PHYSICAL REVIEW B **95**, 214205 (2017)

Cotunneling and polaronic effect in granular systems

A. S. Ioselevich[1,2,*] and V. V. Sivak[3,4]
[1]*Condensed Matter Physics Laboratory, National Research University Higher School of Economics, Moscow 101000, Russia*
[2]*L. D. Landau Institute for Theoretical Physics, Moscow 119334, Russia*
[3]*Department of Physics, Yale University, New Haven, Connecticut 06520, USA*
[4]*Moscow Institute of Physics and Technology, Moscow 141700, Russia*

(Received 10 April 2017; revised manuscript received 24 May 2017; published 13 June 2017)

We theoretically study the conductivity in arrays of metallic grains due to the variable-range multiple cotunneling of electrons with short-range (screened) Coulomb interaction. The system is supposed to be coupled to random stray charges in the dielectric matrix that are only loosely bounded to their spatial positions by elastic forces. The flexibility of the stray charges gives rise to a polaronic effect, which leads to the onset of Arrhenius-type conductivity behavior at low temperatures, replacing conventional Mott variable-range hopping. The effective activation energy logarithmically depends on temperature due to fluctuations of the polaron barrier heights. We present the unified theory that covers both weak and strong polaron effect regimes of hopping in granular metals and describes the crossover from elastic to inelastic cotunneling.

DOI: 10.1103/PhysRevB.95.214205

I. INTRODUCTION

In this paper we are discussing a multiparticle cotunneling mechanism of conductivity in a granular metal with flexible charges, randomly placed in the insulating matrix. The flexibility of disorder gives rise to a sort of "random polaronic effect," which dramatically affects the temperature dependence of conductivity. A study of this subject requires a combination of different physical concepts and methods, so we start from a brief review of the necessary ingredients.

A. Variable-range hopping

The stretched-exponential temperature dependence of conductivity

$$\sigma \propto \exp\{-(T_0/T)^{\alpha}\} \tag{1}$$

is characteristic of variable-range hopping (VRH) in homogeneously disordered materials (such as amorphous solids [1] or doped semiconductors [2]) at low temperatures. In the case when the long-range interactions do not play any significant role, the exponent $\alpha = 1/(d+1)$ (where d is the space dimensionality), $T_0 \equiv T_{\mathrm{M}} \sim [\nu_F \xi^d]^{-1}$ (where ν_F is the density of states at the Fermi level and ξ is the inverse decrement of electronic wave functions), and the corresponding dependence is known as the Mott law [3]. In the opposite case, when the long-range Coulomb interaction is crucial and gives rise to the soft Coulomb gap in the electronic density of states [4], the exponent $\alpha = \frac{1}{2}$, $T_0 \equiv T_{\mathrm{ES}} \sim e^2/\kappa\xi$ (where κ is the dielectric constant) and this dependence is known as the Efros-Shklovskii law.

B. Variable-range cotunneling in granular systems

A similar behavior of conductivity was also observed in arrays of metallic and semiconducting quantum dots in the temperature range 1–200 K (see reviews [5,6] and some old [7] and more recent [8] experimental papers). An explanation of the stretched-exponential T dependence of conductivity in granular materials attracted interest of theorists. Many of early theories [9] were based on special assumptions about the distribution of the random parameters of grains (sizes, etc.) and were criticized (see, e.g., [10]) because of their *ad hoc* character and the lack of universality. Some other theories [11,12] correctly indicated the important role of sequential tunneling of electrons through a chain of intermediate grains, but did not give a correct multiparticle description for this tunneling and a valid recipe for evaluation of the tunneling amplitude. Such a prescription was worked out in [13], where the simple idea, introduced in [11], was generalized to take into account both multiparticle character (i.e., the cotunneling, see [14,15]) of the process and Coulomb effects.

In contrast with the standard single-particle-like VRH scenario, where a particle would travel from one end of the chain to the other, consequently hopping through all intermediate grains, the multiple cotunneling scenario developed in [13] involves all possible sequences of hops between neighboring grains in the chain. In general, these hops are executed by different electrons; the intermediate states of the process involve many grains with altered charges. However, upon the completion of the process there is only one net electron that is transferred between the terminal grains of the chain, while the charges of all intermediate grains return to their initial values. It does not mean that all the processes with different sequences of hops lead to the same final state of the system: the final states of some grains may be identical to the initial ones (elastic cotunneling), while for other grains the initial and final states may differ by an electron-hole pair (inelastic cotunneling). As a result (see [13] for details), the law (1) is reproduced with

$$T_0 \sim \mathcal{L}(T) E_c, \tag{2}$$

where

$$\mathcal{L}(T) \sim \begin{cases} \ln(E_c/g\delta), & T \ll T_{c0} \\ \ln\left(E_c^2/gT^2\mathcal{L}^2\right), & T \gg T_{c0} \end{cases} \tag{3}$$

and the Coulomb charging energy is $E_c \sim e^2/2C$, C being typical capacitance in the system of grains. A typical level

---

[*] iossel@itp.ac.ru

spacing in a grain $\delta \sim (\nu_F a^3)^{-1}$, a being the grains size, and $g \ll 1$ is the typical dimensionless conductance between adjacent grains. Note that the logarithmic factor $\mathcal{L}$ is large. The crossover between the elastic and inelastic cotunneling regimes takes place at

$$T \sim T_{c0} = \sqrt{E_C\delta}/\mathcal{L}. \tag{4}$$

Thus, the stretched exponential law in the case of granular materials is slightly modified in the intermediate temperature range ($T > T_{c0}$) due to additional logarithmic T dependence of T_0. This deviation from the Mott-Efros-Shklovskii law is, however, not easy to detect experimentally.

C. Hard gap and polaronic effect in homogeneously disordered systems

It is well known that at relatively high temperatures the VRH is not operative, it is changed to the nearest-neighbor hopping (NNH), so that the stretched-exponential law (1) is replaced by the Arrhenius law for conductivity $\sigma \propto \exp\{-\varepsilon_3/T\}$ (see [2]). What is much less trivial, in some cases [16] the reentrance of the Arrhenius law

$$\sigma \sim \exp\{-E_H/T\} \tag{5}$$

is observed also at *low* temperatures! This reentrance is usually attributed to polaronic effect: a “hard gap” E_H is supposed to be related to the energy, necessary for the creation in advance of a polaronic cloud, that then will accommodate a hopping electron at a new position.

In principle, the hopping electron can take along the necessary energy while hopping from the initial position to the new one, therefore, at still lower temperatures the activational mechanism of polaron hopping is substituted by the tunneling one (see [1]) and the stretched exponential law (1) is again restored at $T \lesssim \omega$, where ω is the characteristic frequency of phonons (or some other species, e.g., magnons, localized electronic excitations, etc.), that constitute the polaronic cloud. The case of magnetic polarons is special: Due to the local conservation of the magnetization the process of the tunneling transfer of the polaronic cloud is strongly suppressed [17], and the classical-quantum crossover is shifted from $T \sim \omega$ to much lower temperatures. It explains why the hard gap phenomenon was experimentally observed predominantly (but not exclusively!) in magnetic systems. In general, the theory of VRH with account for polaronic effect was developed in [17–20] for different types of polarons.

It is important that if the strength of the polaronic effect randomly varies from place to place, then it does not necessarily have to lead to the activation Arrhenius law. If the distribution of barrier heights has a power-law tail at zero, one can expect the dependence (1) with $\alpha \neq 1$. In particular, for homogeneously disordered solids it was shown in [20] that, if the barrier distribution is constant in the vicinity of zero, the Mott conductivity should have the exponent $\alpha = 2/(d+2)$.

D. Polaronic effect in granular systems

The onset of Arrhenius-type behavior of conductivity at low temperatures in granular materials is reported far less often. For example, in experiments [21] the Arrhenius behavior was observed below ~5 K in two-dimensional arrays of semiconducting Ge/Si quantum dots. Above that temperature the conductivity followed Efros-Shklovskii law. In paper [22] the observation of “almost” Arrhenius conductivity was reported in the temperature range $20\ \mathrm{K} < T < 30\ \mathrm{K}$ in the array of metallic Co nanoparticles. This dependence was well described by (1) with $\alpha = 1.1$. At higher temperatures $45\ \mathrm{K} < T < 80\ \mathrm{K}$ Efros-Shklovskii law was found. There are also experiments [23] in which $\alpha = \frac{2}{3}$ was observed in the broad temperature range $7\ \mathrm{K} < T < 200\ \mathrm{K}$ in the arrays of ZnO nanocrystals.

It is tempting to attribute these findings to some kind of random polaronic effect. But, what does the polaron effect mean in the case where we deal not with single charge carriers, but with grains, containing many electrons? And, what are the effective degrees of freedom that constitute here a polaronic cloud? To our knowledge, no models of the polaronic effect in granular systems have been discussed so far. In this paper, we introduce such a model for metallic grains and study its implications for the transport properties within the framework of multiple cotunneling concept.

E. Structure of the paper

The paper is organized in the following way. We discuss the basic concepts of our model in Secs. II and III. We start from the general model, including (i) general (possibly long-range) interaction in the system of grains and impurities, (ii) elasticity of the impurities, (iii) dynamics of impurities, and (iv) intergrain hopping of electrons. Then, we simplify the initial model, first by discarding the long-range part of the interaction and second by assuming the impurities to be so heavy that, according to the Franck-Condon principle, the positions of impurities may be treated as “quasistatic” from the point of view of intergrain electronic transitions.

The detailed calculation of electron transition rate between the pair of distant resonant grains in the case of short-range interaction and quasistatic impurities is carried out in Secs. IV-VII. We deal with different parameter ranges and also provide physical interpretation of results. Based on these findings, we analyze the temperature behavior of conductivity in Secs. VIII and IX. Section X contains the summary of our results and in Sec. XI we discuss the limitations and possible future directions of research.

II. DYNAMICAL FLUCTUATIONS OF THE OFFSET CHARGES: GENERAL MODEL

The main source of disorder in granular systems ar the “stray charges,” hardly removable charged impurities and defects, trapped in the insulating part of the system (see Fig. 1). They produce random Coulomb fields acting on the grains, so that the Coulomb energy of the system is

$$E_C(\vec{N},\vec{Q}) = \tfrac{1}{2}\{(\vec{N}-\vec{Q})\hat{U}(\vec{N}-\vec{Q}) - \vec{Q}\hat{U}\vec{Q}\}, \tag{6}$$

where we have introduced vectorial notation $\vec{N}\hat{U}\vec{N} \equiv \sum_{jj'} U_{jj'}N_jN_{j'}$. Here, integer N_j denotes a number of excess electrons on jth grain, $\hat{U} \equiv e^2\hat{C}^{-1}$ is the inverse matrix of capacitances, and the components Q_j of the vector $\vec{Q}$ are the so-called “offset charges” (not necessarily integers!). It should

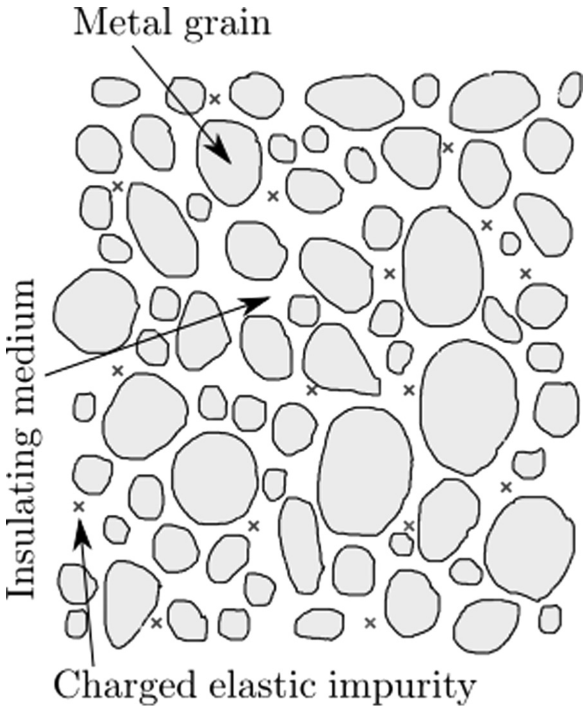


FIG. 1. Sample of granular material with trapped elastic impurities.

be noted that each offset charge Q_j can not be identified with certain unique impurity: all impurities that effectively interact with a given grain j contribute to Q_j. Vice versa, each impurity may effectively contribute to many different variables Q_j.

