Enhancement of hopping conductivity by spontaneous fractal ordering of low-energy sites

Tianran Chen[1] and Brian Skinner[2]

[1]*West Chester University, West Chester, Pennsylvania 19383, USA*
[2]*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

Variable-range hopping conductivity has long been understood in terms of a canonical prescription for relating the single-particle density of states to the temperature-dependent conductivity. Here we demonstrate that this prescription breaks down in situations where a large and long-ranged random potential develops. In particular, we examine a canonical model of a completely compensated semiconductor, and we show that at low temperatures hopping proceeds along self-organized, low-dimensional subspaces having fractal dimension $d = 2$. We derive and study numerically the spatial structure of these subspaces, as well as the conductivity and density of states that result from them. One of our prominent findings is that fractal ordering of low energy sites greatly enhances the hopping conductivity and allows Efros-Shklovskii type conductivity to persist up to unexpectedly high temperatures.

I. INTRODUCTION

In disordered systems with localized electrons, there is no electronic conductivity at zero temperature. At nonzero temperature, however, a finite conductivity is enabled by the process of phonon-assisted quantum tunneling, or "hopping," of electrons between localized states. When the temperature T is relatively high, phonons are abundant and electrons hop primarily between nearest-neighboring sites, with an activation energy that is given by the average difference in energy between neighboring sites. On the other hand, when the temperature is low, nearest-neighbor hopping events become extremely rare and the conductivity is dominated by long-distance hops, in which an electron tunnels directly to a distant target site with energy closely matched to that of its initial site. This process is known as variable-range hopping (VRH) and was first described by Mott nearly 50 years ago [1].

Mott's initial theory was eventually generalized into a prescription for determining the temperature-dependent conductivity $\sigma(T)$ from the energy-dependent single-particle density of states (DOS) $g(\varepsilon)$. (Here and below, ε is defined relative to the Fermi level.) In this prescription, one first considers that at a fixed temperature T electrons hop within some band of energies $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ around the Fermi level [2]. The typical spatial concentration of sites within this band is $N(\varepsilon_0) = \int_{-\varepsilon_0}^{\varepsilon_0} g(\varepsilon) d\varepsilon$ and the typical hopping distance can therefore be estimated as $r(\varepsilon_0) = [N(\varepsilon_0)]^{-1/d}$, where d is the system dimensionality. The temperature-dependent conductivity is then calculated by maximizing the typical electron tunneling rate

$$\Gamma \propto \exp\left[-\frac{2r(\varepsilon_0)}{\xi} - \frac{\varepsilon_0}{T}\right] \tag{1}$$

with respect to ε_0, where ξ is the localization length and T is taken to have units of energy. In Eq. (1), the first term in the exponential indicates suppression of tunneling with distance due to exponentially small overlap between electron wave functions, and the second term indicates suppression with energy due to vanishing phonon abundance. The value of Γ is proportional to σ to within algebraic prefactors. This

prescription for determining $\sigma(T)$ from $g(\varepsilon)$ can be called the "Mott doctrine" for understanding VRH, and it has been at the heart of our understanding of hopping conduction for the last half century. Most generally, if the single-particle DOS varies with energy according to $g(\varepsilon) \propto |\varepsilon|^\alpha$, then the Mott doctrine gives

$$\ln \sigma \propto T^{-(\alpha+1)/(d+\alpha+1)}. \tag{2}$$

While Mott initially assumed a constant DOS around the Fermi level ($\alpha = 0$), it was later shown by Efros and Shklovskii [3] (ES) (after earlier suggestions by Pollak [4], Knotek [5], and Srinivasan [6]) that interaction-induced correlations between electrons generically guarantee a vanishing DOS at $\varepsilon \to 0$. One can understand this vanishing DOS by noticing that, in the ground state arrangement, a filled site and an empty site cannot simultaneously have an energy difference ε and a spatial separation $r < e^2/\varepsilon$. If they did, then the system would be unstable with respect to the process of removing the electron from the filled site (upon which the energy of the empty site would drop, since the contribution to the Coulomb potential by the electron at the formerly-filled site would vanish) and adding it to the empty site. In other words, a filled site having energy $\varepsilon_{\rm f}$ and an empty site having energy $\varepsilon_{\rm e}$ must satisfy

$$\varepsilon_{\rm e} - \varepsilon_{\rm f} > e^2/r_{\rm ef}, \tag{3}$$

where $r_{\rm ef}$ is the distance between the two sites. (Here, e^2 denotes the squared electron charge divided by the dielectric constant, in cgs units.) Equation (3) is usually referred to as the "ES stability criterion." It guarantees that sites which are close together in energy cannot be close together in space, and therefore that the total concentration of sites with energy smaller than ε is limited to $N \lesssim (e^2/\varepsilon)^{-d}$. Defining the DOS $g = dN/d\varepsilon$ gives [7]

$$g(\varepsilon) \lesssim \frac{|\varepsilon|^{d-1}}{e^{2d}}. \tag{4}$$

Equation (4) is called the "Coulomb gap;" its typical effect on the DOS in a three-dimensional (3D) system is illustrated in Figs. 1(a) and 1(b). Using the ideas of the Mott doctrine outlined above [Eq. (2)], the conductivity resulting from the
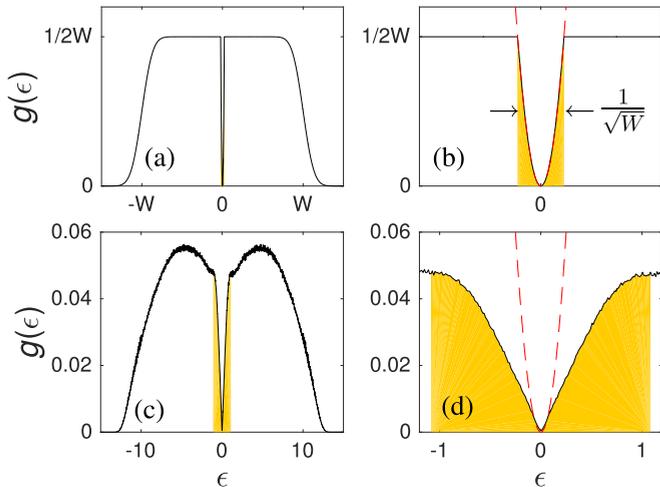
FIG. 1. Single-particle DOS as a function of energy. (a) and (b) show the typical situation, where electron-hole correlations arising from the ES stability criterion dig a narrow Coulomb gap out of a roughly constant DOS. If the total width of the DOS is $\sim 2W$, then the Coulomb gap width is $\sim 1/\sqrt{W}$. (c) and (d) show the results from our numeric simulation for $\Delta = 10$. The red dashed line shows the 3D Coulomb gap, Eq. (4). Shaded regions in these plots show the band of energies, as measured by our numeric simulations, that are used at the highest temperature at which we observe ES VRH [Eq. (5)]. Units of energy in this plot are e^2/a_0 and $g(\varepsilon)$ is plotted in units of $(e^2 a_0^2)^{-1}$.

