate on Si:S here, our method is straightforwardly applicable to other types of impurities (Si:P; Si:As; Ge:Sh), hole doping (Si:B), and codoping (Si:PB; Ge:Ga,As).

With this approach we observe the formation of the impurity band (IB), upon increasing n, and its eventual merger with the conduction band (CB). States in the IB become delocalized, as measured directly via multifractal statistics of wave functions [42], and we observe and characterize the

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

Published by the American Physical Society
FIG. 1. Critical concentrations $n_c$ and exponents $\nu$ as a function of the energy $\epsilon$ from the Fermi level $\epsilon_F$, for $q = 0$ (red circles) and $q = 1$ (blue triangles). Solid and open symbols denote, respectively, the results for $\lambda = 1/4$ and $\lambda = 1/2$ coarse grainings. The error bars, shown only for $\lambda = 1/4$ and if larger than the symbol size, represent the 95% confidence level on the fit parameters. The error bars for $\lambda = 1/2$ are of the same order of magnitude as for $\lambda = 1/4$ and are omitted for clarity.

MIT. In Fig. 1 we plot how $n_c$ and $\nu$ vary for energies $\epsilon$ in the IB below the Fermi energy $\epsilon_F$. For $\epsilon \sim \epsilon_F$, the values are $\nu \sim 0.5$, while deeper in the IB the exponents increase to about $\nu \sim 1$, reaching values around 1.5. As we will show below, our simulations of an uncompensated semiconductor suggest that the reduction in $\nu$ at $\epsilon_F$ is due to the hybridization of IB and CB. Deep in the IB the physics of the Anderson transition reemerges with $\nu$ reaching the range of its proposed universal value [24,37,43]. Experiments can readily access these higher values by moving $\epsilon_F$ via compensation [26]—intentional or otherwise.

Density functional theory (DFT) calculations are now the leading method for ab initio solid state materials characterization [38] and discovery [39]. With the choice of Si:S, we can observe the transition in systems of up to $11 \times 11 \times 11$ unit cells, i.e., 10,648 atoms. These large system sizes can in principle be reached by linear-scaling DFT [44], but despite this, the necessity to average over many hundreds of disorder realizations makes repeated DFT calculations impractical for our purposes [45]. We therefore devise a hybrid approach: linear-scaling DFT calculations are performed, using the ONETEP code [46], on prototype systems of $8 \times 8 \times 8$ diamond-cubic unit cells (4096 atoms), employing geometry optimization to allow for the lattice to accommodate single or multiple S impurities. We include nine optimized nonorthogonal local orbitals (with radius 10 bohrs) for each site (in atomic Si, atomic orbitals are occupied up to level $3p$; for better convergence we additionally consider the five $3d$ orbitals) and use the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional [47] and a periodic cardinal sine (psinc) grid with an 800-eV plane-wave cutoff [48]. This gives an accuracy equivalent to plane-wave DFT for Si and other materials [49]. When embedded in silicon, sulfur, as the other chalcogens, acts as a deep donor. Such defects have highly localized potentials that are well described in a local orbital basis [50]. The impurity distribution is generated by randomly substituting the impurity atoms onto lattice sites. This follows the experimental techniques used to achieve high S concentrations, combining ion implantation with nanosecond pulsed-laser melting and rapid resolidification [40]. The impurities are effectively trapped in the substitutional sites [41].

The resulting Hamiltonians and overlap matrices, represented in terms of the nonorthogonal local orbital basis $\phi_a$, are used to construct three catalogs of local Hamiltonian blocks (cf. Fig. 2). The first catalog describes the Si host material, i.e., a set of on-site energies and hopping terms, starting at a central Si atom and extending to ten shells of Si neighbors. The second corresponds to the energies and hopping terms

