Coulomb Gap Triptych in a Periodic Array of Metal Nanocrystals

Tianran Chen, Brian Skinner, and B. I. Shklovskii
Fine Theoretical Physics Institute, University of Minnesota, Minneapolis, Minnesota 55455, USA
(Received 25 April 2012; published 19 September 2012)

The Coulomb gap in the single-particle density of states (DOS) is a universal consequence of electron-electron interaction in disordered systems with localized electron states. Here we show that in arrays of monodisperse metallic nanocrystals, there is not one but three identical adjacent Coulomb gaps, which together form a structure that we call a "Coulomb gap triptych." We calculate the DOS and the conductivity in two- and three-dimensional arrays using a computer simulation. Unlike in the conventional Coulomb glass models, in nanocrystal arrays the DOS has a fixed width in the limit of large disorder. The Coulomb gap triptych can be studied via tunneling experiments.

DOI: 10.1103/PhysRevLett.109.126805 PACS numbers: 73.21.La, 72.20.Ee, 73.23.Hk, 73.63.Bd

Granular metals and arrays of metallic nanocrystals (NCs) represent interesting composite systems, wherein the unique properties of individual NCs are combined with the collective, correlation-driven effects between the NCs to produce novel material properties [1,2]. One of the most important properties is the electron conductivity, which, in the limit of weak coupling between NCs, proceeds by phonon-assisted tunneling, or "hopping," between the NCs through the insulating gaps that separate them. In relatively dense NC arrays, electron conduction can occur both through nearest-neighbor hopping and through "cotunneling" of electrons between distant NCs via a chain of intermediate virtual states [2–5]. In the presence of some disorder, the latter mechanism dominates at low temperatures, where the length of the hops grows to optimize the conductivity. This transport mechanism was introduced by Mott [6], and is called variable range hopping. When the Coulomb interaction between localized electrons is taken into account, it can be shown that at a sufficiently low temperature, variable range hopping conductivity obeys the Efros-Shklovskii (ES) law [7]

$$\sigma = \sigma_0 \exp[-(T_{\mathrm{ES}}/T)^{1/2}], \tag{1}$$

where σ_0 is a constant (or a weak power-law function of temperature) and T_{ES} is a characteristic temperature. Equation (1) has been observed in a number of granular metal systems at low temperatures (see Ref. [2] and the references therein). In these systems, as in lightly doped semiconductors and other "Coulomb glasses," ES conductivity can be seen as the result of a vanishing single-particle density of states (DOS) at the Fermi level μ. This vanishing DOS is the consequence of a very general stability criterion of the ground state [8], and it implies that in a system of d dimensions the DOS $g(E)$ satisfies

$$g(E) < \frac{A_d}{e^{2d}} |E|^{d-1}. \tag{2}$$

Here, A_d is some numerical constant of order unity, E is the electron energy relative to the Fermi level, and e is the electron charge. Equation (2) is called the "Coulomb gap."