In the context of VRH, the offset charges are usually treated as static random variables, but in this paper we are going to take into account their dynamics. Indeed, Q_j depend on the positions of the charged impurities, that are not absolutely rigid, but can deviate from their equilibrium places. In the harmonic approximation, these deviations are governed by the Hamiltonian $\hat{H}_{\rm dev} = \hat{E}_{\rm dev} + \hat{\mathcal{H}}_{\rm kin}$,

$$\hat{E}_{\rm dev} = \tfrac{1}{2}(\vec{Q} - \vec{Q}^{(0)})\hat{K}(\vec{Q} - \vec{Q}^{(0)}), \qquad (7)$$

$$\hat{\mathcal{H}}_{\rm kin} = -\frac{\hbar^2}{2}\partial_{\vec{Q}}\hat{M}^{-1}\partial_{\vec{Q}}, \qquad (8)$$

where the positively defined matrix of effective masses $\hat{M}$ is related to the masses of impurities. The vector $\vec{Q}^{(0)}$ describes the set of equilibrium values of offset charges for the case of neutral grains: $\vec{N} \equiv 0$. The matrix $\hat{K}$ in (7) is related to the stiffness of the system with respect to displacements of the charged impurities. It contains both the "mechanical" part (due to deformation of surrounding medium) and the "electrostatic" part [the variation of the electrostatic energy of grains (6) due to the displacement of impurities]. The second part depends on the set $\vec{N}$; namely, since $K_{jj'}(\vec{N})$ is linearly related to the second derivative of the energy with respect to the coordinates of impurities, $\hat{K}$ should be a quadratic polynomial in $\vec{N}$:

$$K_{jj'}(\vec{N}) = K^{(0)}_{jj'} + \sum_k K^{(1)}_{jj'k}N_k + \sum_{kk'} K^{(2)}_{jj'kk'}N_k N_{k'}. \qquad (9)$$

The mechanical part contributes only to the first, N-independent, term in (9). Therefore, in the most natural case, when the mechanical stiffness dominates over the electrostatic one, the N dependence of $\hat{K}$ can be neglected. Anyway, even if the electrostatic part of $\hat{K}$ is considerable, at low temperatures it seems reasonable to ignore the "live" $\vec{N}$ dependence of $\hat{K}$, and replace the function $K_{jj'}(\vec{N})$ by its equilibrium value

$$K_{jj'}(\vec{N}) = K^{(0)}_{jj'} + \sum_k K^{(1)}_{jj'k}N^{(\rm eq)}_k + \sum_{kk'} K^{(2)}_{jj'kk'}N^{(\rm eq)}_k N^{(\rm eq)}_{k'}, \qquad (10)$$

where $\vec{N}^{(\rm eq)}$ are the charges that the grains acquire at the equilibrium.

A. Classical ground state

Since the effective masses are large, in the leading approximation, the kinetic energy term (8) in the Hamiltonian can be neglected, and the ground state of the system corresponds to the minimum of the total potential energy of the system of charges

$$\begin{aligned} E_{\rm ch}(\vec{N},\vec{Q}) &= E_C(\vec{N},\vec{Q}) + E_{\rm dev}(\vec{Q}) \\ &= \tfrac{1}{2}\{(\vec{N} - \vec{Q})\hat{U}(\vec{N} - \vec{Q}) - \vec{Q}\hat{U}\vec{Q}\} \\ &\quad + \tfrac{1}{2}(\vec{Q} - \vec{Q}^{(0)})\hat{K}(\vec{Q} - \vec{Q}^{(0)}). \end{aligned} \qquad (11)$$

Minimizing (11) with respect to $\vec{Q}$ at fixed $\vec{N}$, we get

$$\vec{Q}_{\rm min}(\vec{N}) = \vec{Q}^{(0)} + \hat{K}^{-1}\hat{U}\vec{N}, \qquad (12)$$

$$\min_{\vec{Q}} E_{\rm ch}(\vec{Q},\vec{N}) = \tfrac{1}{2}\vec{N}\hat{\mathcal{U}}\vec{N} - \vec{N}\hat{U}\vec{Q}^{(0)}, \qquad (13)$$

where the symmetric matrix

$$\hat{\mathcal{U}} = \hat{U} - \hat{U}\hat{K}^{-1}\hat{U} \qquad (14)$$

has a meaning of the inverse capacitance matrix, renormalized due to the effects of finite elasticity of the system.

For the stability of the system's ground state, the matrix $\hat{\mathcal{U}}$ must be positively defined. A violation of this requirement would mean that we have incorrectly chosen the ground-state set $\vec{N}^{(\rm eq)}$, which turned out to be unstable; the system eventually will move to a different state, where the stability will be restored due to nonlinearity of the problem, expressed in the $\hat{K}(\vec{N})$ dependence (9).

Further minimization of (13) with respect to $\vec{N}$ gives the equilibrium values of charges as integers, closest (in a sense, see below) to $\hat{\mathcal{U}}^{-1}\hat{U}\vec{Q}^{(0)}$. As a result,

$$\vec{N}^{(\rm eq)} = \hat{\mathcal{U}}^{-1}\hat{U}\vec{Q}^{(0)} + \vec{\gamma}, \qquad (15)$$

$$\vec{Q}^{(\rm eq)} = \vec{Q}^{(0)} + \hat{K}^{-1}\hat{U}\vec{N}^{(\rm eq)}, \qquad (16)$$

where $\vec{\gamma}$ is the set of "effective residual offset charges." Note that in the case of rigid impurities (when $\hat{K} \to \infty$) $\hat{\mathcal{U}} \to \hat{U}$, and the effective offset charges are reduced to the standard ones: $\vec{\gamma} \to$ noninteger part of $\vec{Q}^{(0)}$.

The common restriction, usually imposed on the residual offset charges, reads as

$$-1/2 \leqslant \gamma_k \leqslant 1/2. \tag{17}$$

Strictly speaking, this is not correct in general case: for symmetric matrix $\hat{\mathcal{U}}$ the all-integer-$\vec{N}$ minimum of the energy functional may, in principle, lay quite far from the unrestricted one, so that some components of $\vec{\gamma}$ may be quite large. Examples are easy to produce (say, a highly anisotropic potential profile with valleys, looking in low-symmetry directions) but all these examples seem to be exotic, if not pathological. At least we were not able to construct any physically relevant matrix $\hat{\mathcal{U}}$ for that the restriction (17) would be violated. Anyway, it is definitely valid for the case which we are going to study in detail below: the screened Coulomb interaction with diagonal matrix $\hat{\mathcal{U}}$. Therefore, in what follows we will consider the restriction (17) granted.

B. Low-energy Hamiltonian

At low temperatures, both the grains' charges $\vec{N}$ and the offset charges $\vec{Q}$ only slightly deviate from the equilibrium values, so one can write

$$\vec{N} = \vec{N}^{(\mathrm{eq})} + \vec{n}, \quad \vec{Q} = \vec{Q}^{(\mathrm{eq})} + \vec{q}, \tag{18}$$

where n_k may take values $(-1,0,1)$. Now, we are prepared to rewrite the Hamiltonian in terms of deviations $\vec{n}$ and $\vec{q}$. Substituting (18) into (11) and omitting the terms, that do not contain deviations, we get

$$E_{\mathrm{ch}} = \tfrac{1}{2}\{(\vec{n} - \vec{q}\,)\hat{U}(\vec{n} - \vec{q}) + \vec{q}\,(\hat{K} - \hat{U})\vec{q}\} + \vec{n}\,\hat{\mathcal{U}}\vec{\gamma}. \tag{19}$$

C. Thermodynamic excitation energies

Aside from the ground state, the low-lying excited states are of great importance for the transport properties of the system. For the state $\vec{n}$ the thermodynamic (i.e., minimized with respect to $\vec{q}$) excitation energy is

$$\tilde{E}(\vec{n}) \equiv \min_{\vec{q}} E_{\mathrm{ch}}(\vec{n},\vec{q}) = \tfrac{1}{2}\vec{n}\,\hat{\mathcal{U}}(\vec{n} + 2\vec{\gamma}). \tag{20}$$

Different branches of spectrum (19) can be visualized as multidimensional paraboloids in the $\vec{q}$ space. The paraboloids are indexed by vector $\vec{n}$. Excitation energy $\tilde{E}(\vec{n})$ is nothing else but the energetic distance between the bottom of corresponding paraboloid and the global ground-state energy (see Fig. 2). For brevity, we have denoted $\tilde{E}_i \equiv \tilde{E}(\vec{n}_i)$ and $\tilde{E}_f \equiv \tilde{E}(\vec{n}_f)$, where i stands for "initial state" and f for "final state." We will reserve these indices for this purpose and will never use them as grain labels.

Also, to avoid further confusion we should mention that in this paper one will encounter two types of indices: (1) those that label the charging states of a system, i.e., they label vectors $\vec{n}$ (as an example one can consider the initial state $\vec{n}_i$); (2) those that label the grains, i.e., the entries in vector $\vec{n}$. For example, $[\vec{n}_i]_m$ refers to the charge of the mth grain in the initial state $\vec{n}_i$.

1. *Single-particle excitations*

Below we will obtain some crucial characteristics of single-particle excited states, in which only one entry $[\vec{n}]_k$ in $\vec{n}$ is nonzero, while all other entries are zeros (namely, $\vec{n} = \pm\vec{z}_k$ where $[\vec{z}_k]_m = \delta_{km}$).

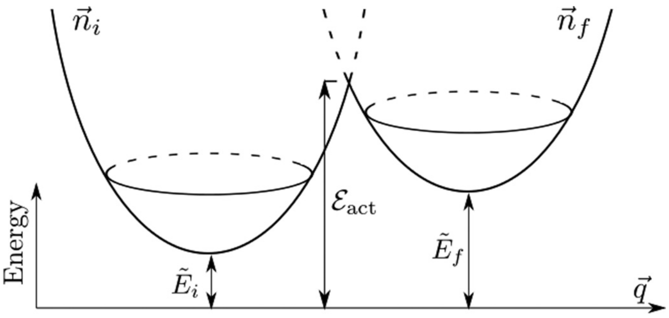


FIG. 2. Energy paraboloids in the $\vec{q}$ space. The paraboloids are indexed by vector $\vec{n}$. Zero-energy level corresponds to the classical ground state. The activation energy $\mathcal{E}_{\mathrm{act}}$ for transition $\vec{n}_i \to \vec{n}_f$ is given by (30).

The thermodynamic excitation energies for such states are

$$\tilde{E}_k^{(\pm)} = \tfrac{1}{2}\vec{z}_k\,\hat{\mathcal{U}}(\vec{z}_k \pm 2\vec{\gamma}) \tag{21}$$

correspond to variation of the "relaxed" energy due to creation (annihilation) of one electron at grain k. By definition, the inequality $\tilde{E}_k^{(\pm)} \geqslant 0$ should hold for all k, which is ensured by (17). In granular systems it is convenient to introduce

$$\varepsilon_k = \begin{cases} \tilde{E}_k^{(+)}, & \text{for } \vec{z}_k\hat{\mathcal{U}}\vec{\gamma} < 0 \\ -\tilde{E}_k^{(-)}, & \text{for } \vec{z}_k\hat{\mathcal{U}}\vec{\gamma} > 0 \end{cases} \tag{22}$$

that has the meaning of the energy of "charged ground state," counted from the global ground state of the grain. Density of such states is sometimes called the density of ground states (DOGS) in the literature (see [11,24] for more details).

As we will see soon, another useful combination that enters the activation exponent of the conductance between two distant grains l (by agreement "left") and r (by agreement "right") is

$$\varepsilon_{lr} \equiv \tfrac{1}{2}\{|\varepsilon_l - \varepsilon_r| + |\varepsilon_l| + |\varepsilon_r|\}. \tag{23}$$

Note that it has a conventional analog in the standard hopping conductivity theory [2].

In the following, we will assume that electronic transitions proceed very fast, so that the slow variables $\vec{q}$ do not have time to change during the process: the Franck-Condon principle (see more discussion on this topic in Sec. IV A). Thus, we introduce additional Franck-Condon excitation energies

$$\begin{aligned} E_k^{(\pm)}(\vec{q}) &\equiv E_{\mathrm{ch}}(\pm\vec{z}_k,\vec{q}) - E_{\mathrm{ch}}(0,\vec{q}) \\ &= \tilde{E}_k^{(\pm)} + \tfrac{1}{2}\vec{z}_k\hat{U}\hat{K}^{-1}\hat{U}\vec{z}_k \mp \vec{z}_k\hat{U}\vec{q}. \end{aligned} \tag{24}$$

2. *Two-particle excitations*

There are four classes of possible two-particle processes, corresponding to an act of the charge $-e$ transfer from grain l to grain r (see Fig. 3).

(a) ($++$ process): transfer of an electron-type single-particle excitation from l to r: $(\vec{n}_i = \vec{z}_l) \longrightarrow (\vec{n}_f = \vec{z}_r)$;

(b) ($--$ process): transfer of a hole-type single-particle excitation from r to l: $(\vec{n}_i = -\vec{z}_r) \longrightarrow (\vec{n}_f = -\vec{z}_l)$;

(c) ($+-$ process): annihilation of a two-particle excitation, consisting of an electron-type excitation in grain l and a hole-type one in grain r: $(\vec{n}_i = \vec{z}_l - \vec{z}_r) \longrightarrow (\vec{n}_f = 0)$;

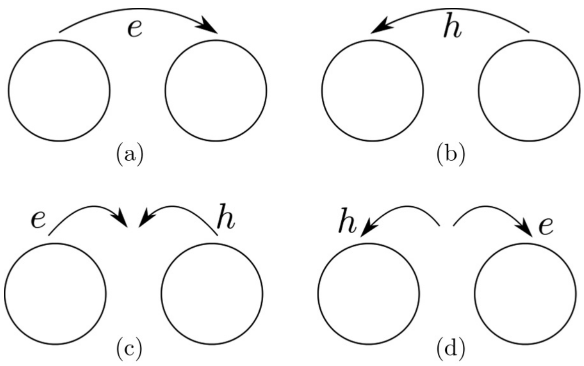


FIG. 3. Four possible ways of the charge $-e$ transfer from grain l to grain r.

(d) ($-+$ process): creation of a two-particle excitation, consisting of an electron-type excitation in grain r and a hole-type one in grain l: $(\vec{n}_i = 0) \longrightarrow (\vec{n}_f = \vec{z}_r - \vec{z}_l)$.