FIG. 2. Schematic description of the workflow. (a) represents the catalog of prototypes. For clarity, we show a projection on the $xy$ plane and distances in units of $a$, the Si lattice parameter. The upper plot depicts one impurity (yellow with orange border) and the neighboring Si atoms (green); the lower plot denotes two impurities at distance $a$ and their Si neighbors (dark green). Gray sites indicate Si atoms unaffected by the impurity potential. In (b) we indicate how we build an effective tight-binding model of 4096 atoms with 29 impurities. The color code is the same as in (a) and indicates which catalog is used. Due to the projection on the $xy$ plane some impurities appear closer than they are. Their three-dimensional (3D) distribution is shown in (c) and (d), where we plot (c) a localized state deep in the IB and (d) an extended state above $\epsilon_F$. We represent the 90% largest wave-function values with spheres of volume proportional to $|\psi|^2$. Opacity and color are proportional to $-\log_L |\psi|^2$, with $L = 16$ here, so that lower (higher) values are in red transparent (violet solid). The box size is as in (b).
when the central atom is S, and the third cluster to pairs of neighboring S atoms. Here, we define a “neighbor” as being at most four shells apart. If two S atoms are five or more shells apart, each S atom is unaffected by the presence of the other [51]. For each system size $L$, concentration of impurities $n$, and disorder realization, we build the effective tight-binding Hamiltonians $H$ and overlap matrices $O$ from these catalogs (cf. Fig. 2) and solve the large generalized eigenvalue problem [52,53]

$$H \psi_j = \epsilon_j O \psi_j, \quad j = 1, \ldots, 9L^3$$

for eigenenergies $\epsilon_j$ and normalized eigenvectors $\psi_j = \sum_k M^j_k \phi_k$, written in a “site” basis by summing over the nine orbital coefficients of each site $k$, i.e., $|\psi_j(k)|^2 = \sum_{a \in k} M^j_a O_{ab} M^b_j$. In Fig. 2, we show examples of localized and extended states. For the $L^3$ = 4096 prototype, we have checked that our $\epsilon_j$'s agree within 0.1%–0.01% with the DFT energy levels. Due to the presence of $O$, and two orders of magnitude more hopping elements in $H$ compared to the Anderson model, we find that 10 648 atoms represent a practical upper limit (with tight-binding matrices of size $203 \times 255$). We average up to 1000 different disorder realizations for each $L$ and $n$ (cf. Table I).

Characterizing the IB and its density of states (DOS) is interesting for its spin and charge transport properties [54,55]. We compute the DOS of the IB from the $\epsilon_j$’s while changing the number of impurities $N_S$. We define $\epsilon_F$ as the midpoint between the highest occupied IB state at energy $\epsilon_{IB}$ and the lowest unoccupied CB state at $\epsilon_{CB}$. To obtain the average DOS for given $N_S$ and $L$, we shift the spectrum of each realization such that $\epsilon_F = 0$. The DOS shown in Fig. 3 is calculated by summing over Gaussian distributions of standard deviation $\sigma = 0.05 \text{meV}$ centered on $\epsilon_j - \epsilon_F$. We find that the IB has a peak at $\epsilon - \epsilon_F \sim -0.1 \text{eV}$ and a tail extending towards the VB with increasing $n$. This agrees with known features of the IB in doped semiconductors [56]. We emphasize that Si:S is particularly interesting for intermediate-band photovoltaic devices, where the efficiency increases when deep IB states can capture low-energy photons [54]. In order to avoid electron-photon recombination, the IB states should be delocalized such that they can contribute to the photocurrent. The determination of $n_c$ and the pronounced tail of the IB as presented in Fig. 3 therefore provide essential information for future device applications.