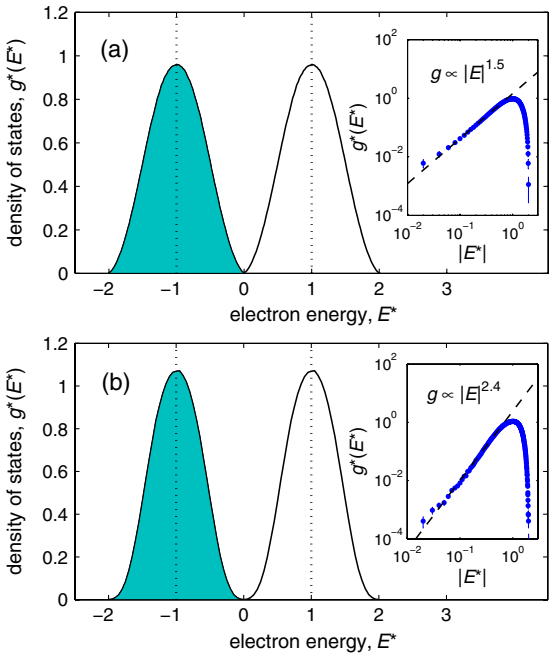

FIG. 1 (color online). The DOS of a regular array of monodisperse NCs where $E^* = E/(e^2/2C_0)$ is the dimensionless single-particle energy and $g^*(E^*) = (e^2 D^d/2C_0)g(E^*)$ the dimensionless DOS; D is the NC diameter. Here, the results are shown from a computer simulation of (a) a two-dimensional (2D) square lattice and (b) a three-dimensional (3D) cubic lattice. The shaded area shows filled electron states, and the empty area indicates empty states. In addition to electron-hole symmetry, the two peaks of the DOS have mirror symmetries across $E^* = \pm 1$ (dotted lines). This symmetry creates from the central Coulomb gap two additional half-gaps at $E^* = \pm 2$, resulting in a "Coulomb gap triptych." The insets show the DOS near the Fermi level $E^* = 0$ in log-log scale.

 © 2012 American Physical Society

In this Letter, we report an additional striking feature of the DOS in periodic arrays of monodisperse metal NCs surrounded by a good insulator with random impurity charges. We show that the Coulomb gap at $E = 0$ necessarily implies the existence of additional, identical Coulomb gaps at energies $E = \pm e^2/C_0$, where C_0 is the self capacitance of each NC. This result is shown in Fig. 1.

In Fig. 1, one can see that to the right of the Fermi level, at $E > 0$, the DOS curve $g(E)$ is reflected across the vertical line $E = e^2/2C_0$ into the interval $e^2/2C_0 < E < e^2/C_0$. Similarly, on the left side of the Fermi level, the portion of the $g(E)$ curve corresponding to $-e^2/2C_0 < E < 0$ is reflected across the vertical line $E = -e^2/2C_0$. We refer to this characteristic shape as a Coulomb gap triptych. As we explain below, its structure is the result of an additional symmetry in the system that arises from the discrete charging spectrum of individual NCs. In this way, the Coulomb gap triptych represents a bridge between the concepts of the Coulomb gap and the Coulomb blockade.

Experimentally, regular arrays of metal NCs can now be reliably synthesized with diameter D in the range 3–7 nm and size dispersion less than 5% [1,2,5]. For such small NCs, the self capacitance C_0 is also small: $C_0 = \kappa D/2$, where κ is the effective dielectric constant of the array, given approximately by the Maxwell-Garnett formula [9,10]. Correspondingly, the Coulomb self-energy $q^2/2C_0$ of an NC with charge q plays a large and important role in electron transport. To see this, one should imagine a hypothetical NC array with no disorder. In such an array, in the ground state all the NCs are neutral and electron conduction requires the thermal excitation of positive-negative NC pairs. Thus, conductivity is activated with an activation energy $e^2/2C_0$. For nanometer-sized NCs, this activation energy can easily exceed the thermal energy k_BT.

In the presence of some finite charge disorder, however, the fluctuating Coulomb potential can cause charging of NCs in the ground state and thus lead to a Coulomb gap in the DOS and ES conductivity. To show how this happens, we adopt the following simplified model in this Letter. We assume that identical, spherical, metallic NCs reside in a regular d-dimensional square lattice with lattice constant D' and that impurity charges $\pm e$ are embedded in the insulator (oxide) between the NCs. Such impurity charges effectively create a fractional donor charge Q_i that resides on each NC i, for reasons that are explained below. The net charge of the NC can then be written as $q_i = Q_i - en_i$, where n_i is the integer number of electrons that reside on the NC relative to its neutral state (n_i can be either positive or negative). Given this model, the Hamiltonian for the system is

$$H = \sum_i \frac{(Q_i - en_i)^2}{2C_0} + \sum_{\langle i,j \rangle} C_{ij}^{-1}(Q_i - en_i)(Q_j - en_j). \quad (3)$$

Here, the first term describes the Coulomb self-energy of each NC and the second term describes the interaction between the charged NCs. The coefficient C_{ij}^{-1} is the inverse of the matrix of electrostatic induction C_{ij}. This Hamiltonian has also been proposed as a model for arrays of large semiconductor NCs [11].