While in the first two processes only single-particle excitations are involved, in the third and fourth ones the two-particle complexes (intergrain electron-hole pairs) are created or annihilated. Because of generally long-range character of interaction matrix $\hat{\mathcal{U}}$ the components of two-particle excitations interact with each other, so that their energies generally are not additive. Thermodynamic excitation energies for the two-particle excitations can be obtained from (20):

$$\tilde{E}_{lr}^{(\pm\pm)} = \tilde{E}_l^{(\pm)} + \tilde{E}_r^{(\pm)}, \tag{25}$$

$$\tilde{E}_{lr}^{(\pm\mp)} = \tilde{E}_l^{(\pm)} + \tilde{E}_r^{(\mp)} - \vec{z}_l \hat{\mathcal{U}} \vec{z}_r \tag{26}$$

and the Franck-Condon energies are

$$\begin{aligned} E_{lr}^{(\pm\pm)}(\vec{q}) &\equiv E_{\text{ch}}(\pm\vec{z}_l \pm \vec{z}_r, \vec{q}) - E_{\text{ch}}(0,\vec{q}) \\ &= E_l^{(\pm)} + E_r^{(\pm)}, \end{aligned} \tag{27}$$

$$\begin{aligned} E_{lr}^{(\pm\mp)}(\vec{q}) &\equiv E_{\text{ch}}(\pm\vec{z}_l \mp \vec{z}_r, \vec{q}) - E_{\text{ch}}(0,\vec{q}) \\ &= E_l^{(\pm)} + E_r^{(\mp)} - \vec{z}_l \hat{U} \hat{K}^{-1} \hat{U} \vec{z}_r. \end{aligned} \tag{28}$$

We will see in Sec. III that the interaction parts in (26) and (28) vanish in the case of short-range interaction.

3. *Potential barrier between resonant grains and activation energy*

A very important role in low-temperature physics is played by the resonant grains, for which either $\tilde{E}_k^{(+)}$ or $\tilde{E}_k^{(-)}$ is anomalously close to zero. The corresponding pairs of states, differing in the charge of the resonant grains, have, therefore, almost identical energies. The transition between such resonant states $\vec{n}_i$ and $\vec{n}_f$ requires, however, a considerable change of the surrounding (i.e., the vector $\vec{q}$), which can only be done continuously. In the course of this change the potential energy of the system also changes: first increases, then decreases, so that the system has to overcome the potential barrier. This can be accomplished either by means of activation over the barrier or by tunneling. For both processes, the height of the barrier W is crucial. To find it we should minimize the energy $E_{\text{ch}}(\vec{n}_i,\vec{q})$ over $\vec{q}$ with additional condition $E_{\text{ch}}(\vec{n}_i,\vec{q}) = E_{\text{ch}}(\vec{n}_f,\vec{q})$, implying that the states are resonant. The result is

$$W = \tfrac{1}{8}(\vec{n}_f - \vec{n}_i)\hat{U}\hat{K}^{-1}\hat{U}(\vec{n}_f - \vec{n}_i). \tag{29}$$

Note that W depends only on the difference $\vec{n}_f - \vec{n}_i$ between the final and initial states. In particular, it is the same for all kinds of processes described in Sec. II C 2.

We can also find the activation energy $\mathcal{E}_{\text{act}}$ for a transition between arbitrary (not necessarily resonant) states $\vec{n}_i$ and $\vec{n}_f$. It can be defined as the lowest point of interception of two paraboloids (see Fig. 2):

$$\mathcal{E}_{\text{act}} = \frac{\tilde{E}_i + \tilde{E}_f}{2} + W + \frac{(\tilde{E}_f - \tilde{E}_i)^2}{16W}, \tag{30}$$

where W is defined by (29). If the states are resonant, we have $\mathcal{E}_{\text{act}} = W$.

III. SHORT-RANGE INTERACTION MODEL

There are some cases when the long-range part of the Coulomb interaction $\hat{U}$ may be neglected:

(1) The systems, where the interaction is screened (say, because of the presence of metallic gate, or due to residual conductivity of the insulating matrix).

(2) The systems with "weak charge disorder," where the concentration of the stray charges is very low (e.g., clean artificial arrays of quantum dots). Here the long-range interaction is not relevant in a wide range of intermediate temperatures. Indeed, in such systems typical values of γ are small and concentration of the resonant grains, proportional to the probability to have $\gamma \approx \pm 1/2$ is low (see [13]): $\tilde{P}(\pm 1/2) \ll 1$. As a result, optimal hops are very long, the width Δ_C of the Coulomb gap is small: $\Delta_C \sim E_c[\tilde{P}(\pm 1/2)]^{1/(d-1)}$. Thus, although the long-range interaction is not suppressed, it becomes essential only at very low temperatures $T \lesssim \Delta_C$. This situation is similar to that for a hopping conductivity of lightly doped semiconductors at low compensation, where (due to the same physical reasons) the crossover from Mott to Efros-Shklovskii law is also shifted to lower temperatures (see [2]).

The above-mentioned systems can be roughly described by the simplest model of "short-range Coulomb interaction" (see, e.g., [15]). In this model we assume the matrices $\hat{U}$ and $\hat{K}$ to be diagonal

$$\hat{U} = \begin{pmatrix} \ddots & 0 & 0 & 0 \\ 0 & U_j & 0 & 0 \\ 0 & 0 & U_{j+1} & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix},$$

$$\hat{K} = \begin{pmatrix} \ddots & 0 & 0 & 0 \\ 0 & K_j & 0 & 0 \\ 0 & 0 & K_{j+1} & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix},$$

but their diagonal entries in general are not identical since different grains have different capacitances, etc. The electrostatic energy (19) can be rewritten in a simple way $E_{\text{ch}} = \sum_j E_j(n_j,q_j)$, where

$$E_j(n,q) = E_j^c \left\{ n^2 + 2n[\gamma_j(1-\alpha_j) - q] + \frac{q^2}{\alpha_j} \right\} \tag{31}$$

and each grain is characterized by three constants:

(i) the standard charging energy $E_j^c \equiv U_j/2 = e^2/2C_j$ with C_j being the capacitance of the grain j;

(ii) the "polaronic coupling constant" $\alpha_j \equiv U_j/K_j$; the inequality $\alpha_j < 1$ is the stability condition;

(iii) the random "effective offset charge" γ_j, distributed in the interval $[-1/2,1/2]$.

The thermodynamic and Franck-Condon excitation energies for this model are

$$\tilde{E}_j^{(\pm)} = E_j^c(1-\alpha_j)(1 \pm 2\gamma_j), \tag{32}$$

$$E_j^{(\pm)}(q) = \tilde{E}_j^{\pm} + 4W_j \mp 2E_j^c q_j, \tag{33}$$

$$W_j = \alpha_j E_j^c/4. \tag{34}$$

For two-particle excitations, in which only charges of two grains l and r are altered, we get $W = W_l + W_r$.

IV. TRANSITION RATE

As we know from [13], the main features of variable-range hopping in granular systems can be revealed already in the simplest model of short-range Coulomb interaction, described in Sec. III, so in this paper we restrict our consideration to this model.

A. Franck-Condon principle

We will also ignore the quantum aspects of the offset charges dynamics (i.e., put $M \to \infty$). In particular, the last assumption means that we can use the Franck-Condon principle in the calculation of transition rates. According to this principle, the set $\vec{q}$ remains unchanged on the cotunneling time scale $\sim E_c^{-1}$. We should calculate the transition rates $w_{lr}(\vec{q}\,)$ between relevant grains at fixed configuration of $\vec{q}$ and only afterwards perform the thermodynamic averaging of the result with respect to $\vec{q}$. The cotunneling time τ_{cotun} can be roughly estimated as the inverse scale of energetic denominators appearing in the perturbation theory

$$\tau_{\text{cotun}} \sim E_c^{-1}. \tag{35}$$

On the other hand, characteristic time scale τ_q for the dynamics of $\vec{q}$ is nothing but the inverse frequency of the impurity oscillations ω_{imp}, which apparently should be of the order of phonon frequencies ω_{ph}. As a result, the applicability criterion for the Franck-Condon principle $\tau_{\text{cotun}} \ll \tau_q$ becomes

$$\omega_{\text{imp}} \ll E_c. \tag{36}$$

Note that in typical granular systems the rough estimate $\sim$100 K holds for E_c as well as for ω_{imp}. Thus, the condition (36) can not be considered automatically fulfilled in all systems; the opposite case $\omega_{\text{imp}} \gg E_c$ is also well probable. This case is, however, not interesting since it implies weak coupling between electronic and vibrational degrees of freedom, leading to a negligible polaronic effect.

As we already noted, we are going to neglect the kinetic term for $\vec{q}$ in the Hamiltonian, which means that $\vec{q}$ can overcome potential barriers only by activation, and not by tunneling. This implies even more stringent condition

$$\omega_{\text{imp}} \ll T \ll E_c \tag{37}$$

that we will consider granted throughout this paper. The low-temperature tunneling case $T \ll \omega_{\text{imp}} \ll E_c$ will be discussed elsewhere.

B. Hamiltonian and some qualitative considerations

Under the above approximations, the Hamiltonian of the system may be written in the form

$$\hat{\mathcal{H}} = \sum_j \{E_j(n_j,q_j) + \hat{\mathcal{H}}_j^{(0)}\} + \sum_{\langle jj'\rangle} \hat{\mathcal{H}}_{\text{tun}}^{(jj')}, \tag{38}$$

where $E_j(n_j,q_j)$ is given by expression (31), and

$$\hat{\mathcal{H}}_j^{(0)} = \sum_{\lambda_j,\sigma} \epsilon_{\lambda_j} \hat{a}_{\lambda_j\sigma}^{\dagger} \hat{a}_{\lambda_j\sigma} \tag{39}$$

is the Hamiltonian of electrons within the jth grain. The index λ_j denotes electronic eigenstates with eigenenergies ϵ_{λ_j}, which are supposed to be σ independent (σ is a spin projection). The level spacing δ_j for electrons at the Fermi level in grain j is small: $\delta_j \ll E_j^c$.

The tunneling Hamiltonian

$$\hat{\mathcal{H}}_{\text{tun}}^{(jj')} = \sum_{\lambda_j,\lambda_{j'},\sigma} t_{\lambda_j\lambda_{j'}} \hat{a}_{\lambda_j\sigma}^{\dagger} \hat{a}_{\lambda_{j'}\sigma} \tag{40}$$

describes the hops of electrons between neighboring grains $\langle jj'\rangle$. We are interested in the case, when the typical dimensionless intergrain conductances

$$g_{jj'} \equiv |t_{jj'}|^2/\delta_j\delta_{j'} \ll 1 \tag{41}$$

are small, and the tunneling Hamiltonian may be treated perturbatively; the necessary order of the perturbation theory, however, appears to be high: the lower the temperature, the higher the order!

The principal idea of any VRH-type calculation is the famous observation of Mott [3]: at low temperatures hopping electrons prefer to visit only resonant sites (in our case for sites stand the grains), where their energies are confined to a narrow strip of width ε near the Fermi energy. Decreasing ε loosens the factor $\exp\{-\varepsilon/T\}$ that suppresses the conductivity due to small number of available excitations. On the other hand, the resonant sites are rare [the typical distance $r(\varepsilon)$ between them grows with decreasing ε], so that the overlap of corresponding wave functions I is small: $I \propto \exp\{-2r(\varepsilon)/\xi\}$ and this small factor becomes still smaller with decreasing ε. Thus, one has to find a compromise between the two exponentially small factors, that results in certain optimal $\varepsilon_{\text{opt}}(T)$ and the conductivity $\sigma \propto \exp\{-\varepsilon_{\text{opt}}(T)/T\}$. In the presence of polaronic effect, the above calculation scenario is somewhat modified, but the main idea remains the same.

While in case of single-electron tunneling evaluation of the overlap exponential factor I is straightforward, and ξ is simply related to the decrement of electronic wave function, in the case of metallic grains the origin of the exponential dependence and explicit form of ξ are more sophisticated. In this section, we will explore this problem, incorporating the additional physics that arises from the effects of the stray-charges flexibility.

C. Amplitudes of multiple cotunneling: Perturbation theory

Thus, we are interested in the amplitude of electronic transition between two distant resonant grains l (left) and r

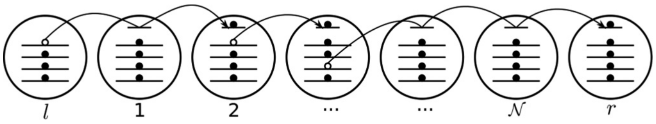


FIG. 4. Cotunneling process from grain l to grain r.

(right). Unless there is a tunneling term in the Hamiltonian, the occupation numbers $\vec{n}$ are preserved. The tunneling term $\hat{\mathcal{H}}_{\text{tun}}$ allows hops of electrons between neighboring grains; to accomplish a hop between distant grains, $\hat{\mathcal{H}}_{\text{tun}}$ has to be applied in some $(\mathcal{N}+1)$-st order of perturbation theory, where $\mathcal{N}$ is a number of intermediate grains, constituting a continuous chain between l and r. A multiparticle process, described by this perturbational approach, is generally known as cotunneling. It was introduced in [14] and applied to the theory of transport in quantum dots [15,25] and granular metals [13,26].