Coulomb gap is

$$\sigma \propto \exp[-(T_{\text{ES}}/T)^{1/2}], \tag{5}$$

regardless of the system dimensionality, where $T_{\text{ES}} \sim e^2/\xi$. Equation (5) is called the ES law of conductivity, and it arises generically in disordered, localized systems at sufficiently low temperature. The universality of Eqs. (4) and (5) has inspired a large number of analytical, numerical, and experimental studies during the past four decades. (For a limited selection of these, see, for example, Refs. [8–21]).

Importantly, the ES law is expected to apply only at sufficiently low temperatures that the band of energies used for hopping conductivity is within the range where Eq. (4) applies. In other words, ES conductivity is supposed to occur only when the states used for conduction are inside the Coulomb gap. Consider, for example, the most common situation, where electrons hop in a three-dimensional (3D) system for which the single-particle energies fall within some wide range $\varepsilon \in (-W, W)$. In this situation the "bare DOS" is determined by the total width of the energy band, $g_0 \sim 1/(a_0^3 W)$, where a_0 is the typical distance between nearest-neighboring sites. The DOS is expected to follow the Coulomb gap expression, $g_{\text{CG}}(\varepsilon) \sim \varepsilon^2/e^6$, only at such small energies that $g_{\text{CG}}(\varepsilon) < g_0$. This condition implies that the Coulomb gap has a width $\varepsilon_{\text{CG}} \sim \sqrt{e^6/a_0^3 W}$ that becomes increasingly narrow as the range W of site energies is increased. Consequently, the ES law, Eq. (5), describes the conductivity only at low temperatures $T \lesssim e^4 \xi/(a_0^3 W)$. This situation is illustrated schematically in Figs. 1(a) and 1(b).

In this paper we demonstrate that this canonical expectation of ES conductivity inside a parabolic Coulomb gap can in

fact be strongly violated. We examine a simple model that describes hopping transport in a completely compensated semiconductor, and we demonstrate that for this system the Mott doctrine itself is apparently violated, so that the usual relation between $g(\varepsilon)$ and $\sigma(T)$ is broken. The result of this breakdown is that the regime of validity of Eq. (5) is enormously extended in temperature, and ES conductivity persists far outside the regime of a quadratic Coulomb gap [as illustrated in Figs. 1(c) and 1(d)]. We explain this surprising result in terms of a spontaneous fractal ordering of low-energy sites, such that for hopping at low temperature the dimensionality d of the system is effectively reduced. Our arguments constitute an update to the theory of transport in completely compensated semiconductors, and can be used to resolve prominent puzzles about the hopping transport in such diverse systems as nanocrystal arrays [22,23] and the bulk transport of doped, compensated topological insulators [24,25], where an anomalous prolongation of ES conductivity to high temperatures was observed but never explained.

The remainder of this paper is organized as follows. In Sec. II we define the model to be studied and discuss our numerical simulations. In Sec. III we present scaling arguments and numerical results to identify the fractal spatial structure of low-energy sites, and we give results for the DOS. Section IV uses these results to derive the temperature-dependent hopping conductivity, which we verify with simulation results. We conclude in Sec. V with a summary and discussions about relevant recent experiments and the generality of our results.

II. MODEL AND NUMERICAL SIMULATIONS

A. Model

The model we consider can be viewed as a lattice-discretized version of a canonical model of completely compensated semiconductors [2,26], in which electrons can reside on "donor" or "acceptor" sites that occur in equal number. We arrange these sites into a cubic lattice, with lattice constant a_0 and size L in each orthogonal direction. Each site is randomly assigned to be either a donor or acceptor type and can be occupied by at most one electron. Donor and acceptor sites are associated with on-site energies $\pm \Delta/2$, respectively, and electrons interact via an unscreened Coulomb interaction. The corresponding (classical) Hamiltonian is

$$H = \frac{\Delta}{2} \sum_i f_i n_i + \sum_{ij} \frac{e^2}{r_{ij}} q_i q_j, \tag{6}$$

where f_i is a binary variable that discriminates between donors ($f_i = 1$) and acceptors ($f_i = -1$), and $n_i = 0, 1$ is the electron occupation of the ith lattice site. Donor sites are neutral when occupied and acceptors are neutral when empty, so that $q_i = (f_i + 1)/2 - n_i$ is the charge of site i.

This model differs notably from the most well-studied model for hopping transport, in which the term $(\Delta/2) f_i$ is replaced by a random on-site energy that is drawn from a wide, continuous distribution (as studied, for example, in Refs. [7,11,19]). As we discuss below, in our model the binary nature of the on-site energy implies a prominent role for the long-ranged Coulomb potential created by electrons and donor charges.

For the remainder of this paper we take units where $e^2 = a_0 = 1$, so that all lengths are written in units of a_0 and all energies are in units of e^2/a_0. Thus, the only physical parameter of the model is the band gap energy Δ. Our focus in this paper is on the limit $\Delta \gg 1$.

B. Pseudoground states

We study this system numerically using a computer simulation, in which we search for the ground state electron occupation n_i of each lattice site for a given random assignment of each f_i and a given choice of Δ (details of the search algorithm are given, e.g., in Ref. [2]). We operate our simulation in the grand-canonical ensemble with fixed chemical potential $\varepsilon = 0$, so that in the ground state all sites with $\varepsilon < 0$ are filled and all sites with $\varepsilon > 0$ are empty.

Of course, finding the exact ground state for a system with reasonable size is a very difficult computational problem, and in general our algorithm does not produce the exact ground state of the system, but only a "pseudoground state" that is stable with respect to the simultaneous change of any two electron occupation numbers [i.e., Eq. (3) is satisfied for all possible choices of filled and empty sites]. Higher-order stability criteria have been studied elsewhere (for example, in Refs. [19,27]), and their effects are generally very small in the range of energies that we consider in this paper.

Once the set of pseudoground state occupation numbers $\{n_i\}$ is known, we calculate the single-electron energy ε_i at each lattice site:

$$\varepsilon_i = \frac{\Delta}{2} f_i - \sum_{j \neq i} \frac{q_j}{r_{ij}}. \tag{7}$$

The DOS is calculated by making a histogram of the values of ε_i. Results presented below for the DOS correspond to simulated systems of size $L = 80$.

C. VRH conductivity

Once the pseudoground state energy at each site is known, we calculate the conductivity of the system using the approach of the Miller-Abrahams resistor network [28]. This process is described in detail elsewhere (for example, in Refs. [2,29]); briefly, it involves modeling the tunneling rate between a given pair of (not necessarily neighboring) sites using a resistance whose value increases exponentially with the sites' spatial separation and with the corresponding activation energy [as suggested by Eq. (1)]. The conductivity of the network is approximated by identifying the minimum value of the resistance such that, when all higher-resistance links are abandoned, current can still percolate through the network.