Owing to the presence of impurity charges, the electrons become redistributed among the NCs from their neutral state so as to screen the disorder Coulomb potential. In order to calculate the DOS and the conductivity, we first attempt to find numerically the set of electron occupation numbers $\{n_i\}$ that minimizes the Hamiltonian. In the numerical simulations that we describe below, we take the approximations that $C_0 = \kappa D/2$ and $C_{ij}^{-1} = 1/\kappa r_{ij}$. These approximations do not affect our main conclusions, as we explain below.

The model of fractional donor charges Q_i was first put forward in Ref. [3]; here, its justification is briefly repeated. When an impurity charge, say, with a charge $+e$, is located close to the point of contact between two NCs labeled A and B, it induces negative image charges $-q_A$ and $-q_B$ in the surfaces of the NCs A and B, respectively. This is shown schematically in Fig. 2. In order to maintain overall neutrality of the NCs, an equal and opposite image charge appears at the center of each NC: $+q_A$ and $+q_B$. (These "image charges at the center" represent a uniform electronic charge at the NC surface.) The values of q_A and q_B are such that together the image charges $-q_A$ and $-q_B$ neutralize the donor charge: $q_A + q_B = e$. Their respective magnitudes are determined by the distance between the impurity and each NC surface. For example, if the impurity sits exactly along the line connecting the centers of NCs A and B, and if the gap $w = D' - D$ between the NCs satisfies $w \ll D$, then $q_A x_B = q_B x_A$, where x_A and x_B are the distances between the impurity and the surfaces of the NCs A and B, respectively. Since the impurity charges and the image charges $-q_A$ and $-q_B$ together form a compact, neutral arrangement, the net effect of the impurity charge is to produce "fractionalized" donor charges, such that $+q_A$ is relayed to the center of NC A and $+q_B$ is relayed to the center of B.

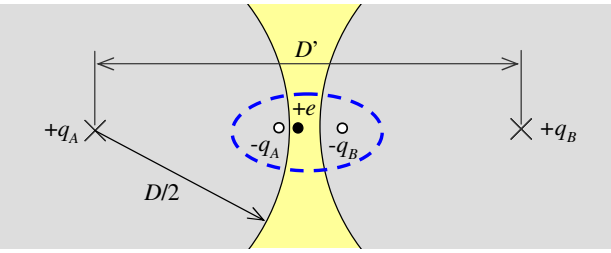

FIG. 2 (color online). A schematic depiction of the fractionalization of a charged impurity (small black circle) between NCs (large gray circles). The positive impurity induces negative image charges (white circles) in nearby metal surfaces and is effectively neutralized, while equal and opposite positive images are conveyed to the center of the NCs (×'s).

In this way, each NC i can be said to have a fractional donor charge Q_i, which is equal to the sum total of the fractionalized charges donated by individual impurities around it. In the limit where there are very many impurity charges surrounding each NC, one can suppose that the random variable Q_i is Gaussian distributed with some standard deviation larger than e. In fact, however, in such cases one can effectively adopt a much simpler model in which the value of Q_i is chosen randomly from the uniform distribution $Q_i \in [-e/2, +e/2]$. To see why this model is valid, consider that each NC minimizes its Coulomb self-energy by minimizing the magnitude of its net charge, $|Q_i - en_i|$. Since n_i can take any integer value, it is generally true that in the ground state $-e/2 \leq Q_i - en_i \leq e/2$. In other words, each NC can effectively adjust to the presence of an arbitrarily strong charge disorder by changing its electron number n_i (say, by drawing electrons from the voltage source) so that its net charge acquires a magnitude smaller than $e/2$. This has important implications for the disorder dependence of conductivity, as we show below.