In the last decade, multifractal analysis [42,57,58] has become the method of choice to reliably and accurately extract the localization properties from wave functions [14,25,59]. In its essence, it describes the scaling of various moments of the spatial distribution of $|\psi_j|^2$, which is encoded in the singularity strengths $\alpha_q$. We coarse grain $|\psi|^2$ [60] by fixing a box size $l < L$ and partitioning the domain in $(L/l)^3 = \lambda^{-3}$ boxes. The amplitudes of the coarse-grained wave function $\mu$ are given by $\mu = \sum_{l \in \mathcal{B}} |\psi_l|^2$, i.e., by summing all $|\psi_l|^2$ pertaining to the same box $\mathcal{B}_l$. After rescaling the amplitudes as $\log \mu_l / \log \lambda$, we compute their arithmetic mean $\alpha_0 = \langle \log \mu_l \rangle / \log \lambda$ and weighted mean $\alpha_1 = \sum_l \mu_l \log \mu_l / \log \lambda$ (proportional to the von Neumann entropy [61]). Finally, for each $n$ and $L$ we take the ensemble average $\alpha_q(n,L)$, where $q = 0$ or 1.

At criticality, the universality class of the transition determines the scaling of $\alpha_q$ with $n$ and $L$. We capture this behavior using the well-established framework of finite-size scaling. Following Ref. [25], we assume that the data for each $L$ meet at the critical point $w = 0$ with a value $\alpha_q^{\text{crit}}$, and scale polynomially with $\rho L^{1/v}$, where $\rho(w) = w + \sum_{m>0} b_m w^m$ includes higher-order dependencies on the dimensionless concentration $w = (n - n_c)/n_c$. We hence fit the data against the function [62]

$$\alpha_q(n,L) = \alpha_q^{\text{crit}} + \sum_{i=1}^{m_L} a_i \rho L^{1/v},$$

with $n_c$, $v$, $\alpha_q^{\text{crit}}$, the $a_i$’s, and the $b_i$’s as fitted parameters, and $m_L$ and $m_p$ as expansion orders [51]. We illustrate the localization and scaling properties of the wave functions using the moments $\alpha_0$ and $\alpha_1$. Figure 1 shows the results of the fits as $\epsilon$ is varied, obtained from (2) (see Tables 1 and 2 in Ref. [51]). Crucially, we only accept estimates of $n_c$ and $v$ after consistently and rigorously checking their robustness against perturbations in $n$ and stability when increasing $m_L$ and $m_p$ [24,25,51].

Following this recipe, we identify the Anderson MIT and reconstruct the energy dependence of the mobility edge $n_c(\epsilon)$...
in Si. It exhibits (i) a maximum close to $\varepsilon_F$ and a decrease until $\varepsilon - \varepsilon_F \approx -0.09$ eV. (ii) For lower energies, $n_\text{c}$ increases again and the mobility edge moves towards the tail of the IB (cf. Fig. 3). These findings suggest a natural split into two different regimes, as also seen in the energy dependence of $\nu$. Values of $\nu$ in regime (i) increase continuously from $\nu \approx 0.5$ at $\varepsilon_F$ to about $\nu \approx 1$. In regime (ii), we find a larger spread of values $1 \lesssim \nu \lesssim 1.5$. This spread is consistent with the statistical uncertainty of each estimated $\nu$ in Fig. 1, which is dominated by the range of $L$ and the ensemble size $N$ (cf. Table I). However, the trend in $\nu$ observed in regime (i) requires a different explanation.