Since the magnitude of the conductivity at a given temperature depends on the value of the localization length ξ, we find it expedient to define dimensionless units that incorporate ξ in such a way that it is eliminated as a free variable. In particular, notice that at a given temperature the conductivity can be written as

$$\sigma = \sigma_0 \exp\left[-\frac{2r}{\xi} - \frac{\varepsilon_a}{T}\right], \tag{8}$$

where ε_a is the activation energy, r is the typical hopping distance, and σ_0 is a prefactor that depends only algebraically on temperature. We therefore introduce the dimensionless temperature (restoring explicitly the units e^2 and a_0)

$$T^* = \frac{2a_0^2 T}{e^2 \xi} \tag{9}$$

and the dimensionless logarithm of the conductivity:

$$(\ln \sigma)^* = \frac{\xi}{2a_0} \ln(\sigma/\sigma_0). \tag{10}$$

In these units Eq. (8) can be written simply as

$$(\ln \sigma)^* = -r - \varepsilon_a/T^* \tag{11}$$

(reverting to units where $e^2 = a_0 = 1$), and ξ is eliminated as a separate variable for describing the conductivity.

D. Nearest-neighbor hopping conductivity

The resistor network method for calculating conductivity is expensive computationally, since it involves calculating the effective resistance between all pairs of sites in the system for each value of the temperature. We are therefore limited to studying VRH numerically at only modestly large system sizes $L = 30$. For reasons discussed in Sec. III, this finite size limitation also prevents us from applying the resistor network method to values of Δ larger than 10. Nonetheless, the high temperature conductivity (which we discuss in Sec. IV) can be calculated using a simpler method. At high temperatures, VRH is abandoned in favor of nearest-neighbor hopping with a finite, temperature-independent activation energy ε_a. The value of ε_a can be found by isolating the set of sites whose energies fall within some range $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ and checking whether those sites produce nearest-neighbor percolation across the system. The smallest value of ε_0 for which such percolation is possible approximates the activation energy ε_a. Using this method we can estimate the high-temperature activation energy for systems as large as $L = 80$ and Δ as large as 20.

E. Fractal dimension

Crucial to our explanation of the nature of hopping conductivity in this system is the existence of a nontrivial fractal dimension (or "correlation dimension") for different sets of low energy sites. In principle, this fractal dimension can be measured directly from our numeric simulations by the method of calculating a correlation integral for a given set of sites. This technique was introduced in Ref. [30], and it can be described heuristically as follows. First, one identifies the set of sites whose energy lies within a given energy window. One can then choose an arbitrary site within the set, and define a sphere of radius s centered at the chosen site. The number of other sites in the set whose spatial location is also inside the sphere is $M(s)$. The proportionality $M \propto s^d$ defines the fractal dimension d. For the results presented below, we average the value of $M(s)$ over all choices of the chosen center site and over many random realizations of the donor and acceptor assignments.

III. DOS AND STRUCTURE OF THE RANDOM POTENTIAL

A. Long-ranged potential

In order to understand the transport in this system, and the apparent failure of the Mott doctrine, one must first consider the structure of the random Coulomb potential. As is generically the case in heavily compensated semiconductors, in our model a large value of $\Delta \gg 1$ implies that a large, random Coulomb potential develops throughout the system [2,26,31]. One can see the origin of this potential by first imagining that all donors (with nominal energy $+\Delta/2$) are empty, while all acceptors (with nominal energy $-\Delta/2$) are filled. In this situation all sites become charged either positively (for donors) or negatively (for acceptors), and consequently a random Coulomb potential develops. While the system is neutral on average, statistical fluctuations in the local concentration of donors or acceptors lead to large fluctuations in the Coulomb potential over long length scales, which can be understood using the following scaling argument [26]. Consider a Gaussian sphere of size $R \gg 1$. Such a sphere encloses $\sim R^3$ charges, which sum to zero on average but whose sum has a typical magnitude $Q \sim R^{3/2}$. The typical potential at the surface of the sphere, relative to the value at its center, is $\phi(R) \sim Q/R \sim \sqrt{R}$. One can thus think that the potential experiences increasingly large fluctuations over increasingly long length scales, as in a random walk (illustrated in Fig. 2).

These fluctuations grow with length scale until some typical length R_g, at which the potential becomes large enough to bring donor or acceptor levels to the Fermi level. Further growth of the potential is truncated, since donors or acceptors that cross the Fermi level change their charge state (from $q = \pm 1$ to $q = 0$) and thereby screen the potential. The resulting range of energies for electron sites is therefore as wide as $\sim(-\Delta, \Delta)$. The length R_g can be estimated by equating $\phi(R_g) \sim \Delta$, which gives (with proper numerical coefficients [2,26])

$$R_g = \frac{\Delta^2}{8\pi}. \tag{12}$$

R_g can be thought of as the correlation length of the random potential.
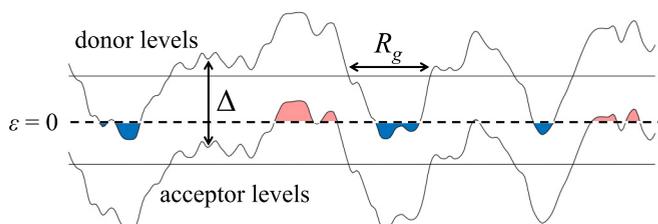
FIG. 2. Schematic illustration of the random potential. The upper (lower) meandering thin lines show the energy of donor (acceptor) sites as a function of spatial coordinate. Screening occurs when these levels cross the Fermi level ($\varepsilon = 0$), at which point donor or acceptor sites become neutralized (indicated by red and blue shaded areas, respectively) and further growth of the potential is truncated. R_g denotes the correlation length of the potential, and horizontal thin lines show $\varepsilon = \pm\Delta/2$.

The wide range of single-particle energies is shown in Fig. 1(c), where we plot the DOS for $\Delta = 10$ as measured by our computer simulation. One can notice from Fig. 1(d) that the DOS follows the 3D Coulomb gap expression, Eq. (4), only over a very narrow range of energies: $|\varepsilon| \lesssim 0.06$. At larger energies $0.06 \lesssim |\varepsilon| \lesssim 1$ the DOS is better described by $g(\varepsilon) \propto |\varepsilon|$. This unusual linear dependence of $g(\varepsilon)$ is one of the central puzzles to be explained.

B. Fractal dimension

Given this picture of a random potential with long-ranged spatial correlations, as illustrated in Fig. 2, one can understand intuitively why the Mott doctrine should fail. Implicit in the Mott doctrine is the assumption that energy levels are drawn randomly from some distribution with no spatial structure. Thus, in the Mott doctrine one estimates the typical distance between sites within some low-energy band in a mean-field way: by calculating the volume-averaged concentration of such sites and raising this concentration to the power $-1/d$. However, in the present problem the gradual development of the random potential implies that sites which are close together in space are more likely than random to be close together in energy. In other words, if two (say) donor sites have a spatial separation $r \ll R_g$, then it is unlikely that they can have a difference in energy that is of order Δ. One can thus expect that sites with similar energy are spatially clustered in such a way that their effective dimension d is smaller than 3.

The value of d can be derived theoretically using the following scaling arguments. In these arguments, our primary goal is to derive the fractal dimension of sites with energy close to the Fermi level. First, however, we find it expedient to derive the fractal dimension of sites with energy *far* from the Fermi level. This derivation provides a useful intermediate result which we then use to derive the fractal dimension of low-energy sites.