Given the ground state configuration for a particular system, defined by the set of electron occupation numbers $\{n_i\}$, one can determine the energy of the highest filled electron level, $E_i^{(f)}$, and the lowest empty electron level, $E_i^{(e)}$, at each NC i. Specifically:

$$E_i^{(f)} = \frac{2e^2 n_i - 2Q_i e - e^2}{2C_0} - e\sum_{j\neq i} C_{ij}^{-1}(Q_j - en_j), \quad (4)$$

$$E_i^{(e)} = \frac{2e^2 n_i - 2Q_i e + e^2}{2C_0} - e\sum_{j\neq i} C_{ij}^{-1}(Q_j - en_j). \quad (5)$$

These energies are defined so that the Fermi level $\mu = 0$ and in the ground state $E_i^{(f)} < 0$ and $E_i^{(e)} > 0$ for all i. The single-particle DOS $g(E)$ is defined by making a histogram of the energy values $E_i^{(f)}$ and $E_i^{(e)}$, and has also been termed the "density of ground states" [3]. More highly excited electron energy states are ignored in this study, as they play no role in conductivity at $k_B T \ll e^2/C_0$.

In order to evaluate numerically the DOS, we use a computer simulation to search for the ground state arrangement of electrons, $\{n_i\}$, in a finite array of NCs. For simplicity, we set the lattice constant $D' = D$; this corresponds to the limit where the gap w between the NCs is very thin while the tunneling transparency of the barrier between them remains much smaller than unity. In our simulation, we search for the ground state by looping over all NC pairs i, j and attempting to move one electron from i to j. If the move lowers the Hamiltonian H, then it is accepted, otherwise it is rejected. Equivalently, one can say that for all i, j we check that the ES ground state criterion is satisfied:

$$E_j^{(e)} - E_i^{(f)} - e^2 C_{ij}^{-1} > 0. \quad (6)$$

It should be noted that this procedure does not in general find the exact ground state, but instead finds only a "pseudoground state" that is stable with respect to single-electron transfers. In principle, the system energy can be lowered further by some simultaneous multielectron transfers. Such processes are generally seen to have only a relatively weak effect on the DOS [12,13] that slightly deepens the Coulomb gap near the Fermi level.

The resulting DOS is shown in Fig. 1(a) for a two-dimensional (2D) simulated system of size 100×100 lattice sites, and in Fig. 1(b) for a three-dimensional (3D) system of size $25 \times 25 \times 25$. Electron energies are plotted in the dimensionless form $E^* = E/(e^2/2C_0)$, and the DOS is plotted in the dimensionless form $g^*(E^*) = (e^2 D^d/2C_0)g(E^*)$. The insets to these figures show a log-log plot of the DOS near $E = 0$, suggesting that in 2D systems the DOS follows $g_{2D}(E) \propto E^{1.5}$ at small energies and in 3D systems $g_{3D}(E) \propto E^{2.4}$. These exponents are somewhat larger than the theoretical ones given in Eq. (1), so that apparently the ES bound is not saturated. This is similar to what happens in the Efros model of the Coulomb glass [14] at disorder strength $A = 1$ [12]. The results of Fig. 1 are generated using a uniform distribution $Q_i \in [-e/2, e/2]$ for the fractional charge. If one instead takes Q_i to be Gaussian distributed, with a standard deviation $\sim e$, the resulting DOS is everywhere equal to that of Fig. 1 to within 0.6%.

Figure 1 also highlights the striking additional symmetry in the DOS in both 2D and 3D, as compared to the DOS in the conventional Coulomb glass problem [8,12]. That is, each peak in the DOS is symmetric with respect to reflections about $E^* = \pm 1$, so that the DOS has identical, repeated Coulomb gaps at $E^* = \pm 2$. The origin of these additional Coulomb gaps can be understood by noting a particular symmetry in the Hamiltonian that is reflected in the filled and empty state energies, $E_i^{(f)}$ and $E_i^{(e)}$. Specifically, by combining Eqs. (4) and (5), one can show that

$$E_i^{*(e)} = E_i^{*(f)} + 2 \quad (7)$$

for all i. Thus, all NCs contribute to the DOS two energy levels—one filled and one empty—separated by e^2/C_0. This implies that as the DOS collapses at $|E^*| \to 0$ (the Coulomb gap), the DOS must also collapse at $|E^*| \to 2$ in identical fashion. That is, the ES stability criterion of Eq. (6) places constraints both on the DOS near $E = 0$ and on the DOS near $E = \pm e^2/C_0$. A similar observation was made for a somewhat different system in the very recent Ref. [15].