Let us consider the transfer of an electron via virtual states on $\mathcal{N}$ intermediate grains, $l \equiv 0,1,\dots,\mathcal{N},\mathcal{N}+1 \equiv r$ being the path of adjacent grains, starting at the initial grain l and terminating at the final grain r, so that k is a neighbor of $k-1$ and $k+1$ (see Fig. 4). In principle, one should sum over all possible paths connecting l and r, but in the case of small tunneling amplitudes t we can expect the sum to be dominated by shortest paths, i.e., those with minimal possible number $\mathcal{N}$ of intermediate grains. Moreover, in most situations only one particular path will be important. The amplitude of such multiple cotunneling event is given by a proper matrix element of

$$A_{\{h_k,e_k\}}(\vec{q}) = (-i)^{\mathcal{N}+1} \int \mathrm{T}\{\hat{S}\hat{\mathcal{H}}_{\text{tun}}(t_{N+1})\dots\hat{\mathcal{H}}_{\text{tun}}(t_1)\} \times \prod_{k=1}^{\mathcal{N}+1} dt_k \tag{42}$$

calculated at given static set of deviations $\vec{q}$.

The amplitude $A_{\{h_k,e_k\}}$ describes the process, at the end of which a hole with the set of quantum numbers h_0 is created in grain l and an electron with the set $e_{\mathcal{N}+1}$ is created in grain r; generally speaking, each of $\mathcal{N}$ intermediate grains $k = 1,\dots,\mathcal{N}$ acquires an electron-hole pair with quantum numbers $\{e_k,h_k\}$ (inelastic cotunneling). However, it is possible to have $e_k = h_k$ for certain k's, then no electron-hole pairs are created in the corresponding grains (elastic cotunneling). Let us denote the set of such k's as $m_1,m_2,\dots,m_{\mathcal{M}}$. If this set is empty ($\mathcal{M} = 0$), one speaks about purely inelastic multiple cotunneling; if it includes all the intermediate grains ($\mathcal{M} = \mathcal{N}$), we deal with purely elastic multiple cotunneling.

Consider the set of indices $\{\alpha_1\beta_1;\dots;\alpha_{\mathcal{N}+1}\beta_{\mathcal{N}+1}\}$, which describes certain sequence of individual tunnelings between pairs of neighboring grains in the chain (in the kth entry a tunneling event occurs from the state α_k to the state β_k). Such a set is some permutation of "the natural sequence" $\{h_0e_1;h_1e_2;\dots;h_{\mathcal{N}}e_{\mathcal{N}+1}\}$, corresponding to progressive motion of electron from the left end of the chain to its right end. All the permuted sets contribute to the amplitude $A_{\{h_k,e_k\}}$ alongside with the natural one. From (42) we get

$$\tilde{A}_{\{h_k,e_k\}} = (-i)^{\mathcal{N}+1} \prod_{k=1}^{\mathcal{N}+1} t_{e_k h_{k-1}} \int \prod_{k=1}^{\mathcal{N}+1} d\tau_k \langle|\mathrm{T}\{\hat{\psi}^{\dagger}_{e_{\mathcal{N}+1}}(\tau_{\mathcal{N}}+\dots+\tau_1)\hat{\psi}_{h_{\mathcal{N}}}(\tau_{\mathcal{N}}+\dots+\tau_1)\dots\hat{\psi}^{\dagger}_{e_1}(0)\hat{\psi}_{h_0}(0)\}|\rangle, \tag{43}$$

where we have used the shortcut notation $|\,\rangle$ for initial and $\langle\,|$ for final state. The tunneling matrix element $t_{e_k h_{k-1}}$ describes the transition of electron from the state h in the $(k-1)$th grain to the state e in the kth grain.

Within the short-range interaction model we only need to time order the $\hat{\psi}$ operators acting in each single-grain subspace:

$$-i\langle|\mathrm{T}\{\psi_{h_k}(\tau_1+\dots+\tau_k)\psi^{+}_{e_k}(\tau_1+\dots+\tau_{k-1})\}|\rangle. \tag{44}$$

As a result, we obtain (45) for inelastic grains and (46) for elastic ones:

$$-in_{h_k}\big[1-n_{e_k}\big]\big(\theta(\tau_k)e^{-i(\epsilon_{h_k}+E_k^{(+)})\tau_k-i(\tau_1+\dots+\tau_{k-1})(\epsilon_{h_k}-\epsilon_{e_k})} - \theta(-\tau_k)e^{-i(\epsilon_{h_k}-E_k^{(-)})\tau_k-i(\tau_1+\dots+\tau_{k-1})(\epsilon_{h_k}-\epsilon_{e_k})}\big) \tag{45}$$

$$-i\big(1-n_{e_m}\big)\theta(\tau_m)e^{-i(\epsilon_{e_m}+E_k^{(+)})\tau_k} + in_{e_m}\theta(-\tau_m)e^{-i(\epsilon_{e_m}-E_k^{(-)})\tau_k}. \tag{46}$$

Here, n_s is the fermionic occupation number for state s (not to be mixed with the occupation numbers of grains!). The two terms in (45) and (46) correspond to different sequences of tunnelings: in one case, the hole is created first and the electron afterwards, and in another case the order is reversed. Let us now collect all the factors containing τ_k for each $k \in 1\dots\mathcal{N}$ in (43) and integrate them out. Then, we eventually arrive at

$$\tilde{A}_{\{h_k,e_k\}} = -in_{h_0}[1-n_{e_{\mathcal{N}+1}}] \prod_{k=1}^{\mathcal{N}+1} t_{e_k h_{k-1}} \prod_{k=1}^{\mathcal{N}} B_k. \tag{47}$$

The factors B_k for inelastic and elastic grains, respectively, are

$$B_k(\vec{q}) = -n_{h_k}[1-n_{e_k}]\left\{\frac{1}{\epsilon_{h_k}+\Delta_k+E_k^{(+)}(q_k)} + \frac{1}{\epsilon_{h_k}+\Delta_k-E_k^{(-)}(q_k)}\right\},$$

$$B_m(\vec{q}) = -\left\{\frac{1-n_{e_m}}{\epsilon_{e_m}+\Delta_m+E_m^{(+)}(q_m)} + \frac{n_{e_m}}{\epsilon_{e_m}+\Delta_m-E_m^{(-)}(q_m)}\right\}, \tag{48}$$

where we have introduced

$$\Delta_k \equiv -\epsilon_{e_{\mathcal{N}+1}} - E_r^{(+)}(q_r) + \sum_{s=k+1}^{\mathcal{N}} (\epsilon_{h_s} - \epsilon_{e_s}). \tag{49}$$

Note that the explicit summation over all tunneling sequences, which we have just performed, is only possible in the short-range interaction model.

According to Fermi's golden rule, the transition rate is

$$w_{lr}(\vec{q}) = 2\pi \sum_{\{h_k,e_k\}} |\tilde{A}_{\{h_k,e_k\}}|^2 \times \delta\left(\epsilon_{h_0} - \epsilon_{e_{N+1}} + \sum_{s=1}^{\mathcal{N}} (\epsilon_{h_s} - \epsilon_{e_s}) - \Delta\right), \tag{50}$$

where the Franck-Condon energy difference

$$\Delta \equiv \Delta(\vec{q}\,) = E_{\text{ch}}(\vec{n}_f,\vec{q}\,) - E_{\text{ch}}(\vec{n}_i,\vec{q}\,) \tag{51}$$

is the difference between electrostatic energies of final and initial states. We should also define the thermodynamic energy difference

$$\tilde{\Delta} = \tilde{E}_f - \tilde{E}_i \tag{52}$$

which is the energy distance between two local minima corresponding to initial $\vec{n}_i$ and final $\vec{n}_f$ states.

Note that it is impossible to distinguish the final states of the systems with the same sets of inelastic quantum numbers and different sets of elastic quantum numbers. For this reason, in principle, one should add the amplitudes of such processes rather than probabilities. However, the arising interference terms are strongly suppressed because of violent sign fluctuations of tunneling elements and, therefore, they were neglected in (50).

V. AVERAGING OF THE TRANSITION RATE

The transition rate (50) depends on both dynamic random variables $\vec{q}$ and static ones $\vec{\gamma},\vec{\alpha}$, etc. There is an important difference between these two groups of variables: we should perform thermodynamic averaging of (50) over $\vec{q}$, while the frozen static variables do not imply averaging, so that, in principle, they remain "live" and characterize a specific surrounding of a particular chain. However, because of typically large number $\mathcal{N}$ of grains in the chain, the static disorder is partly self-averaged and washed out. As we will see, only the dependence on few static random variables (mainly those characterizing the terminal grains l and r) remains live.

A. Thermodynamic averaging

Gibbs thermodynamic averaging over $\vec{q}$ has the form

$$\langle(\ldots)\rangle_{\vec{q}} = \frac{1}{Z_c} \sum_{\vec{n}} \int (\ldots) \exp\left\{-\frac{E_{\text{ch}}(\vec{n},\vec{q})}{T}\right\} d\vec{q}, \tag{53}$$

where Z_c is the partition function:

$$Z_c = \left[\det \frac{\hat{K}}{2\pi T}\right]^{-1/2} \sum_{\vec{n}} \exp\{-\tilde{E}(\vec{n})/T\}. \tag{54}$$

In averaging over electronic states we will assume an equilibrium noncorrelated distribution, so that $\langle n_{s_k}\rangle = f_F(\epsilon_{s_k})$ and $\langle n_{s_k} n_{s_{k'}}\rangle = \langle n_{s_k}\rangle\langle n_{s_{k'}}\rangle$, with f_F being the Fermi function.

Further simplification can be made if we note that the characteristic value of inelastic energies ϵ_{h_k}, ϵ_{e_k} is controlled by the combination of Fermi functions and delta function in (50). We can see that $\epsilon_{\text{inel}} \sim \Delta/L$, where $L = \mathcal{N} - \mathcal{M}$ is the number of inelastic grains in the string. This means that typically $\epsilon_{\text{inel}} \ll E_c$. The same is also true for Δ_k. Therefore, these quantities in the denominators in (48) can be neglected compared to $E_k^{(\pm)}$. However, this is not true for the elastic energies ϵ_m since they do not enter the delta function. These energies are of order E_c and should be kept in the denominators. For the same reason (as $\epsilon_{h_m} \sim E_c \gg T$), we can substitute $1 - \theta(\epsilon_{h_m})$ instead of $f_F(\epsilon_{h_m})$. The summation over spin projections should be performed only for inelastic grains. It is straightforward and yields the factor 2^{L+1}.

Also, we can substitute $t_{e_k h_{k-1}}$ by its "coarse-grained" value at the Fermi level $t_{k,k-1}$ (i.e., averaged over an interval of energies, large compared to the level spacing δ, but small compared to any other relevant scale). It allows for replacement of the summation over electronic states by integration $\sum_{\epsilon_s} \to \int d\epsilon_s/\delta_k$, which immediately allows to integrate out the elastic energies, and we arrive at

$$\overline{w}_{lr} = 2\pi\left\langle 2^{L+1} \prod_{k=1}^{\mathcal{N}+1} |t_{k,k-1}|^2 \sum_{\{h_l,e_l\}} f_F(\epsilon_{h_0})[1 - f_F(\epsilon_{e_{\mathcal{N}+1}})] \prod_{k,\text{inelastic}} f_F(\epsilon_{h_k})[1 - f_F(\epsilon_{e_k})]\left[\frac{1}{E_k^{(+)}} - \frac{1}{E_k^{(-)}}\right]^2 \times \prod_{m,\text{elastic}} \left[\frac{1}{E_m^{(+)}} + \frac{1}{E_m^{(-)}}\right] \delta\left(\epsilon_{h_0} - \epsilon_{e_{\mathcal{N}+1}} + \sum_{s=1}^{\mathcal{N}} (\epsilon_{h_s} - \epsilon_{e_s}) - \Delta\right)\right\rangle_{\vec{q}}. \tag{55}$$

The variables q_k for intermediate grains enter $\overline{w}_{lr}$ only through the denominators in square brackets of both types (elastic and inelastic), and the result of integration over dq_k in thermal averaging formally diverges at resonances, where either $E_k^{(-)}$ or $E_k^{(+)}$ goes to zero. This divergency, however, should be cut off at $E_k^{\pm} \sim |t|$, where the perturbation theory ceases to be valid, thus, the integration in the vicinity of the resonance gives finite result. So, each integration over dq_k gives three contributions: one from the vicinity of thermal equilibrium $q_k \approx 0$, and the other two from the vicinities of the resonances $q_k^{(\pm)}$, defined by conditions $E_k^{(\pm)}(q_k^{(\pm)}) \approx 0$. The relative magnitudes of the resonant contributions are exponentially suppressed $\sim E_c/|t| \exp\{-E_c/T\}$ and thus are negligible in the low-temperature VRH regime. As a result, thermodynamic fluctuations of q at intermediate grains are not relevant, and one can simply neglect them, setting $q_k \approx 0$, so that $E_k^{(\pm)}(q_k) \to E_k^{(\pm)}(0)$ in (55).

B. Self-averaging of intermediate grains

To treat the products of large number of random factors, occurring in (55), we will apply the Cental Limit Theorem self-averaging rule in the form

$$\prod_{j=1}^{n} X_j \approx \exp\{n\,\overline{\ln X}\}, \quad n \gg 1 \tag{56}$$

and hence

$$|t_{kk-1}|^2 \to t^2 \equiv e^{\overline{\ln |t_{kk-1}|^2}}, \quad \delta_k \to \delta \equiv e^{\overline{\ln \delta_k}},$$
$$E_k^{(c)} \to E_c \equiv e^{\overline{\ln E_k^{(c)}}},$$
$$\left(\frac{1}{E_m^{(+)}(0)} + \frac{1}{E_m^{(-)}(0)}\right) \to \frac{A_1}{E_c}, \tag{57}$$
$$\left(\frac{1}{E_k^{(+)}(0)} - \frac{1}{E_k^{(-)}(0)}\right)^2 \to \frac{A_2}{E_c^2}, \tag{58}$$

where

$$A_1 \equiv 2\exp\{-\overline{\ln[1-4(1-\alpha)^2\gamma^2]}\},$$
$$A_2 \equiv 4A_1^2\exp\{\overline{\ln[(1-\alpha)\gamma]}\}. \tag{59}$$

The factors A_1, A_2 depend on the specific distribution functions of random parameters α and γ and their correlation. They are model-dependent numbers of order unity and we will never refer to their exact values in this paper.