1. Fractal dimension of sites with energy far from zero

In order to understand the clustering of sites into fractal subspaces, let us first consider the set of sites having energy far from the Fermi level. For example, take the set of sites with energy in the infinitesimal window $(\Delta/2, \Delta/2 + d\varepsilon)$; such sites are donors experiencing an electric potential close to zero. (One could likewise consider the equivalent set of acceptor sites with energy close to $-\Delta/2$.) Imagine now the process of drawing a box of size s around a given site within this set, and let $M(s)$ denote the expected number of additional sites within the set that will also be inside the box. In a system with no spatial correlations, one would estimate the number of sites within the box to be given by the volume-averaged concentration of sites within the set multiplied by the box volume: $M(s) \sim g(\Delta/2)d\varepsilon s^3$. This result is indeed correct at distances $s \gg R_g$, where the correlated structure of the potential is lost due to screening effects. Substituting $g(\Delta/2) \sim 1/\Delta$ therefore gives $M(s) \sim s^3/\Delta$ at $s \gg R_g$.

On the other hand, at distances $s \ll R_g$, there is no screening, and the structure of the random potential produces a nontrivial fractal dimension d of sites, $M(s) \propto s^d$. To understand this dimension, consider that when the box size $s \ll R_g$, the typical potential at the edge of the box is $\sim\sqrt{s}$

(in our dimensionless units). One can therefore imagine that the energy of donor sites within the box is chosen from a random (Gaussian) distribution with width $\sim\sqrt{s}$ and mean $\Delta/2$. The proportion of donor sites having energy in the window $(\Delta/2, \Delta/2 + d\varepsilon)$ is therefore $\sim d\varepsilon/\sqrt{s}$. Since there are $\sim s^3$ total sites in the box, the total number of sites within the box that are also within the energy window is $\sim s^3 d\varepsilon/\sqrt{s}$, and therefore $M(s) \sim s^{5/2} d\varepsilon$. One can therefore conclude that sites with energy near $\pm\Delta/2$ have a fractal dimension $d = 5/2$ at length scales $s \ll R_g$. Notice in particular that this result for $M(s)$ matches the one derived in the previous paragraph for $s \gg R_g$ when $s \sim R_g \sim \Delta^2$.

2. Fractal dimension of sites with energy close to zero

We now turn our attention to those sites with energy very close to zero. In particular, let us consider the set of sites with energy $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$, where $\varepsilon_0 \ll \Delta$; we refer to this set as the "ε_0 set." Constraints on the value of ε_0 will be discussed below. The spatial arrangement of sites within the ε_0 set is influenced by the structure of the random potential (including the process of nonlinear screening) and by the correlations between filled and empty sites implied by the ES criterion [Eq. (3)]. To derive the effective spatial dimension of these sites, let us first note that for the ε_0 set there are two relevant length scales (in addition to the lattice constant). On one hand, at length scales $s \gg R_g$, the random potential loses its correlations by the process of nonlinear screening, and sites within the set must be three-dimensionally distributed. One can therefore say that at distances $s \gg R_g$ the behavior of the function $M(s)$ is given by $M(s) \sim [\int_{-\varepsilon_0}^{\varepsilon_0} g(\varepsilon) d\varepsilon] s^3 \sim \varepsilon_0 g(\varepsilon_0) s^3$.

While the length R_g is associated with potential fluctuations of order Δ, there is another, much shorter length scale $R_0 \sim \varepsilon_0^2$ associated with fluctuations of order ε_0. That is, over scales $s \ll R_0$ the potential does not develop sufficiently to bring donor or acceptor sites from energy $\sim\varepsilon_0$ to the Fermi level. One can therefore expect that at distances $s \ll R_0$ the function $M(s)$ behaves equivalently as for the sites that have energy far from zero, and $M(s) \sim \varepsilon_0 s^{5/2}$.

At intermediate length scales, $R_0 \ll s \ll R_g$, sites within the ε_0 set are arranged with some fractal dimension d that is to be determined. Let us therefore write the function $M(s)$ in piecewise form

$$M(s) \sim \begin{cases} \varepsilon_0 s^{5/2}, & s \ll R_0 \\ A s^d, & R_0 \ll s \ll R_g \\ \varepsilon_0 g(\varepsilon_0) s^3, & s \gg R_g, \end{cases} \quad (13)$$

where A is an unknown constant to be determined. In order to solve for d, let us first note that if a d-dimensional system having DOS $g_d(\varepsilon)$ (with units of energy per length to the power d) is embedded inside a three-dimensional volume having size R_g, then its 3D DOS (with units of energy per volume) is $g(\varepsilon) \sim g_d(\varepsilon)/R_g^{3-d}$. Let us then make the crucial assumption that within the d-dimensional space the DOS is as large as it can be, limited only by the ES bound [Eq. (4)]. In other words, we set $g_d(\varepsilon) \sim |\varepsilon|^{d-1}$, and therefore

$$g(\varepsilon) \sim \frac{|\varepsilon|^{d-1}}{R_g^{3-d}}. \quad (14)$$

Given this assumption, one can derive the value of the fractal dimension d (and the constant A) by demanding continuity of the function $M(s)$ at both $s \sim R_0$ and $s \sim R_g$. By examining Eq. (13), and substituting $R_0 \sim \varepsilon_0^2$ and $R_g \sim \Delta^2$, we get $A \sim \varepsilon_0^2$ and $d = 2$. So we conclude that the fractal dimension of the ε_0 set at length scales $R_0 \ll s \ll R_g$ is 2.

One can now ask the question: What is the largest value of ε_0 for which this derivation is applicable? In particular, consider that in a volume of size R_g a two-dimensional (2D) subspace contains only $\sim R_g^2$ total sites. Sites within the ε_0 set must therefore comprise only $\sim 1/R_g$ of the total. This gives a constraint $\varepsilon_0 g(\varepsilon_0) \lesssim 1/R_g$, and using Eq. (14) we arrive at $\varepsilon_0 \lesssim 1$.

To summarize, our scaling analysis leads us to the following dramatic conclusion: Sites with energy $|\varepsilon| \lesssim 1$ are arranged along a two-dimensional subspace over the parametrically wide range of length scales $1 \ll s \ll R_g$. This theoretical prediction can be compared directly to results from our numerical simulations, from which we can measure the function $M(s)$ for sites with energy in the band $(-1, 1)$ (see Sec. II E). The results are shown in Fig. 3(a) for a system of size $L = 80$ and gap $\Delta = 20$, and they compare favorably with the derived function in Eq. (13). A drawing of the typical spatial configuration of sites with $|\varepsilon| < 1$ is displayed in Fig. 3(b).

C. Low-energy DOS

Using the results of the preceding subsection, one can understand the low-energy DOS as follows. At vanishingly small energies, the DOS is constrained by the 3D Coulomb gap equation, $g(\varepsilon) \sim \varepsilon^2$. Such a parabolic depletion of the DOS arises from the ES stability criterion, Eq. (3), applied over distances much longer than the correlation length, $r \gg R_g$. This relation therefore constrains the density of states only at energies $|\varepsilon| \sim 1/r \ll 1/R_g$.