In Fig. 4, we present the distribution of states resolved in both energy $\varepsilon$ and $\alpha_0$. Perfectly extended states correspond to $\alpha_0 = 3$, while increasing localization results in $\alpha_0 \to \infty$. The data for $N_S = 40$ ($n = 4.9 \times 10^{20}$ cm$^{-3}$) show metallic states of the CB with $\alpha_0 \approx 3$ at $\varepsilon \approx \varepsilon_F$. The IB is characterized by (i) a majority region of states with $\alpha_0 \approx 3.4$ for $(\varepsilon - \varepsilon_F) \in [-0.25 \text{ eV}, -0.05 \text{ eV}]$ and (ii) a tail region of more localized $\alpha_0$ values ($\geq 3.4$) for $(\varepsilon - \varepsilon_F) \lesssim -0.25$ eV. However, for $N_S = 100$ ($n = 1.2 \times 10^{21}$ cm$^{-3}$) IB and CB have lost their identities. In fact, the two bands overlap in $\varepsilon$ as shown in Fig. 3, and also change their localization properties—the bands have hybridized, with $\alpha_0$ decreasing towards 3 (the metallic limit) close to $\varepsilon_F$. This observation is intriguing when tensioned against the simultaneous decrease in the value of $\nu$ at $\varepsilon \approx \varepsilon_F$ (cf. Fig. 1). Apparently, the localization of the IB states is substantially modified by the presence of the states from the CB. In Fig. 5, we show the $\alpha_0$ data for $N = 4096$ as a function of $\varepsilon$ and $n$. For small impurity concentrations, the IB consists of localized states with some of the largest values of $\alpha_0 \approx 3.6$, while the CB contains delocalized states with $\alpha_0 \gtrsim 3$. Upon increasing $n$, the IB develops and its states become more delocalized. Initially, this trend is most pronounced where the DOS of the IB is large (see Fig. 3), i.e., around $\varepsilon - \varepsilon_F \approx -0.12$ eV. Simultaneously, states at the top of the IB exhibit $\alpha_0$ values close to those denoting extended states in the CB, even before the band gap has fully closed. When reliable scaling is possible, we eventually see how the two mobility edges emerge. At the lower mobility edge, we find values of $\nu \sim 1$–1.5. At the upper mobility edge, we observe lower estimates for $\nu$ coinciding with lower $\alpha_0$ values at the transition due to the strong hybridization of IB and CB. Let us discuss how this observed hybridization and the resulting enhanced metallic behavior can affect the value of $\nu$. The leading scaling behavior from (2) is $\alpha_0^{\text{crit}} - \alpha_0 \sim w L^{1/\nu}$ for $w > 0$. A decrease in the effective $\alpha_0$ yields an increase in $\alpha_0^{\text{crit}} - \alpha_0$, which is consistent with a reduced exponent $\nu$ as observed in Fig. 1 for $(\varepsilon - \varepsilon_F) \gtrsim -0.1$ eV. An argument similar to the famous “gang of four” result [23] can be made directly for the transport experiments, where an increase in $\sigma \sim w^{\nu}$ for $0 \lesssim w \ll 1$, i.e., close to the critical point, is also consistent with a reduced $\nu$.

Let us reiterate our main point: Our simulations of the Anderson MIT in an uncompensated doped semiconductor find an effective $\nu \approx 0.5$ near $\varepsilon_F$, where the IB and the CB hybridize. Larger values of $\nu$, around 1–1.5, can be observed when the level of compensation is increased. These results provide a possible explanation for the observation of Itoh et al. [26] that in experiments a change from 0.5 to $\nu \approx 1$ can be induced by compensation. Taken together, compensation and band hybridization provide two important pieces to complete the “exponent puzzle”: Modeling the Anderson transition in doped semiconductors needs to include the CB (VB for hole-doped materials) together with the IB provided by the Anderson model—the experiments obviously include both and hence find $\nu$ values which, depending on their state of compensation, can be different from predictions based solely on the Anderson model of the IB.

How exactly the hybridization changes the effective value of $\nu$, as well as whether the value of $\nu$ deep in the IB is different from the noninteracting predictions, remain challenges for future high-precision studies. This includes
understanding how the hybridization of two bands, each with its own characteristic length scale, fits the hypotheses of
Chayes’ theorem [63].

Still, the approach we present here exploits and transfers the accuracy and versatility of modern \textit{ab initio} simulations to the study of Anderson localization in doped semiconductors—at a fraction of the computational cost. Beyond bulk semiconductors, other disordered systems [64], two-dimensional (2D) [65–67] and layered materials [68] are also well within reach, as is the investigation of the influence of many-body physics by, e.g., studying the interaction-enabled MIT in 2D Si:P [69,70]. We find that the critical concentration agrees quantitatively with a previous experiment in Si:S by Winkler et al. [40]. Our approach is hence capable of modeling fundamental physical phenomena while also making material-specific predictions.

The research data for this Rapid Communication are openly available [71].

\textit{Note added.} A first-principles study in Ti-doped Si using the dynamical cluster approximation was recently reported in Ref. [72].