C. Back to thermodynamic averaging

Finally, we have to perform the remaining summation over the energies $\epsilon_{e_s}, \epsilon_{h_s}$ of the components of electron-hole pairs, created in the inelastic grains, and over the vector $\vec{q}$.

Let us introduce the new temperature scale

$$T_* = \sqrt{(A_1/2A_2)E_c\delta} \tag{60}$$

and more convenient notations $\epsilon_s = \epsilon_{e_s}$, $\epsilon_{\mathcal{N}+s} = -\epsilon_{h_s}$ for all "inelastic" intermediate grains, and $\epsilon_{2\mathcal{N}+1} = \epsilon_{e_{\mathcal{N}+1}}$, $\epsilon_{2\mathcal{N}+2} = -\epsilon_{h_0}$ for terminal grains. Then, we can rewrite (55) as

$$\overline{w}_{lr} = 4\pi g T_* \left(\frac{A_1 g\delta}{E_c}\right)^{\mathcal{N}} \sum_{L=0}^{\mathcal{N}} C_{\mathcal{N}}^L \langle I_L[\Delta(q_l,q_r)]\rangle_{\vec{q}}, \tag{61}$$

where

$$I_L(\Delta) = T_* \int_{-\infty}^{\infty} \prod_{s=1}^{2L+2} \frac{d\epsilon_s}{T_*}[1 - f_F(\epsilon_s)]\delta\left(\sum_{s=1}^{2L+2} \epsilon_s + \Delta\right). \tag{62}$$

The binomial coefficient $C_{\mathcal{N}}^L \equiv \mathcal{N}![(\mathcal{N}-L)!L!]^{-1}$ appears in the formula (61) as a number of possible partitions of the string into elastic and inelastic subsets. Note that in (62) we have used the relation (41) and expressed the result in the terms of the average dimensionless conductance between adjacent grains $g \equiv (|t|/\delta)^2 \ll 1$.

As it was shown in [13], in the absence of the polaronic effect (for $W \to 0$) the characteristic scale ϵ_{inel} for the energies of electron-hole pairs, created in the acts of inelastic cotunneling, is much larger than temperature. It allows for an evaluation of the multiple integral in (62), leading to the result

$$I_L(\Delta) = \frac{(|\Delta|/T_*)^{2L+1}}{(2L+1)!}\exp\left\{-\frac{\Delta + |\Delta|}{2T}\right\}, \quad \frac{\Delta}{L} \gg T. \tag{63}$$

A physical meaning of this result is clear: if the number of inelastic grains is L and their total energy is Δ, then the characteristic energy of one electron (or one hole) created in the process is $\epsilon_{\text{inel}} \sim \Delta/L$. The phase volume, corresponding to the processes with $2L+2$ particles with energies $\epsilon \sim \epsilon_{\text{inel}}$ and finite density of states, is then proportional to $\epsilon_{\text{inel}}^{2L+2}$, which (with the help of Stirling formula) explains formula (63).

We will see soon that in the presence of the polaronic effect (namely, for $\Delta \lesssim W$) ϵ_{inel} becomes comparable to temperature. Finding $I_L(\Delta)$ for $\epsilon_{\text{inel}} \lesssim T$ is a much more difficult problem, which, however, is possible to resolve, using a trick proposed in [27].

We should Fourier transform the δ function in (62) to decouple the integrals over ϵ_s:

$$\frac{1}{2\pi}\int e^{-i\Delta t}dt \prod_{s=1}^{2L+2}\int \frac{e^{-it\epsilon_s}}{1+e^{-\epsilon_s/T}}d\epsilon_s. \tag{64}$$

The trick is to multiply it by $1 = \exp(-\frac{\Delta}{2T} - \frac{1}{2T}\sum\epsilon_s)$, which will make the integrals convergent, and they can be easily calculated via residue theory:

$$\frac{1}{2\pi}e^{-\Delta/2T}\int e^{-i\Delta t}dt \prod_{s=1}^{2L+2}\int \frac{e^{-it\epsilon_s}}{2\cosh\frac{\epsilon_s}{2T}}d\epsilon_s$$
$$= \frac{1}{2\pi}e^{-\Delta/2T}\int e^{-i\Delta t}dt\left[\frac{\pi T}{\cosh \pi t T}\right]^{2L+2}. \tag{65}$$

As a result,

$$I_L(\Delta) = \frac{T_*}{2\pi}e^{-\Delta/2T}\int e^{-i\Delta t}dt\left(\frac{\pi T/T_*}{\cosh \pi t T}\right)^{2L+2}. \tag{66}$$

The integral (66) can be evaluated exactly, but it is more convenient first to perform summation over L, which in this representation turns out to be trivial:

$$\sum_{L=0}^{\mathcal{N}} C_{\mathcal{N}}^L I_L(\Delta) = \frac{T_*}{2\pi}e^{-\Delta/2T}\int e^{-i\Delta t}dt\left[1 + \left(\frac{\pi T/T_*}{\cosh \pi t T}\right)^2\right]^{2\mathcal{N}}$$
$$\times\left(\frac{\pi T/T_*}{\cosh \pi t T}\right)^2. \tag{67}$$

We are left with the thermodynamic averaging $\langle \exp\{-\Delta(1/2T + it)\}\rangle_{\vec{q}}$, which is reduced to the Gaussian integration that can be easily performed. As a result,

$$\overline{w}_{lr} \propto \left(\frac{A_1 g\delta}{E_c}\right)^{\mathcal{N}} \exp\left\{-\frac{\varepsilon}{T} - \frac{D^2}{16WT}\right\}$$
$$\times \int \exp\left\{-\frac{W}{T}\left(2tT - i - i\frac{D}{4W}\right)^2\right\}$$
$$\times\left[1 + \left(\frac{\pi T/T_*}{\cosh \pi t T}\right)^2\right]^{N} \frac{dt}{\cosh^2 \pi t T}, \tag{68}$$

where

$$D \equiv D_{lr} = |\tilde{\Delta}_{lr}| - 4W, \quad W \equiv W_l + W_r,$$
$$\varepsilon \equiv \varepsilon_{lr} \equiv \tfrac{1}{2}\{|\varepsilon_l - \varepsilon_r| + |\varepsilon_l| + |\varepsilon_r|\}, \quad \tilde{\Delta}_{lr} = \varepsilon_l - \varepsilon_r,$$

and we have omitted all preexponential factors.

VI. CASE STUDIES: TRANSITION RATES AT DIFFERENT STRENGTH OF POLARONIC EFFECT

The general formula (68), in principle, contains the answers for all possible questions concerning different modes of the charge transfer between two particular grains. However, for understanding of the physical origin of each particular mode and the corresponding T dependencies, it is necessary to consider the limiting cases. In this section, we undertake such a case study.

To evaluate the integral over t one can use the steepest descent method. The saddle point is located below the lowest pole on the imaginary axis of the complex t plane at $t = i(1-\xi)/2T$, where ξ satisfies the equation

$$D + 4W\xi = 2\pi\mathcal{N}T y(\xi)\cot(\pi\xi/2) \tag{69}$$

and

$$y(\xi) \equiv \left[1 + \left(\frac{T_* \sin(\pi\xi/2)}{\pi T}\right)^2\right]^{-1} = \frac{\overline{L}}{\mathcal{N}} \tag{70}$$

determines the typical number $\overline{L}$ of inelastic grains in the string. In terms of ξ, the transition rate can be written as

$$\overline{w}_{lr} \propto \left(\frac{A_1 g\delta}{E_c}\right)^{\mathcal{N}} \left[1 + \left(\frac{\pi T}{T_* \sin(\pi\xi/2)}\right)^2\right]^{\mathcal{N}} \times \exp\left\{\frac{W}{T}\xi^2 - \frac{\varepsilon}{T} + \frac{D}{2T}\xi\right\}. \tag{71}$$

We have omitted in (71) the preexponential factor $\sin^{-2}\pi\xi$ because it becomes essential only at extremely low temperatures in the case of purely elastic cotunneling. This factor may be important for systems where the length of strings $\mathcal{N}$ is restricted from above (as in small arrays of quantum dots, or single-electron transistors), which is not the case as long the VRH conductivity of a large sample is concerned.

We will see that the entire range of the parameter D $(-4W < D < \infty)$ may be split into three intervals with different types of approximations applicable:

(1) weak polaronic effect ($D > 0$, $D \gg \Delta D$);
(2) strong polaronic effect ($D < 0$, $|D| \gg \Delta D$);
(3) narrow transition range ($|D| \lesssim \Delta D$).

The transition range width

$$\Delta D = 8\sqrt{W\overline{L}T} \lesssim 8\sqrt{W\mathcal{N}T} \ll 4W. \tag{72}$$

The latter inequality is valid because, as we will see in Sec. VIII, $\max\{W,|\tilde{\Delta}|\}/T \gg \mathcal{N}$ for typical strings of grains that contribute to the conductivity. Below we discuss these three intervals separately.

A. Weak polaronic effect: "Electron hopping"

For $D > 0$ the parameter $\xi \ll 1$, so that all trigonometric functions in (69) and (70) can be expanded. Aside from that, as long as $D \gg \Delta D$ one can neglect the second term on the left-hand side of (69), so that the latter is reduced to the cubic equation

$$(T_*/2T)^2\xi^3 + \xi = \frac{4\mathcal{N}T}{D} \tag{73}$$

and one can easily express ξ via y:

$$\xi = \frac{2T}{T_*}\sqrt{(1-y)/y}. \tag{74}$$

The transition rate then can be rewritten as

$$\overline{w}_{lr} \propto \left(\frac{A_1 g\delta}{E_c(1-y)}\right)^{\mathcal{N}} e^{\frac{D}{2T}\xi(y)} \exp\left\{-\frac{\varepsilon}{T}\right\}. \tag{75}$$

Substituting (74) to (73), we arrive at the equation

$$y^3 = (1-y)z^2, \quad \text{where} \quad z = \frac{D}{2T_* N} \tag{76}$$

which implicitly determines the function $y(z)$. Finally, for the transition rate we obtain

$$\overline{w}_{lr} \sim \left[\frac{A_1 g\delta}{E_c}\exp\left\{\Theta_1\left(\frac{D}{2T_*\mathcal{N}}\right)\right\}\right]^{\mathcal{N}} \exp\left\{-\frac{\varepsilon}{T}\right\} \tag{77}$$

the function $\Theta_1(z)$ being defined as

$$\Theta_1(z) = 2y(z) - \ln[1 - y(z)]. \tag{78}$$

The asymptotics of $\Theta_1(z)$ are

$$\Theta_1(z) \approx \begin{cases} 3z^{2/3}, & \text{for } z \ll 1 \\ 2\ln(ze), & \text{for } z \gg 1. \end{cases} \tag{79}$$

Note that $\Theta_1(z)$ coincides with $\varphi(z)$ which was described in [13], so that for $D > 0$ their result coincides with the result of this paper up to a slight modification $|\tilde{\Delta}| \to D$ in the definition of the argument z.

One can easily check that the parameter ξ is indeed small under conditions $D > 0$, $\max\{W,|\tilde{\Delta}|\} \gg \overline{L}T$, no matter whether $T < T_*$ or $T > T_*$.

B. Strong polaronic effect: "Polaron hopping"

For negative and not very small D we can neglect the right-hand side of Eq. (69) (it can definitely be done if $|D| \gg \delta D$), and then

$$\xi = |D|/4W, \tag{80}$$

where ξ is not necessarily small. Substituting (80) into (71), we get

$$\overline{w}_{lr} \propto \left[\frac{A_1 g\delta}{E_c}\exp\left\{\Theta_2\left(\frac{T_*}{\pi T}\sin\frac{\pi|D|}{8W}\right)\right\}\right]^{\mathcal{N}} \times \exp\left\{-\frac{\varepsilon}{T} - \frac{D^2}{16WT}\right\} \tag{81}$$

and also

$$y = \frac{1}{1+z^2}, \qquad z = \frac{T_*}{\pi T}\sin\frac{\pi|D|}{8W}. \tag{82}$$

In order to write the transition rate in a compact form, like (77), we have introduced the new function $\Theta_2(z)$:

$$\Theta_2(z) = \ln(1 + 1/z^2). \tag{83}$$

C. Narrow transition range

Looking at the results of the two preceding subsections, we conclude that $\xi \ll 1$ for all positive D and also for small negative D (such that $|D| \ll 4W$). However, for very small $|D|$ (both positive and negative) the first term on the left-hand side of (69) can not be neglected. The condition of “very small” $|D|$ reads as $|D| \sim \xi W$, where ξ is given by the solution of (73). It yields $|D| \sim \Delta D$, where ΔD is given by (72). To find ξ in the narrow transition range $|D| \lesssim \Delta D$, one would have to solve a quartic equation

$$[(T_*/2T)^2\xi^3 + \xi]\left[1 + \frac{4\xi W}{D}\right] = \frac{4\mathcal{N}T}{D}. \tag{84}$$

To justify the numerical coefficients in (72), we note that the saddle-point equation (69) for small ξ can be written in the form

$$D + 4W\xi = \frac{4\overline{L}T}{\xi}, \tag{85}$$

where $\overline{L} \equiv y(\xi)\mathcal{N}$ itself depends on ξ. A formal “solution” of this equation is

$$\xi = -\frac{D}{8W}\left(1 \pm \sqrt{1 + \frac{64W\overline{L}T}{D^2}}\right). \tag{86}$$

From the result (86) immediately follows the estimate for the width of transition range $\Delta D \sim 8\sqrt{W\overline{L}T}$. Since the requirement $\overline{L} \leqslant \mathcal{N}$ is always satisfied, we arrive at the inequality in (72).