At larger energies $1/R_g \ll \varepsilon \ll 1$, the DOS is constrained by the ES criterion applied among pairs of sites with separation $r \ll R_g$; i.e., among sites within the same 2D cluster. This constraint produces the DOS presented in Eq. (14) (with $d = 2$). Putting these two results together gives the following result for the low-energy DOS:

$$g(\varepsilon) \sim \begin{cases} \varepsilon^2, & \varepsilon \ll 1/R_g \\ |\varepsilon|/R_g, & 1/R_g \ll \varepsilon \ll 1. \end{cases} \quad (15)$$

These equations coincide closely with simulation results at different values of Δ, as demonstrated in Fig. 4.

The implications of this DOS for the conductivity are discussed in detail in the following section. However, already from the derivation of the DOS one can understand the reason for the persistence of ES conductivity outside the regime of a parabolic DOS. Within the range of temperatures where the activation energy ε_a satisfies $1/R_g \ll \varepsilon_a \ll 1$, hopping takes place not along a set of randomly-positioned sites in three dimensions, but rather within a 2D subspace of sites having $|\varepsilon| \lesssim 1$. The existence of this subspace greatly enhances the conductivity and also allows the Coulomb gap to play a much more prominent role in the conductivity.
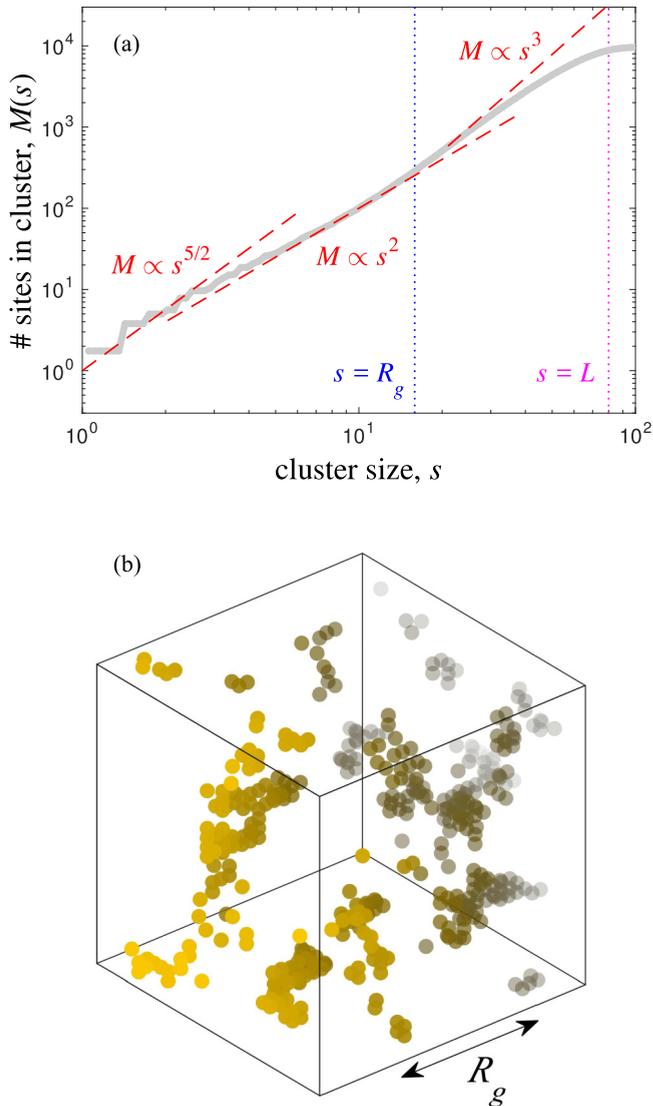
FIG. 3. Fractal ordering of sites with $|\varepsilon| < 1$ for a system with size $L = 80$ and gap $\Delta = 20$. (a) The function $M(s)$ is plotted, as measured by our numeric simulation (thick gray line) and as derived theoretically in Eq. (13) (red dashed lines). The dashed lines use the derived values $d = 2$ and $A \sim 1$ and have no free parameters. The vertical dotted lines indicate the length scales R_g and L. (b) An example of the spatial arrangement of sites with $|\varepsilon| < 1$. Points show the locations of such sites in a typical pseudoground state. (Only a subvolume of the total simulated system is shown.) The points are colored according to their distance from the front left face of the system.

IV. TEMPERATURE-DEPENDENT CONDUCTIVITY

The central result of this paper, derived in the preceding section, is that in this system the long-ranged Coulomb potential induces sites with energy $|\varepsilon| \lesssim 1$ to arrange along 2D subspaces of the sample volume. Applying the ES stability criterion to such sites gives a DOS that varies linearly with energy. One can therefore understand the conductivity in a relatively simple way, by adapting the Mott doctrine to account for the reduced dimensionality of the system over length scales $1 \ll r \ll R_g$. In particular, inserting $\alpha = 1$ and $d = 2$ into
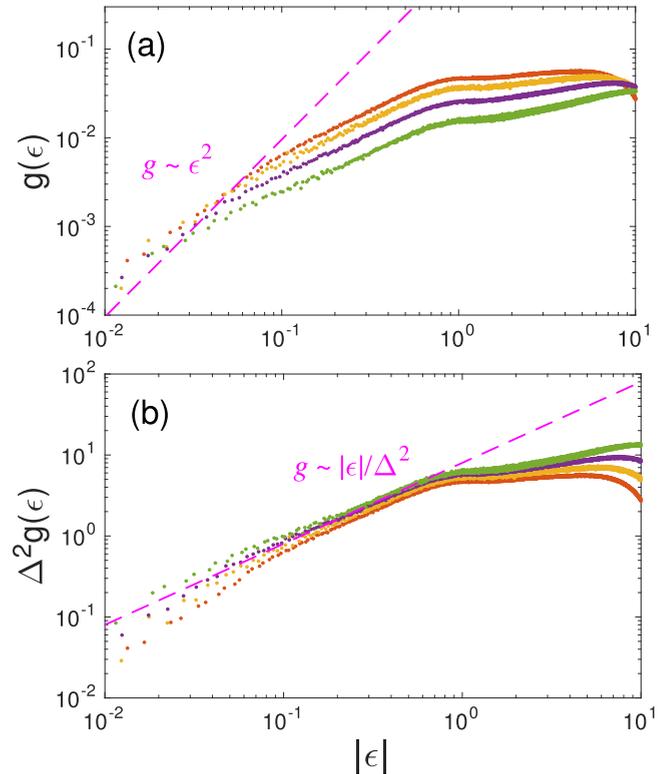
FIG. 4. (a) Density of states as a function of energy for different values of Δ, as measured by our numeric simulations. From highest to lowest, the curves correspond to $\Delta = 10, 12, 15$, and 20. The dashed line shows the 3D Coulomb gap equation. (b) When the DOS is plotted as $\Delta^2 g(\varepsilon)$ versus energy, all the data collapses onto a single curve $g(\varepsilon) \sim |\varepsilon|/\Delta^2$ within the energy range $1/R_g \ll \varepsilon \ll 1$.