One can also note that states with $E_i^{*(f)} < -2$ or $E_i^{*(e)} > 2$ are prohibited, since by Eq. (7) these would imply that some NCs have $E_i^{(e)} < 0$ or $E_i^{(f)} > 0$. Thus, $g(E)$ is strictly zero at $|E^*| > 2$. This is a markedly different situation than in the conventional Efros model [14], where the width of the DOS reflects the characteristic strength of the disorder.

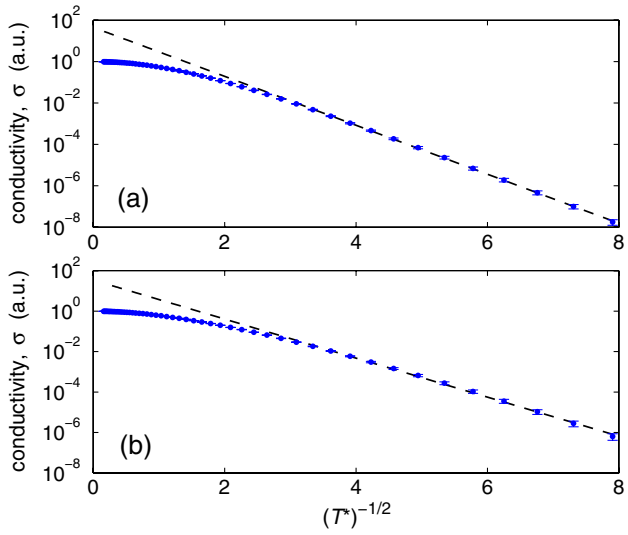

FIG. 3 (color online). The temperature dependence of the conductivity in (a) 2D and (b) 3D. In both cases, the conductivity follows the ES law [Eq. (1)] at small temperatures, $T^* \ll 1$, as shown by the dashed lines.

In the present problem, for a large enough disorder the DOS has a saturated width e^2/C_0. This saturation occurs because the number of electrons n at each site can adjust to screen an arbitrarily large Coulomb disorder. Thus, one can expect that for large disorder, the conductivity also becomes independent of disorder strength.

In order to evaluate the conductivity directly, we employ the approach of the Miller-Abrahams network [16], in which each pair ij of NCs is said to be connected by some equivalent resistance R_{ij}. The value of R_{ij} increases exponentially with the distance r_{ij} between the NCs and the activation energy ΔE_{ij} required for the electron hopping between i and j according to $R_{ij} \propto \exp[2r_{ij}/\xi + \Delta E_{ij}/k_B T]$, where ξ is the electron localization length [3] and the value of ΔE_{ij} is determined by the ground state energies $\{E_i^{(f)}\}$ and $\{E_i^{(e)}\}$ [11]. The resistance of the system as a whole is found using a percolation approach. Specifically, we find the minimum value R_c such that if all resistances R_{ij} with $R_{ij} < R_c$ are left intact while others are eliminated (replaced with $R = \infty$), then there exists a percolation pathway connecting opposite faces of the simulation volume. The conductivity σ of the system is equated with $1/(R_c D^{d-2})$.

Our results for the conductivity are shown in Fig. 3, plotted as a function of the dimensionless temperature $T^* = 4DC_0 k_B T/(e^2\xi)$ raised to the power $-1/2$. The results indicate that the conductivity is well described by the ES law of Eq. (1) at relatively small temperatures $T^* \ll 1$, both in 2D and 3D [17]. This behavior is consistent with the prominent Coulomb gaps seen in Fig. 1. In both 2D and 3D, replacing the uniform distribution of Q_i with a distribution with larger variance—for example, by taking Q_i as the sum of three or more independent fractional charges—did not affect the conductivity to within our numerical accuracy. This insensitivity to the disorder strength stands in contrast to the Efros model [14], where a large disorder widens the DOS, so that ES conductivity exists only when the temperature is sufficiently small and the electron hops are confined to within the parametrically narrow window of energies in which $g(E)$ is constrained by the Coulomb gap [8]. On the contrary, in arrays of monodisperse metallic NCs, the DOS becomes essentially independent of disorder strength, so that even at large disorders, the Coulomb gap plays a prominent role and the conductivity follows the ES law.