We will not discuss the transition region in detail since the corresponding range of random parameters D, W is narrow and does not give any considerable contribution to physical observables.

VII. PHYSICAL INTERPRETATION

Activation exponential factors in formulas (77) and (81) coincide with the corresponding factors from standard polaron hopping theory. Additional modifications, specific for our problem, arise only in the power-law factors (effective overlap integrals) due to the many-particle nature of the cotunneling process.

A. Main exponential dependence

Activation exponential factors for conventional polarons were obtained previously by other researchers. Still, we would like to present some physical arguments that qualitatively explain the origin of these factors.

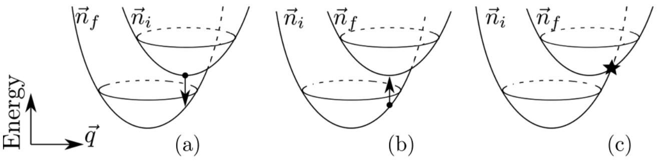


FIG. 5. Different types of transitions: (a) exothermic, (b) endothermic, (c) polaron transition.

Let us first consider the exothermic transition with $\Delta < 0$, i.e., $E_{\text{ch}}(\vec{n}_f,\vec{q}) < E_{\text{ch}}(\vec{n}_i,\vec{q})$ [see Fig. 5(a)]. Since the total electrostatic energy decreases, the electron-hole pairs will be created in the inelastic intermediate grains to ensure energy conservation. Thus, in this situation with exponential accuracy the probability of transition is just the probability to find the system in state $\vec{n}_i$. Its maximum value is

$$\overline{w} \propto \exp\{-\tilde{E}_i/T\}. \tag{87}$$

Now, let us consider the endothermic transition with $\Delta > 0$, i.e., $E_{\text{ch}}(\vec{n}_f,\vec{q}) > E_{\text{ch}}(\vec{n}_i,\vec{q})$ [see Fig. 5(b)]. Since the electrostatic energy increases, the electron-hole pairs will be annihilated in some intermediate grains to make up the shortfall. Such pairs are difficult to find at low temperature, which is accounted for by the additional exponential factor $\exp(-\Delta/T)$. The transition rate also contains $\exp\{-E_{\text{ch}}(\vec{n}_i,\vec{q})/T\}$: the probability to find the system in state $\vec{n}_i$. Together these two contributions yield $\overline{w} \propto \exp\{-E_{\text{ch}}(\vec{n}_f,\vec{q})/T\}$, and we should take its maximum value

$$\overline{w} \propto \exp\{-\tilde{E}_f/T\}. \tag{88}$$

But what if $\vec{q}$, that delivers the minimum of $E_{\text{ch}}(\vec{n}_f,\vec{q})$, violates the requirement $\Delta > 0$? The result (88) will be incorrect in this case since there will be no exponential factor $e^{|\Delta|/T}$ for $\Delta < 0$! In this case, the true minimum of the activation energy should lay on the boundary of the regions with $\Delta > 0$ and $\Delta < 0$, i.e., at $\Delta = 0$ [see Fig. 5(c)]:

$$\overline{w} \propto \exp\{-\min E_{\text{ch}}(\vec{n}_f,\vec{q})/T\}, \quad \vec{q}: \ \Delta = 0. \tag{89}$$

Finding the minimum, we obtain

$$\overline{w} \propto \exp\{-\mathcal{E}_{\text{act}}/T\}, \tag{90}$$

where $\mathcal{E}_{\text{act}}$ is given by (30).

To determine which of the above-described solutions gives the largest contribution to the probability of transition, we should simply compare the corresponding exponents. It is easily seen that the contribution of the boundary minimum (90) dominates when $|\tilde{E}_i - \tilde{E}_f| < 4W$ and explains the main exponential dependence in polaron hopping regime (81). In the opposite case $|\tilde{E}_i - \tilde{E}_f| > 4W$ we have to choose the maximum of (87) and (88), which explains the result $\overline{w}_{lr} \propto \exp\{-\varepsilon_{lr}/T\}$ from (77).

B. Energies of electron-hole pairs

In the electron hopping regime, the Franck-Condon transition energy Δ (that coincides with the aggregate energy of electron-hole pairs involved) does not differ much from its “relaxed” value $\tilde{\Delta}$. Hence,

$$\varepsilon_{\text{inel}} \sim \Delta/\overline{L} \approx \tilde{\Delta}/\overline{L} \sim \mathcal{L}T \gg T, \tag{91}$$

where $\mathcal{L}$ is the large logarithmic factor (see later).

On the contrary, according to the above-described physical picture, in polaron hopping regime the transition occurs at $\Delta = 0$ [see (89)]. However, presented in Sec. VI more careful calculation (which takes into account the Δ dependence of the power-law factor) shows that Δ is indeed small, but finite, in contrast with conventional polaron transitions (where it is exactly zero). In fact, $\Delta \sim \overline{L}T$, which means that in the polaron hopping regime

$$\varepsilon_{\text{inel}} \sim \Delta/\overline{L} \sim T. \tag{92}$$

Another way to see this is to note that in the presence of polaronic effect, the integral (68) converges at $t \lesssim 1/T$, which means that the integrals over the inelastic energies (64) converge solely due to the denominators, which means at $\varepsilon_{\text{inel}} \sim T$.

Thus, we conclude that during the crossover from electron to polaron hopping the characteristic energies of electron-hole pairs decrease from $\varepsilon_{\text{inel}} \sim \mathcal{L}T$ to $\varepsilon_{\text{inel}} \sim T \ll \mathcal{L}T$.

VIII. CONDUCTIVITY: GENERAL CONSIDERATION

We now finally turn to the calculation of conductivity of a sample of granular metal. According to the general philosophy of the hopping, we introduce the Miller-Abrahams network of conductances $g_{lr} \propto \overline{w}_{lr}$, connecting each pair of grains. As usual, at low temperatures the conductivity is dominated by distant pairs of resonant grains, so that we can use the results of the preceding section, and represent the conductances in (almost) standard form

$$g_{lr} \sim \exp\left\{-\frac{\mathcal{E}_{lr}}{T} - \frac{2|\mathbf{r}_l - \mathbf{r}_r|}{a_{lr}}\right\}, \tag{93}$$

where $\mathbf{r}_l$, $\mathbf{r}_r$ are the positions of the centers of the grains and the "effective decay length" a_{lr} is given by

$$\frac{2a_0}{a_{ij}} = \ln\left(\frac{E_c}{A_1 g\delta}\right) - \begin{cases}\Theta_1\left(\frac{D_{lr}}{2T_*\mathcal{N}_{lr}}\right), & D_{lr} > 0\\ \Theta_2\left(\frac{T_*}{\pi T}\sin\frac{\pi|D_{lr}|}{8(W_l+W_r)}\right), & D_{lr} < 0\end{cases} \tag{94}$$

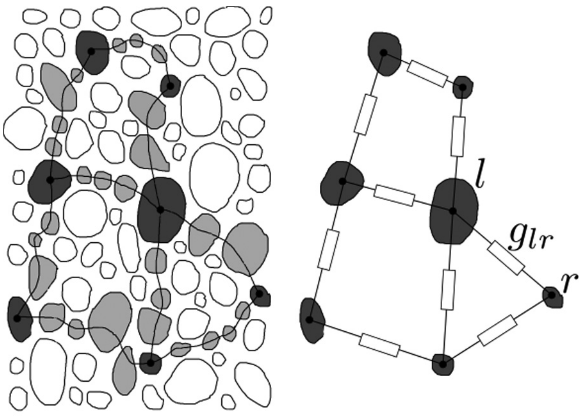
where a_0 is the average distance between neighboring grains, so that $|\mathbf{r}_l - \mathbf{r}_r| \equiv a_0\mathcal{N}_{lr}$. The functions Θ_1 and Θ_2 are given by (79) and (83).

The dependence of a_{lr} on both r_{lr} and characteristics of grains seems to be unusual, however, it is a distinct feature of hopping in granular materials. This dependence is only logarithmic and therefore it can be taken into account perturbatively (see later).

The activation energy is standard for a polaronic problem:

$$\mathcal{E}_{lr} = \varepsilon_{lr} + \begin{cases}0 & (D_{lr} > 0),\\ \frac{D_{lr}^2}{16(W_l+W_r)} & (D_{lr} < 0).\end{cases} \tag{95}$$

Thus, we have come to a system of grains, each of them being characterized by the position $\mathbf{r}_j$ of its center, by the energy ε_j, and by the barrier W_j. Obviously, at low T hopping electrons prefer to "make stops" only at resonant grains, those with small ε and W, while the nonresonant grains with typically large ε and W may serve only as intermediate places, where electrons occur only virtually, staying there for very short times, governed by the quantum uncertainty (see Fig. 6). The random variables ε_j and W_j have the distribution function $\tilde{\nu}(\varepsilon, W)$ and may be correlated or uncorrelated, depending on the underlying physics. In this paper, we will consider only the noncorrelated case, in which for small enough ε



(a) (b)

FIG. 6. (a) Sample of granular material. Resonant grains are shown as dark shapes. Electron cotunneling paths between resonant grains go through gray intermediate grains. (b) Equivalent Miller-Abrahams network of conductances.

$$\tilde{\nu}(\varepsilon, W) = \nu_0 P(W), \tag{96}$$

where $P(W)$ is the barrier distribution function and ν_0 is the density of states (DOGS) near $\varepsilon = 0$:

$$\nu_0 = n_g\tilde{P}(\gamma = \pm 1/2)/E_c, \tag{97}$$

where $\tilde{P}(\gamma = \pm 1/2)$ is the probability density to have $\gamma = 1/2$ (the same as $\gamma = -1/2$), and n_g is the concentration of grains.

Statistics of barriers

Within the paradigm of short-range interaction, it is natural to assume that the interaction between grains and impurities is also a short-range one. It means that only the impurities, situated immediately at the surface of particular grain j, contribute to both ε_j (through the offset charge γ_j) and W_j (through finite elasticity α_j).

In general, the number of such impurities p_j fluctuates from grain to grain. Different impurities give to ε_i contributions of different signs, depending on their charges and positions. Therefore, ε_j would fluctuate from grain to grain already in the hypothetical case, where all p_j are the same.

Because of the random signs of different contributions to ε_j, the density of states $\nu_0(\varepsilon)$ can easily be nonzero at $\varepsilon = 0$. However, it is not the case for the barriers W_j. Each barrier is additive with respect to different impurities

$$W_j = \sum_{k_j=1}^{p_j} W_{k_j}. \tag{98}$$

Since the individual contributions are positively defined, the random charges of different impurities do not matter, and the distribution function $P(W)$ should vanish at $W = 0$. At the same time, for any positive barrier height $P(W)$ should be nonzero: there is no physical reason to expect a hard gap in $P(W)$.

1. *Gaussian distribution of barriers*

The contributions of individual impurities W_{k_j} are assumed to be noncorrelated random variables with identical relatively narrow distributions. Thus, for the most interesting case, in which the average number of impurities coupled to each grain is large $\overline{p}_j \gg 1$, the central part of the distribution $P(W)$ should be a Gaussian one:

$$P(W) = \frac{1}{\sqrt{2\pi}\delta W} \exp\left\{-\frac{(W-\overline{W})^2}{2\delta W^2}\right\}, \tag{99}$$

where $\overline{W}$ is the average barrier, δW is its standard deviation, and by "central part" we mean the range of barriers W, where $|W-\overline{W}| \ll \overline{W}$. We will stick to the case $\overline{p} \gg 1$, when $\overline{W} \propto \overline{p}$ and $\delta W \propto \overline{p}^{1/2}$, so that the distribution is narrow: $\delta W \ll \overline{W}$.

The condition $\overline{p}_j \gg 1$ also allows to neglect the correlations between ε_j and W_j. Note that for $p_j \sim 1$ the correlations are strong: for instance, the absence of impurities, associated with given grain (i.e., $p_j = 0$), leads to $W_j = 0$ and, simultaneously, $\gamma_j = 0$.

Thus, in this paper we are going to analyze the case where the joint density of states (96) has the polaronic factor $P(W)$ defined in (99). We have physically justified such choice, although other distribution functions can also be considered.