Eq. (2) gives an ES-like temperature exponent, $\ln \sigma \propto T^{-1/2}$, while a naive use of the system dimensionality $d = 3$ would give an erroneous temperature exponent $\ln \sigma \propto T^{-2/5}$.

Further, applying the Mott argument to this reduced-dimensional space suggests that ES conductivity persists all the way until $T^* \sim 1$. The correctness of this conclusion is verified explicitly in Fig. 5(a), where we plot the logarithm of the conductivity against $(T^*)^{-1/2}$. Note that the appearance of ES conductivity at relatively high temperatures $T^* \sim 1$ is in strong contrast to the result one would derive from the measured DOS using the conventional Mott doctrine. Indeed, such a calculation would suggest that ES conductivity occurs only at extremely small temperatures, $T^* \lesssim 1/\Delta^4$, due to the tiny domain of the parabolic Coulomb gap.

From our simulation we can also measure the numerical value of the characteristic temperature T_{ES}, defined by Eq. (5). This measurement gives $T_{ES}/(e^2/\xi) = 6.4 \pm 0.1$. This result is consistent with the value of T_{ES} reported for 2D systems, for which numerical studies [32] give $T_{ES}/(e^2/\xi) \approx 6$ and a self-consistent theory [33] gives $T_{ES}/(e^2/\xi) = 6.5$. This agreement lends further support to our picture of hopping on fractal 2D subspaces. In principle, at extremely small temperatures $T^* \ll 1/\Delta^4$ the typical hop length becomes much longer than R_g, and VRH is 3D in nature rather than 2D. Such 3D ES conductivity is known to correspond to a different numerical value of $T_{ES}/(e^2/\xi) \approx 2.8$ [2]. Unfortunately, we
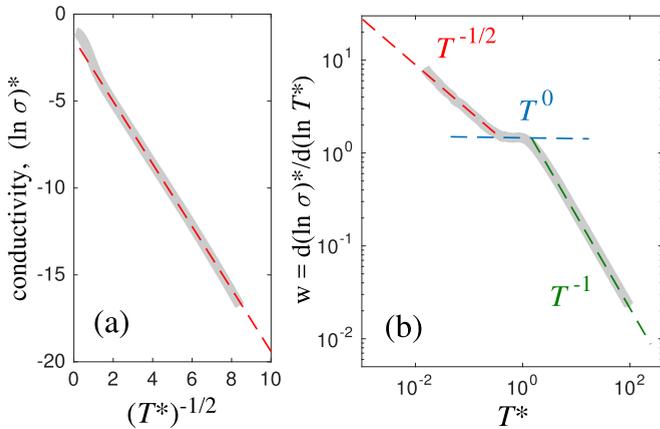
FIG. 5. Temperature-dependent conductivity for $\Delta = 10$, as measured by our numeric simulation. (a) The dimensionless logarithm of the conductivity, $(\ln \sigma)^*$, as a function of $(T^*)^{-1/2}$. The thick gray curve shows our numerical result, and the dashed red line is a linear fit. The linear dependence at $(T^*)^{-1/2} \gtrsim 1$ implies that the conductivity follows Eq. (5) at all $T^* \lesssim 1$. (b) Zabrodskii-Zinov'eva analysis of conductivity at different temperature regimes [34,35]. The vertical axis shows the reduced activation energy $w = d(\ln \sigma)^*/d(\ln T^*)$, which has a power-law dependence on T^* with an exponent that reflects the VRH mechanism. The thick gray curve is our numerical result, and dashed lines indicate the results of Eq. (16).

FIG. 6. Activation energy for nearest-neighbor hopping ε_a as a function of the band gap Δ. Black points are our numerical results, and the solid line is a linear fit. The dash-dotted line is the result reported in Ref. [25].

are unable to confirm this change in T_{ES} at low temperatures due to finite size limitations, which restrict our simulations to shorter hop lengths.

While the primary purpose of this paper is to study and explain the low temperature conductivity, it is worth commenting on the conductivity at higher temperatures as well. In particular, one can ask about the crossover between VRH at low temperature and nearest-neighbor activated conductivity at higher temperature. Our picture of the low-temperature conductivity is that at $T^* \ll 1$ conductivity occurs through VRH on 2D subspaces having low energy $|\varepsilon| \lesssim 1$. As the temperature is raised to the point that $T^* \sim 1$, the typical hopping length becomes of order unity. It is therefore natural to ask whether these 2D subspaces percolate across the system, in the nearest-neighbor sense. If they do, then at high temperatures electrons can hop across the system using only chains of nearest-neighboring sites with $|\varepsilon| \lesssim 1$, and the corresponding activation energy will be of order unity, regardless of the value of Δ. On the other hand, if the 2D subspaces do not produce nearest-neighbor percolation across the system, then at high temperatures the current will be forced to pass through high energy sites, and the activation energy will increase with increasing Δ.

In order to distinguish between these two possibilities, we measure the activation energy ε_a for nearest-neighbor hopping as a function of Δ, using the method described in Sec. II D. The result is plotted in Fig. 6. Our results suggest that ε_a increases linearly with Δ as $\varepsilon_a \approx 0.17\Delta + 0.45$. This is consistent with an earlier theoretical study of the bulk transport in completely compensated topological insulators [25], which studied the case $\Delta = 10$ (for a model where donors and acceptors are placed randomly, rather than on a lattice) and found that $\varepsilon_a \approx 0.15\Delta$.
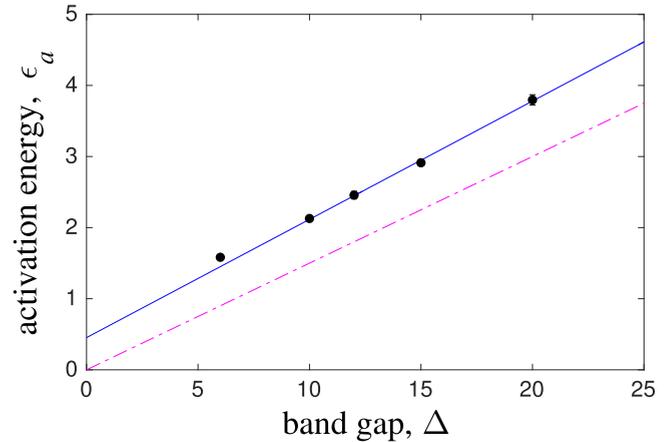
The observed increase of ε_a with Δ implies that there is no nearest-neighbor percolation of 2D subspaces. Instead, one can envision that the 3D system has many 2D subspaces embedded within it, each of size $\sim R_g$, separated by small gaps of order one lattice site [as suggested by the picture in Fig. 3(b)]. When the typical hopping length is slightly larger than unity, electrons can hop across these gaps, and the activation energy is $\varepsilon_a \sim 1$. However, if the hopping length is reduced to precisely one lattice constant, then the activation energy increases sharply to $\varepsilon_a \sim 0.17\Delta$, which is (by our measurement) the energy required to traverse the gaps between 2D subspaces and produce nearest-neighbor percolation.