The triptych structure of the DOS should have observable consequences for a number of experiments on metal NC arrays. It is possible, for example, that the DOS can be probed directly by tunneling experiments, similar to the ones that have directly observed the Coulomb gap in doped semiconductors [18]. For systems with a finite dispersion δC in NC self-capacitance, the repeated Coulomb gaps will be smeared over some finite energy interval rather than collapsing to zero exactly at $E^* = \pm 2$. We simulated this behavior numerically by adding a stochastic spatial variation to C_0, and for a root mean square deviation $\delta C \ll C_0$, we arrived at $g(E^* = \pm 2)/g(E^* = \pm 1) \approx 3(\delta C/C_0)^2$. This implies that for a system with 5% dispersion in the NC diameter, the collapse of the DOS at $E^* = \pm 2$ is complete to within 1%, and the resulting $g^*(E^*)$ curve would not be distinguishable from that of Fig. 1 if added to the plot.

The authors would like to thank E. Aydil, A. L. Efros, A. Frydman, Yu. M. Galperin, M. Goethe, A. Kamenev, U. Kortshagen, M. Müller, A Möbius, M. Palassini, and L. Wienkes for helpful discussions. This work was supported partially by the MRSEC Program of the National Science Foundation under Grant No. DMR-0819885. T. C. was partially supported by the FTPI.

[1] D. V. Talapin, J.-S. Lee, M. V. Kovalenko, and E. V. Shevchenko, Chem. Rev. **110**, 389 (2010).
[2] I. S. Beloborodov, A. V. Lopatin, V. M. Vinokur, and K. B. Efetov, Rev. Mod. Phys. **79**, 469 (2007).
[3] J. Zhang and B. I. Shklovskii, Phys. Rev. B **70**, 115317 (2004).
[4] I. S. Beloborodov, A. V. Lopatin, and V. M. Vinokur, Phys. Rev. B **72**, 125121 (2005).
[5] H. Moreira, Q. Yu, B. Nadal, B. Bresson, M. Rosticher, N. Lequeux, A. Zimmers, and H. Aubin, Phys. Rev. Lett. **107**, 176803 (2011).
[6] N. Mott, J. Non-Cryst. Solids **1**, 1 (1968).
[7] A. L. Efros and B. I. Shklovskii, J. Phys. Chem. Solids **8**, L49 (1975).
[8] A. L. Efros and B. I. Shklovskii, *Electronic Properties of Doped Semiconductors* (Springer-Verlag, New York, 1984).

[9] J. C. Maxwell, *A Treatise on Electricity and Magnetism* (Clarendon, Oxford, 1891), 3rd ed., Vol. 2, p. 57.
[10] T. Chen, B. Skinner, and B. I. Shklovskii, Phys. Rev. B **84**, 245304 (2011).
[11] B. Skinner, T. Chen, and B. I. Shklovskii, Phys. Rev. B **85**, 205316 (2012).
[12] A. Möbius, M. Richter, and B. Drittler, Phys. Rev. B **45**, 11568 (1992).
[13] A. L. Efros, B. Skinner, and B. I. Shklovskii, Phys. Rev. B **84**, 064204 (2011).
[14] A. L. Efros, J. Phys. Chem. Solids **9**, 2021 (1976).
[15] J. Mitchell, A. Gangopadhyay, V. Galitski, and M. Müller, Phys. Rev. B **85**, 195141 (2012).
[16] A. Miller and E. Abrahams, Phys. Rev. **120**, 745 (1960).
[17] In fact, if one repeats the original ES derivation [7] using the DOS shown in Fig. 1, one arrives at a slightly different temperature dependence $\ln\sigma \propto T^{-\gamma}$ at low temperature, with $\gamma \approx 0.56$ in 2D and $\gamma \approx 0.53$ in 3D. Due to finite size limitations, our conductivity data (Fig. 3) cannot discriminate between these exponents and $\gamma = 1/2$.
[18] J. G. Massey and M. Lee, Phys. Rev. Lett. **75**, 4266 (1995).