2. *Rectangular distribution of barriers*

For example, in the paper [20] a factorized rectangular distribution

$$P(W) = \frac{\theta(W-W_{\text{min}})\theta(W_{\text{max}}-W)}{W_{\text{max}}-W_{\text{min}}} \tag{100}$$

was studied by the method similar to the original percolational approach to the Mott VRH, proposed in [28]. The most spectacular result was obtained for the case $W_{\text{min}} \ll \overline{\varepsilon}(T) \ll W_{\text{max}}$, when for relevant $\varepsilon \sim \overline{\varepsilon}(T)$ within the Mott strip of width $\overline{\varepsilon}(T)$ the"generalized density of states" $\tilde{\nu}(\varepsilon,W) = \tilde{\nu}_0$ may be treated as ε and W independent. Under this condition, the temperature dependence of conductivity is given by

$$\sigma \propto \exp\{-(T_M/T)^{2/(d+2)}\}, \tag{101}$$

$$T_M \sim (\tilde{\nu}_0 a^d)^{-1/2}, \quad \overline{\varepsilon}(T) \sim T^{d/(d+2)} T_M^{2/(d+2)}. \tag{102}$$

For low temperatures, such that $W_{\text{min}} > \overline{\varepsilon}(T)$, the low-energy gap in the distribution (100) becomes essential and the conductivity acquires a hard gap as well:

$$\sigma \propto \exp\{-W_{\text{min}}/T\}, \quad T \ll W_{\text{min}}\left(\frac{W_{\text{min}}}{T_M}\right)^{\frac{1}{d+1}}. \tag{103}$$

As we have already noted, in reality the gap in the distribution $P(W)$ is not hard, it is most likely to have an exponential tail at zero. In the rest of this paper, we will adopt the Gaussian distribution of barriers (99).

A. Percolation problem

The conductivity of Miller-Abrahams network with exponentially wide distribution of conductances should be found by means of percolation theory [2,28]. We have come to a kind of colored percolation problem on random sites, homogeneously distributed in $(\mathbf{r},\epsilon)$ space with the density (DOS) ν_0 [29]. Each site k, besides its position $(\mathbf{r}_k,\epsilon_k)$, is characterized by a positive variable W_k, distributed according to (99). The variables ϵ_k and W_k are interpreted as two components of a composite color of a site k, and $n(\epsilon,W) \equiv \nu(\epsilon)P(W)$ is the density of sites with given color.

By definition, a pair of sites $\langle kj\rangle$ is "ξ connected" if

$$\xi_{kj} \equiv \frac{2|\mathbf{r}_k-\mathbf{r}_j|}{a} + \frac{E_{kj}}{T} < \xi, \tag{104}$$

where

$$E_{kj} \equiv E(\varepsilon_k W_k|\varepsilon_j W_j) = \varepsilon_{kj} + \Lambda_{kj}, \tag{105}$$

$$\Lambda_{kj} \equiv \frac{D_{kj}^2}{16(W_k+W_j)}\theta(-D_{kj}), \tag{106}$$

$$D_{kj} \equiv |\varepsilon_k-\varepsilon_j| - 4(W_k+W_j). \tag{107}$$

As it is usual for the problems of VRH type, the dc charge transfer processes at low temperatures are confined to certain critical subnetwork of resonant grains with small energies ε within certain narrow strip of width $\overline{\varepsilon}(T)$. This width is T dependent and has to be defined self-consistently. It is more convenient to proceed with calculations in different temperature ranges separately.

IX. CONDUCTIVITY IN DIFFERENT TEMPERATURE RANGES

There are three principal energy scales in our problem: (i) the average barrier $\overline{W}$, (ii) the barrier dispersion $\delta W \ll \overline{W}$, and (iii) the width of the Mott strip $\overline{\varepsilon}(T)$. Correspondingly, there are three ranges of temperature with different dominating physics.

A. Electron hopping: Standard Mott VRH

In this case $\overline{W} \ll \overline{\varepsilon}(T)$, so that the polaron effect is negligible, and the standard Mott law is valid

$$\sigma \propto \exp\left\{-\left(\frac{T_M}{T}\right)^{\frac{1}{d+1}}\right\}, \qquad T_M = \frac{\beta_M}{\nu_0 a^d}. \tag{108}$$

The width of effective energy strip is

$$\overline{\varepsilon}(T) \sim T(T_M/T)^{\frac{1}{d+1}}. \tag{109}$$

The condition $\overline{W} \ll \overline{\varepsilon}(T)$ is equivalent to

$$T \gg T_{c1}, \qquad T_{c1} \sim \overline{W}\left(\frac{\overline{W}}{T_M}\right)^{\frac{1}{d}} \tag{110}$$

being the temperature of crossover between the electron and the polaron regimes.

B. Polaron hopping, grains with typical barriers

Here,

$$\delta W \ll \overline{\varepsilon}(T) \ll \overline{W} \tag{111}$$

so that the polaron effect is dominant, but the fluctuations of the barriers are still negligible. In this range, the results of the model with identical barriers for all grains are applicable:

$$\sigma \propto \exp\left\{-\frac{2\overline{W}}{T} - \left(\frac{T_M'}{T}\right)^{\frac{1}{d+1}}\right\},$$

$$T_M' = \frac{\beta_M'}{\nu_F a^d}, \qquad \overline{\varepsilon}(T) \sim T\left(\frac{T_M'}{T}\right)^{\frac{1}{d+1}}. \tag{112}$$

Note that the second term in the exponent is small compared to the first one. The condition $\delta W \ll \overline{\varepsilon}(T) \ll \overline{W}$ is equivalent to $T_{c2} \ll T \ll T_{c1}$, where

$$T_{c2} \sim \delta W \left(\frac{\delta W}{T'_M}\right)^{\frac{1}{d}} \sim T_{c1}\left(\frac{\delta W}{\overline{W}}\right)^{\frac{d+1}{d}} \ll T_{c1}. \tag{113}$$

In principle, one can also write a general formula that describes the behavior of the conductivity in the entire range $T \gg T_{c2}$:

$$\sigma \propto \exp\left\{-\frac{W_{\text{eff}}(T)}{T}\right\}, \quad W_{\text{eff}}(T) = \overline{W}[2 + F_1(T/T_{c1})], \tag{114}$$

$F_1(x)$ being a universal function, with a shape depending only on the space dimensionality d. Its asymptotics are

$$F_1(x) \approx x^{\frac{d}{d+1}} \begin{cases} c_1 & (x \ll 1), \\ c_2 & (x \gg 1), \end{cases} \tag{115}$$

where c_1, c_2 are some universal constants. An interpolation formula for the function $F_1(x)$ was proposed in [20].

C. Polaron hopping, grains with the barriers in the Gaussian tail of distribution

Here,

$$\overline{\varepsilon}(T) \ll \delta W \ll \overline{W} \tag{116}$$

and the grains with anomalously low barriers W_j from the tail of the distribution (99) dominate the critical network:

$$\sigma \propto \exp\left\{-\frac{2\overline{W}}{T} + \frac{\delta W}{T}\left[8d\,\ln\left(\frac{T_{c2}}{T}\right)\right]^{1/2}\right\},$$
$$\overline{\varepsilon}(T) \sim \delta W\left[8d\,\ln\left(\frac{T_{c2}}{T}\right)\right]^{-1/2}. \tag{117}$$

Namely, the critical subnetwork consists of grains with W_j from a narrow strip

$$\left|W_j - \frac{1}{2}W_{\text{eff}}(T)\right| \lesssim \overline{\varepsilon}(T),$$
$$W_{\text{eff}}(T) \equiv 2\overline{W} - \delta W\left[8d\,\ln\left(\frac{T_{c2}}{T}\right)\right]^{1/2}. \tag{118}$$

The derivation of results (117) and (118) is given in Appendix B. It is based on the possibility to reduce the initial multiparametrical percolation problem to certain universal one, governed by a single parameter.

Note that the result (117) is only valid, if the second term in the exponent is small compared to the first one. In the terms of temperature it means

$$T_{\min} \ll T \ll T_{c2}, \qquad T_{\min} \sim T_{c2}\exp\left\{-\frac{1}{8d}\left(\frac{\overline{W}}{\delta W}\right)^2\right\}. \tag{119}$$

In principle, one can also combine the results in the entire range $T_{\min} \ll T \ll T_{c1}$ in one formula:

$$\sigma \propto \exp\left\{-\frac{W_{\text{eff}}(T)}{T}\right\}, \quad W_{\text{eff}}(T) = 2\overline{W} + \delta W\,F_2(T/T_{c2}), \tag{120}$$

where $F_2(x)$ is, again, a universal function, depending only on the space dimensionality d. Its asymptotics are

$$F_2(x) \approx \begin{cases} -\sqrt{8d\,\ln(1/x)} & (x \ll 1), \\ c_3 x^{\frac{d}{d+1}} & (x \gg 1), \end{cases} \tag{121}$$

where c_3 is a constant. Note that the high-temperature asymptotics of (120) coincides with the low-temperature asymptotics of (114).

The inequality $T \gg T_{\min}$ was imposed to secure the condition of relatively small relevant fluctuations with $W - \overline{W} \ll \overline{W}$. This condition is needed to justify their universal Gaussian distribution. At $T \sim T_{\min}$, the relevant fluctuations are large: $W - \overline{W} \sim \overline{W}$ and their distribution is strongly model dependent.

The range of exponentially low temperatures $T \lesssim T_{\min}$ may only be of academic interest because at such low T, the the barrier would rather be penetrated due to quantum tunneling mechanism than due to the activational one, which we are discussing here.

D. Effective radius of the wave functions

Now we have to specify the "effective radius of the wave function" a, entering the criterion (104). In the electron hopping regime, when $|\varepsilon| \gg W$, the actual "radius" $a = a_{kj}$ depends on the characteristics of particular grains k and j. This logarithmic dependence can be easily taken into account. This was done in [13] with the help of the perturbational method in the percolation theory. As a result, the effective radius a appeared to be inversely proportional to the large logarithm $\mathcal{L}(T)$:

$$a = a(T) = \frac{2a_0}{\mathcal{L}(T)}, \tag{122}$$

where we define a_0 as the mean distance between the grains: $r_{kj} = a_0\mathcal{N}_{kj}$. As a consequence, the parameter T_M in (108) acquires a logarithmic temperature dependence

$$T_M = T_M(T) = \frac{\beta'}{\nu_F a(T)^d} = \frac{\beta}{\nu_F}\left(\frac{\mathcal{L}(T)}{2a_0}\right)^d. \tag{123}$$

Fortunately, for $|\varepsilon| \ll W$, in all the polaron hopping regimes, including both weak fluctuations and strong fluctuations cases, a does not depend on i and j, but only on temperature. This can be seen from the expression for transition rate, which contains the factor

$$e^{-2r_{kj}/a} \equiv \left\{\frac{A_1 g\delta}{E_c}\left[1 + \left(\frac{\pi T}{T_*\sin\frac{\pi|D|}{8W}}\right)^2\right]\right\}^{\mathcal{N}_{kj}}$$
$$\approx \left\{\frac{A_1 g\delta}{E_c}\left[1 + \left(\frac{\pi T}{T_*}\right)^2\right]\right\}^{\mathcal{N}_{kj}}. \tag{124}$$

Thus, no perturbational method is needed in this case and

$$a = a'(T) = \frac{2a_0}{\ln\left(\frac{E_c}{A_1 g\delta}\right) - \ln\left[1 + \left(\frac{\pi T}{T_*}\right)^2\right]} \tag{125}$$

so that the parameter T'_M in (112) also becomes logarithmically T dependent:

$$T'_M = T'_M(T) = T'_M(0)\left(1 - \frac{\ln\left[1 + \left(\frac{\pi T}{T_*}\right)^2\right]}{\ln\left(\frac{E_c}{A_1 g\delta}\right)}\right)^d,$$

$$T'_M(0) = \frac{\beta'}{\nu_F a'(0)^d} = \frac{\beta'}{\nu_F}\left(\frac{\ln\left(\frac{E_c}{A_1 g\delta}\right)}{2a_0}\right)^d. \tag{126}$$

X. SUMMARY

We have considered a granular metal with charged impurities (stray charges) that are not rigidly fixed at certain positions, but can be slightly displaced from their equilibrium positions due to electrostatic forces and thermal fluctuations. The flexibility of the system of impurities leads to a "polaronic effect" that can be either weak (for $T \gg T_{c1}$) or strong (for $T \ll T_{c1}$), where the crossover temperature

$$T_{c1} \sim \overline{W}(\overline{W}/T_M)^{\frac{1}{d}} \tag{127}$$

and $T_M \approx T_M(T_{c1})$ is given by (134) below.

In this paper, we have studied a model with short-range interaction (e.g., screened Coulomb) and taken into account only the thermoactivational mechanism of the penetration of the polaronic barrier by the configurational degrees of freedom, not the tunneling one. The latter may become relevant at very low temperatures (the larger the effective mass M of the impurities, the lower the corresponding crossover temperature).

Experimentalists often introduce the differential activation energy according to

$$E_{\text{act}}(T) \equiv \frac{\partial \ln(1/\sigma(T))}{\partial(1/T)}. \tag{128}$$

Below we discuss the temperature dependence of $E_{\text{act}}(T)$ in different temperature ranges.

A. Weak polaronic effect

In the range $T \gg T_{c1}$, the hopping can be with minor modifications described by the electronic multiple cotunneling [13]

$$E_{\text{act}}(T) \approx \frac{T}{d+1}\left(\frac{T_M(T)}{T}\right)^{\frac{1}{d+1}}. \tag{129}$$

B. Strong polaronic effect

The range $T \ll T_{c1}$, where the cotunneling is dominated by polaronic effect, is split into two subranges with the crossover temperature

$$T_{c2} \sim T_{c1}(\delta W/\overline{W})^{\frac{d+1}{d}} \ll T_{c1}. \tag{130}$$

(i) For $T_{c2} \ll T \ll T_{c1}$, the spatial fluctuations of polaronic barriers W_j can be neglected and the choice of resonant grains that constitute the effectively conducting network is dictated exclusively by the values of ε_j: they should lie in the Mott-type strip $|\varepsilon_j| \lesssim \overline{\varepsilon}(T)$, while the particular values of W_j are irrelevant. Here the approach, similar to the one proposed in [20], is applicable, and

$$E_{\text{act}}(T) \approx 2\overline{W} + \frac{T}{d+1}\left(\frac{T'_M(T)}{T}\right)^{\frac{1}{d+1}}, \tag{131}$$

where the second term is relatively small, compared to the first one.