This picture suggests that the activation energy for conductivity has an essentially discontinuous dependence on the hopping length r when r approaches 1. If one takes this picture literally, then the temperature dependence of the conductivity is given by

$$(\ln \sigma)^* \sim \begin{cases} -(1/T^*)^{1/2}, & T^* \ll 1 \\ -1, & 1 \ll T^* \ll 0.17\Delta \\ -(0.17\Delta/T^*), & T^* \gg 0.17\Delta. \end{cases} \quad (16)$$

These three regimes of temperature dependence are consistent with our simulation results, as shown in Fig. 5(b).

The intermediate temperature regime in Eq. (16), in which $\ln \sigma$ is roughly temperature independent, is somewhat unusual in the context of hopping transport. (In general, our results suggest that σ will have only a power-law dependence on temperature in the intermediate temperature range.) It remains unclear whether such a flat temperature dependence can persist in more general models where electron sites are randomly positioned, or whether it is an artifact of our use of a discrete lattice. It is worth noting, however, that a similar progression of temperature exponents, including an intermediate regime where $(\ln \sigma)^* \sim 1$, has been reported experimentally [36,37].

V. SUMMARY AND DISCUSSION

In this paper we have studied the problem of VRH in a simple model of a completely-compensated semiconductor,

and shown that hopping conductivity is greatly enhanced at low temperatures relative to the canonical expectation using the Mott doctrine. This enhancement arises from fractal ordering of low-energy sites, which are pushed by the long-ranged random Coulomb potential into low-dimensional subspaces having fractal dimension $d = 2$. We have derived this fractal dimension using simple scaling arguments, which we confirm with numeric results. The resulting picture allows us to derive theoretically the DOS and the temperature-dependent conductivity [Eqs. (15) and (16), respectively]. Both results are in close agreement with numerical simulations. Our findings represent a significant update to the theory of electronic transport in completely compensated semiconductors, which was initialized over 45 years ago [26,38].

The most prominent experimental consequence of our results is that ES conductivity persists to a much higher temperature than the naive expectation: $T^* \sim 1$ rather than $T^* \sim 1/\Delta$. Such enhancement has in fact been seen recently in a number of contexts. For example, a series of works recently explored the hopping conductivity in assemblies of semiconductor nanocrystals [22,23,39–41]. In such systems conductivity takes place through hopping of electrons between nanocrystals. Within each nanocrystal, the conduction band is split into a sequence of discrete energy shells associated with quantum confinement of electrons. When the doping level in the assembly is such that on average an integer number of these shells is completely filled, the system resembles a completely compensated semiconductor, with the gap between quantum confinement shells playing the role of the band gap Δ. This situation was studied extensively both experimentally [22] and theoretically [23], and it was shown [22] that ES conductivity persists up to temperatures as high as $T^* \sim 2$. The DOS $g(\varepsilon)$ was also found using computer modeling [23] to have a highly linear dependence on energy up to $|\varepsilon| \sim 1$, which further confirms the picture presented here.

Another prominent example of enhanced hopping conductivity has arisen in the study of 3D topological insulators (TIs). As grown, 3D TI crystals are usually *n*-type semiconductors, and a significant experimental effort has been expended with the goal of silencing their bulk conductivity. Such efforts usually proceed through compensation of donor electrons by acceptors, so that optimally resistive TIs are completely compensated semiconductors [24]. However, even in the completely compensated state, the bulk conductivity is often frustratingly large experimentally, and this large conductivity has been explained in terms of the band-bending effects discussed in Sec. III [25,31,42]. While much attention was paid to the unexpectedly small activation energy for conductivity

at high temperatures (as examined in Fig. 6), these systems also display an unusually robust VRH at low temperatures. In particular, VRH conductivity with temperature exponent close to 1/2 has been seen to persist up to temperatures as high as 100 K in compensated samples of Bi_2Se_3 [24]. This is in strong contrast to the conventional expectation (as outlined in the Introduction), which instead predicts ES conductivity to end at temperatures of order 10 K. This puzzling observation is explained by the theory developed here, and indeed in Bi_2Se_3 the temperature $T^* = 100$ K corresponds to $T^* \sim 1.5$, so that this observation is completely consistent with our results.

Finally, let us remark on the generality of our results. Most generally, one can expect our results to apply for any system that has both a finite concentration of unscreened charged impurities and a chemical potential that lies in the middle of a large energy gap. For such systems, a large, random Coulomb potential invariably arises and produces a nontrivial spatial structure in the energy landscape (as illustrated in Fig. 2) and therefore leads to a violation of the Mott doctrine. Of course, our model has not included the presence of "diagonal disorder" (i.e., uncorrelated on-site disorder), which generically exists with some finite amplitude W. (For example, in nanocrystal arrays such disorder can arise from random variations in the nanocrystal diameter.) However, if this diagonal disorder amplitude is small enough that $W \ll \Delta$, then its effects are overwhelmed by those of the long-ranged Coulomb potential, and our model applies. In particular, our primary conclusion that ES conductivity persists up to temperatures $T^* \sim 1$ remains unchanged. On the other hand, for systems with incomplete compensation, such that either donor or acceptor type impurities predominate, the chemical potential shifts toward either the conduction or valence band edge and our theory can lose its validity. Since the typical amplitude of the disorder potential in a system with degree of compensation K is $\Gamma \sim e^2 N_i^{1/3}(1-K)^{-1/3}$ [2,43], one can expect our results to apply so long as $1 - K \ll 1$ and $\Gamma \gg W$. A full analysis of the case of finite compensation and finite diagonal disorder is left for a future study.

[1] N. F. Mott, Conduction in glasses containing transition metal ions, J. Non-Cryst. Solids **1**, 1 (1968).

[2] B. I. Shklovskii and A. L. Efros, *Electronic Properties of Doped Semiconductors* (Springer-Verlag, New York, 1984).

[3] A. L. Efros and B. I. Shklovskii, Coulomb gap and low temperature conductivity of disordered systems, J. Phys. C: Solid State Phys. **8**, L49 (1975).

[4] M. Pollak, Effect of carrier-carrier interactions on some transport properties in disordered semiconductors, Discuss. Faraday Soc. **50**, 13 (1970).

[5] M. L. Knotek and M. Pollak, Correlation effects in hopping conduction: A treatment in terms of multielectron transitions, Phys. Rev. B **9**, 664 (1974).

[6] G. Srinivasan, Statistical mechanics of charged traps in an amorphous semiconductor, Phys. Rev. B **4**, 2581 (1971).

[7] A. L. Efros, Coulomb gap in disordered systems, J. Phys. C: Solid State Phys. **9**, 2021 (1976).

[8] S. D. Baranovskii, A. L. Efros, B. L. Gelmont, and B. I. Shklovskii, Coulomb gap in disordered systems: computer simulation, J. Phys. C **12**, 1023 (1979).

[9] A. I. Larkin and D. E. Khmelnitskii, Activation conductivity in disordered systems with large localization length, Sov. Phys. JETP **56**, 647 (1982).