We are grateful to Amnon Aharony, Siddharta Lal, David Quigley, and Alberto Rodriguez for discussions. We thank EPSRC for support via the ARCHER RAP project e420 and the MidPlus Regional HPC Centre (EP/K000128/1).
[47] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).

[48] C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez, and M. C. Payne, Phys. Rev. B 66, 035119 (2002).

[49] C.-K. Skylaris and P. D. Haynes, J. Chem. Phys. 127, 164712 (2007).

[50] P. Y. Yu and M. Cardona, Fundamentals of Semiconductors, Graduate Texts in Physics No. 11 (Springer, Berlin, 2010), p. 793.

[51] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.99.081201 for further details of the DFT simulations, hand gap closure, and scaling of the multifractal exponents, including full details of the finite-size scaling fits.

[52] O. Schenk, M. Bollhöfer, and R. A. Römer, SIAM J. Sci. Comput. 28, 963 (2006).

[53] M. Bollhöfer and Y. Notay, Comput. Phys. Commun. 177, 951 (2007).

[54] A. Luque and A. Martí, Phys. Rev. Lett. 78, 5014 (1997).

[55] T. Wellens and R. A. Jalabert, Phys. Rev. B 94, 144209 (2016).

[56] F. Cyrot-Lackmann and J. P. Gaspard, J. Phys. C 7, 1829 (1974).

[57] F. Milde, R. A. Römer, and M. Schreiber, Phys. Rev. B 55, 9463 (1997).

[58] F. Evers and A. D. Mirlin, Rev. Mod. Phys. 80, 1355 (2008).

[59] A. Richardella, P. Roushan, S. Mack, B. Zhou, D. A. Huse, D. D. Awschalom, and A. Yazdani, Science 327, 665 (2010).

[60] We drop the eigenstate index $j$ from here on.

[61] S. Chakravarty, Int. J. Mod. Phys. B 24, 1823 (2010).

[62] Note that we employ Wegner's relation [73] in assuming that in 3D the localization exponent equals the conductivity exponent.

Although strictly speaking a hypothesis for interacting systems, it is still widely used in this context.

[63] J. T. Chayes, L. Chayes, D. S. Fisher, and T. Spencer, Phys. Rev. Lett. 57, 2999 (1986).

[64] D. Chakraborty, R. Sensarma, and A. Ghosal, Phys. Rev. B 95, 014516 (2017).

[65] A. K. Geim and K. S. Novoselov, Nat. Mater. 6, 183 (2007).

[66] G. R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M. S. Strano, V. R. Cooper, L. Liang, S. G. Louie, E. Ringe, W. Zhou, S. S. Kim, R. R. Naik, B. G. Sumpter, H. Terrones, F. Xia, Y. Wang et al., ACS Nano 9, 11509 (2015).

[67] S. Das, J. A. Robinson, M. Dubey, H. Terrones, and M. Terrones, Annu. Rev. Mater. Res. 45, 1 (2015).

[68] N. R. Wilson, P. V. Nguyen, K. Seyler, P. Rivera, A. J. Marsden, Z. P. Laker, G. C. Constantinescu, V. Kandyba, A. Barinov, N. D. Hine, X. Xu, and D. H. Cobden, Sci. Adv. 3, e1601832 (2017).

[69] E. Abrahams, S. V. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. 73, 251 (2001).

[70] A. Punnoose and A. M. Finkel’stein, Science 310, 289 (2005).

[71] E. G. Carnio, N. D. M. Hine, and R. A. Römer, Data for “Resolution of the exponent puzzle for the Anderson transition in doped semiconductors,” http://wrap.warwick.ac.uk/id/eprint/92910 (2018).

[72] H. Terletskya, Y. Zhang, K.-M. Tam, T. Berlijn, L. Chioncel, N. S. Vidhyadhiraja, and M. Jarrell, Appl. Sci. 8, 2401 (2018).

[73] F. J. Wegner, Z. Phys. B: Condens. Matter Quanta 25, 327 (1976).