[10] J. H. Davies, P. A. Lee, and T. M. Rice, Properties of the electron glass, Phys. Rev. B **29**, 4260 (1984).

[11] A. Möbius, M. Richter, and B. Drittler, Coulomb gap in two- and three-dimensional systems: Simulation results for large samples, Phys. Rev. B **45**, 11568 (1992).

[12] F. G. Pikus and A. L. Efros, Critical Behavior of Density of States in Disordered System with Localized Electrons, Phys. Rev. Lett. **73**, 3014 (1994).

[13] Q. Li and P. Phillips, Unexpected activated temperature dependence of the conductance in the presence of a soft Coulomb gap in three dimensions, Phys. Rev. B **49**, 10269 (1994).

[14] C. C. Yu, Time-Dependent Development of the Coulomb Gap, Phys. Rev. Lett. **82**, 4074 (1999).

[15] M. Lee, J. G. Massey, V. L. Nguyen, and B. I. Shklovskii, Coulomb gap in a doped semiconductor near the metal-insulator transition: Tunneling experiment and scaling ansatz, Phys. Rev. B **60**, 1582 (1999).

[16] V. Yu. Butko, J. F. DiTusa, and P. W. Adams, Coulomb Gap: How a Metal Film Becomes an Insulator, Phys. Rev. Lett. **84**, 1543 (2000).

[17] M. Müller and L. B. Ioffe, Glass Transition and the Coulomb Gap in Electron Glasses, Phys. Rev. Lett. **93**, 256403 (2004).

[18] B. Surer, H. G. Katzgraber, G. T. Zimanyi, B. A. Allgood, and G. Blatter, Density of States and Critical Behavior of the Coulomb Glass, Phys. Rev. Lett. **102**, 067205 (2009).

[19] M. Goethe and Matteo Palassini, Phase Diagram, Correlation Gap, and Critical Properties of the Coulomb Glass, Phys. Rev. Lett. **103**, 045702 (2009).

[20] A. Amir, Y. Oreg, and Y. Imry, Variable range hopping in the Coulomb glass, Phys. Rev. B **80**, 245214 (2009).

[21] E. Bardalen, J. Bergli, and Y. M. Galperin, Coulomb glasses: A comparison between mean field and Monte Carlo results, Phys. Rev. B **85**, 155206 (2012).

[22] H. Liu, A. Pourret, and P. Guyot-Sionnest, Mott and efros-shklovskii variable range hopping in CdSe quantum dots films, ACS Nano **4**, 5211 (2010).

[23] B. Skinner, T. Chen, and B. I. Shklovskii, Theory of hopping conduction in arrays of doped semiconductor nanocrystals, Phys. Rev. B **85**, 205316 (2012).

[24] Z. Ren, A. A. Taskin, S. Sasaki, K. Segawa, and Y. Ando, Optimizing $Bi_{2-x}Sb_xTe_{3-y}Se_y$ solid solutions to approach the intrinsic topological insulator regime, Phys. Rev. B **84**, 165311 (2011).

[25] B. Skinner, T. Chen, and B. I. Shklovskii, Why Is the Bulk Resistivity of Topological Insulators So Small? Phys. Rev. Lett. **109**, 176801 (2012).

[26] B. I. Shklovskii and A. L. Efros, Completely compensated crystalline semiconductor as a model of an amorphous semiconductor, Sov. Phys. JETP **35**, 610 (1972).

[27] A. L. Efros, B. Skinner, and B. I. Shklovskii, Coulomb gap in the one-particle density of states in three-dimensional systems with localized electrons, Phys. Rev. B **84**, 064204 (2011).

[28] A. Miller and E. Abrahams, Impurity conduction at low concentrations, Phys. Rev. **120**, 745 (1960).

[29] A. L. Efros and B. I. Shklovskii, Coulomb interaction in disordered systems with localized electronic states, in *Electron-electron Interactions in Disordered Systems*, Modern Problems in Condensed Matter Sciences, Vol. 10, edited by A. L. Efros and M. Pollak (North Holland, Amsterdam, 1985), p. 409.

[30] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, Phys. D: Nonlinear Phenom. **9**, 189 (1983).

[31] B. Skinner, T. Chen, and B. I. Shklovskii, Effects of bulk charged impurities on the bulk and surface transport in three-dimensional topological insulators, J. Exp. Theor. Phys. **117**, 579 (2013).

[32] D. N. Tsigankov and A. L. Efros, Variable Range Hopping in Two-Dimensional Systems of Interacting Electrons, Phys. Rev. Lett. **88**, 176602 (2002).

[33] E. I. Levin, V. L. Nguen, B. I. Shklovksii, and A. L. Efros, Coulomb gap and hopping electric conduction. Computer simulation, Sov. Phys. JETP **65**, 842 (1987).

[34] A. G. Zabrodskii and I. S. Shlimak, Influence of doping inhomogenities on the conductivity of heavily doped and compensated semiconductors, Sov. Phys. Semicond. **9**, 391 (1975).

[35] A Zabrodskii and K Zinov'eva, Low-temperature conductivity in metal-insulator transition in compensated n-Ge, Sov. Phys. JETP **59**, 425 (1984).

[36] L. R. Wienkes, C. Blackwell, and J. Kakalios, Electronic transport in doped mixed-phase hydrogenated amorphous/nanocrystalline silicon thin films, Appl. Phys. Lett. **100**, 072105 (2012).

[37] L. R. Wienkes, C. Blackwell, T. Hutchinson, and J. Kakalios, Conduction mechanisms in doped mixed-phase hydrogenated amorphous/nanocrystalline silicon thin films, J. Appl. Phys. **113**, 233707 (2013).

[38] B. I. Shklovskii and A. L. Efros, Impurity band and conductivity of compensated semiconductors, Sov. Phys. JETP **33**, 468 (1971).

[39] K. Sharma, A. S. Al-Kabbi, G. S. S. Saini, and S. K. Tripathi, Electrical conduction mechanism in nanocrystalline CdTe (nc-CdTe) thin films, Appl. Phys. A **108**, 911 (2012).

[40] T. Chen, B. Skinner, W. Xie, B. I. Shklovskii, and U. R. Kortshagen, Carrier transport in films of alkyl-ligand-terminated silicon nanocrystals, J. Phys. Chem. C **118**, 19580 (2014).

[41] T. Chen, K. V. Reich, N. J. Kramer, H. Fu, U. R. Kortshagen, and B. I. Shklovskii, Metal-insulator transition in films of doped semiconductor nanocrystals, Nat. Mater. **15**, 299 (2016).

[42] N. Borgwardt, J. Lux, I. Vergara, Z. Wang, A. A. Taskin, K. Segawa, P. H. M. van Loosdrecht, Y. Ando, A. Rosch, and M. Grüninger, Self-organized charge puddles in a three-dimensional topological material, Phys. Rev. B **93**, 245149 (2016).

[43] T. Chen and B. I. Shklovskii, Anomalously small resistivity and thermopower of strongly compensated semiconductors and topological insulators, Phys. Rev. B **87**, 165119 (2013).