(ii) For $T \ll T_{c2}$, both ε_j and W_j are important, the effective network is formed by the grains for which both ε_j and W_j are anomalously small: $|\varepsilon_j| \lesssim \overline{\varepsilon}(T)$ and simultaneously $|W_j - \frac{1}{2}W_{\text{eff}}(T)| \lesssim \overline{\varepsilon}(T)$. Here,

$$\overline{\varepsilon}(T) \sim \delta W\left[8d\,\ln\left(\frac{T_{c2}}{T}\right)\right]^{-1/2}. \tag{132}$$

In this temperature range,

$$E_{\text{act}}(T) \approx W_{\text{eff}} \equiv 2\overline{W} - \delta W\left[8d\,\ln\left(\frac{T_{c2}}{T}\right)\right]^{1/2}. \tag{133}$$

Thus, due to the tail in the distribution of the barrier fluctuations, the effective activation energy continues to decrease even at lowest temperatures, although this decrease becomes very slow.

C. Elastic versus inelastic cotunneling

The above-mentioned crossovers discriminate different modes of hopping with respect to the strength of polaronic effect (crossover at $T \sim T_{c1}$) and to the importance of the fluctuations of the latter (crossover at $T \sim T_{c2}$). There is, however, one additional crossover at $T \sim T_{c0}$ that discriminates different modes with respect to the character of cotunneling: elasic or inelastic. For $T \gg T_{c0}$, the "constants" $T_M(T)$ and $T'_M(T)$ logarithmically increase with the lowering of temperature, while at $T \lesssim T_{c0}$ they saturate and become T independent:

$$T_M(T) = \frac{\beta\mathcal{L}^d(T)}{\nu_F(2a_0)^d}, \quad T'_M = \frac{\beta'\mathcal{L}^d(0)}{\nu_F(2a_0)^d}, \tag{134}$$

$$\mathcal{L}(T) \approx \ln\left(\frac{E_c}{g\delta}\frac{T_{c0}^2}{T_{c0}^2 + T^2}\right). \tag{135}$$

The explicit form of T_{c0} depends on the relation between T_{c0} and T_{c1}:

$$T_{c0} \sim (E_c\delta)^{1/2}\begin{cases} 1/\mathcal{L}(0), & T_{c0} \gg T_{c1} \\ 1/\pi, & T_{c0} \ll T_{c1}. \end{cases} \tag{136}$$

For $T_{c0} \gg T_{c1}$, the crossover between elastic and inelastic cotunneling takes place within the weak polaron effect domain, while for $T_{c0} \ll T_{c1}$ it happens within the strong polaron effect domain.

D. Long-range Coulomb interaction: Some qualitative estimates

If the long-range part of the interaction $U_{jk} \propto 1/|\mathbf{r}_{jk}|$ is not screened, the problem becomes not analytically solvable, but some qualitative conclusions still may be drawn. Although the characteristics of grains are spatially correlated in this case, at low temperature a typical length of a hop exceeds the corresponding correlation length, so that variables ε_j and W_j of the terminal grains in the string still may be treated as independent random variables. The distribution of ε_j has a Coulomb gap $\nu \propto \varepsilon_j^{d-1}$ (see [11,13]), while the distribution of W_j apparently remains Gaussian, with certain $\overline{W}$ and δW. Then, as in the case of short-range interaction, one can define two crossover temperatures $T_{c1}^* \sim \overline{W}^2/T_0$ and $T_{c2}^* \sim \delta W^2/T_0$.

(a) As it was already argued in [13], in the range of weak polaronic effect $T \gg T_{c1}^*$, the Efros-Shklovskii law $\sigma \propto \exp\{-(T_0/T)^{1/2}\}$ should be valid, where $T_0(T)$ is given by (2). This law corresponds to $E_{\text{act}}(T) \sim (T T_0)^{1/2}$.

(b) Within the range $T_{c2}^* \ll T \ll T_{c1}^*$, the temperature dependence of conductivity is mainly determined by the activation over typical polaronic barriers: $E_{\text{act}}(T) \approx 2\overline{W} + \text{const} \times (T T_0)^{1/2}$.

(c) At still lower $T \ll T_{c2}^*$, the hopping proceeds through the grains with anomalously low barriers, and the formula (133) is valid, where T_{c2} should be replaced by T_{c2}^*.

A special interesting question is the role of long-range interaction in the case of a dense two-dimensional array of grains. There is certain spatial scale r_c such that for $|\mathbf{r}_{jk}| < r_c$ the interaction $U_{jk} \propto (1/r_c)\ln(|\mathbf{r}_{jk}|/r_c)$ describes the logarithmic two-dimensional Coulomb law, while the standard three-dimensional Coulomb law $U_{jk} \propto 1/|\mathbf{r}_{jk}|$ is restored only at $|\mathbf{r}_{jk}| \gg r_c$ (see [30] for detailed discussion). Such form of the interaction leads to modification of the Coulomb gap and, therefore, may result in an unusual behavior of transport.

XI. CONCLUSION

In this paper, we have introduced a concept of polaronic effect in granular systems, related to the flexibility of the random charges, trapped in the insulating matrix. We have explained how this effect is manifested in the conductivity of the system, the latter being controlled by multiple cotunneling of electrons through long "strings" of adjacent grains. The basic line of our reasoning was similar to that of every VRH-like calculation, and could be split into two basic steps: (i) We calculate the carrier transition rate between distant resonant sites, taking into account all the necessary physics (the cotunneling and coupling to flexible impurities). The transition rate is given by the expression (68) and its simplifications (77) and (81). (ii) Afterwards, we find the conductivity of Miller-Abrahams network of conductances with the help of percolation theory. We distinguish two important temperature ranges, namely, $T > T_{c1}$ (electron hopping) and $T < T_{c1}$ (polaron hopping), which differ in both physical characteristics of transport and T dependence of conductivity.

Aside from the crossover between electron and polaron hopping, we have also studied another crossover between elastic and inelastic cotunneling regimes. It takes place at the temperature $T = T_{c0}$, given by (136). In the presence of strong polaronic effect, this temperature turns out to be the same as for single quantum dot, in contrast with the electron hopping regime, where it is much lower.

A few important questions remain open and are subject to future research.

(i) In this work, we have treated the configurational degrees of freedom (i.e., coordinates of the charged impurities) as classic ones. This can be justified only if the temperature T is higher than characteristic frequencies ω_i of the impurities vibrations; at lower T, the polaronic barriers would be penetrated by means of quantum tunneling. In principle, it should lead to the reentrance of Mott law at $T < \overline{\omega}$. However, since ω_i may be different for different impurities, the crossover to tunneling may occur not simultaneously at all grains, and that may give rise to some interesting new physics.

(ii) We have studied here the case of short-range interaction; only some qualitative ideas about the effects of the long-range Coulomb interaction were given is Sec. X D.

(iii) The lack of experimental data on polaron effect in granular systems does not allow to reliably choose the distribution function of the polaronic barriers W. It is not clear whether it should be wide or narrow, have a power-law tail at $W \to 0$, or exponential one.

(iv) It is also unclear if the present model, in which the occupation numbers n are coupled to oscillators, is adequate for real-world applications. Another possible option would be the coupling to the tunneling two-level systems (TTLS, see [31,32]). Physically the two-level system may be represented by impurity atoms that can tunnel between two adjacent potential wells. Such systems play an important role in physics of glasses, and they also contribute to dephasing in qubits [33].

We believe that the formalism developed in this paper will allow us to answer at least the first question from the list above. As for the other questions, some novel approaches may be required.

ACKNOWLEDGMENTS

We are indebted to M. V. Feigel'man and L. B. Ioffe for helpful discussions. This work was supported by Basic research program of National Research University Higher School of Economics.

APPENDIX A: LIST OF NOTATIONS

Here, in order to avoid confusion, we produce a list of notations for most important characteristic energies, appearing in this work.

$E_C(\vec{N},\vec{Q})$	Coulomb energy of the system
$E_{\text{dev}}(\vec{Q})$	Elastic energy of impurities
$E_{\text{ch}}(\vec{N},\vec{Q})$	Potential energy of a system of charges
$\Delta_{\text{if}}(\vec{q}) \equiv E_{\text{ch}}(\vec{n}_f,\vec{q}) - E_{\text{ch}}(\vec{n}_i,\vec{q})$	Franck-Condon energy difference
$E_k^{(\pm)}(\vec{q})$, $E_{lr}^{(\pm\pm)}(\vec{q})$	Single- and two-particle Franck-Condon excitation energies
$\tilde{E}(\vec{n})$	Thermodynamic (i.e., minimized with respect to q) energy
$\tilde{\Delta}_{\text{if}} \equiv \tilde{E}(\vec{n}_f) - \tilde{E}(\vec{n}_i)$	Thermodynamic energy difference
$\tilde{E}_k^{(\pm)}$, $\tilde{E}_{lr}^{(\pm\pm)}$	Single- and two-particle thermodynamic excitation energies
$E_k^c = e^2/2C_k$	Charging energy of grain k
W_k	Polaronic barrier's height at grain k

APPENDIX B: PERCOLATION WITH GAUSSIAN DISTRIBUTION OF BARRIERS. DERIVATION OF THE FORMULA (117)

We are interested in the low-T behavior of the system, when the critical subnetwork, responsible for the conductivity, is comprised of the rare grains with anomalously small ϵ_j and anomalously low barriers W_j from the tail of the distribution (99). As we will see later, these parameters are confined in narrow strips

$$\epsilon_j \lesssim \bar{\varepsilon}_\epsilon, \quad W_j - \xi T \lesssim \bar{\varepsilon}_W, \quad \text{with } \bar{\varepsilon}_\epsilon, \bar{\varepsilon}_W \ll \xi T. \tag{B1}$$

Now, we will adopt the inequalities (B1) as a conjecture, the real values of $\bar{\varepsilon}_\epsilon, \bar{\varepsilon}_W$ will be determined later, and these values will justify (B1).

Based on the assumption (B1), the expression (107) for Λ_{kj} can be expanded and we get

$$E(\epsilon_k W_k | \epsilon_j W_j) = \frac{|\epsilon_k| + |\epsilon_j|}{2} + W_k + W_j. \tag{B2}$$

The result (B2) means that indeed each site may be characterized by one "composite color" $\tilde{W}_k \equiv W_k + |\epsilon_k|/2 > 0$ not by two independent colors ϵ_k and W_k. The density of this composite color is

$$\tilde{n}(\tilde{W}) = 2\int_0^{2\tilde{W}} n(\epsilon, \tilde{W} - \epsilon/2) d\epsilon \tag{B3}$$

(the factor 2 arises due to symmetry $\varepsilon \to -\varepsilon$), and

$$\xi_{kj} \equiv \frac{2|\mathbf{r}_k - \mathbf{r}_j|}{a} + \frac{E_{ij}}{T}, \quad E_{kj} \equiv E(\tilde{W}_k | \tilde{W}_j) = \tilde{W}_k + \tilde{W}_j. \tag{B4}$$

Now, we introduce dimensionless variables

$$\mathbf{x}_k \equiv \frac{2\mathbf{r}}{\xi a}, \qquad \delta_k \equiv \frac{1}{2} - \frac{\tilde{W}_k}{\xi T} \tag{B5}$$

and arrive at the dimensionless percolation problem

$$\tilde{\xi}_{kj} \equiv |\mathbf{x}_k - \mathbf{x}_j| - \delta_k - \delta_j < 0 \tag{B6}$$

with the density in the $(\mathbf{x}, \delta)$ space

$$\tilde{n}(\delta) = 2\xi T \left(\frac{\xi a}{2}\right)^d \int_0^{\xi T\left(\frac{1}{2} - \delta\right)} n\left[\epsilon, \xi T\left(\frac{1}{2} - \delta\right) - \frac{\epsilon}{2}\right] d\epsilon$$
$$\approx \frac{4\beta' T}{T'_M} (\xi/2)^{d+1} \Phi\left[\frac{\overline{W} - \xi T\left(\frac{1}{2} - \delta\right)}{\delta W}\right], \tag{B7}$$

where

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left\{-\frac{\zeta^2}{2}\right\} d\zeta. \tag{B8}$$

In particular, we will need the asymptotics

$$\Phi(x) \approx \frac{1}{x\sqrt{2\pi}} \exp\{-x^2/2\} \qquad (x \gg 1). \tag{B9}$$

It is convenient to write

$$\sigma \propto \exp\left\{-\frac{W_{\text{eff}}(T)}{T}\right\}, \qquad W_{\text{eff}}(T) \equiv \xi T, \tag{B10}$$

and note that $w \equiv 2\overline{W} - W_{\text{eff}}(T) \ll \overline{W}$ and $\xi T \delta \ll \overline{W}$, so that

$$\tilde{n}(\delta) \approx \frac{4\beta' T}{T'_M} \left(\frac{\xi}{2}\right)^{d+1} \Phi\left(\frac{w/2 + 2\overline{W}\delta}{\delta W}\right),$$

$$\Phi\left(\frac{w/2 + 2\overline{W}\delta}{\delta W}\right) \approx \frac{1}{\sqrt{2\pi}} \frac{\delta W}{w/2 + 2\overline{W}\delta}$$
$$\times \exp\left\{-\frac{(w/2)^2}{2\delta W^2} - \frac{w}{\delta W^2}\overline{W}\delta - \frac{(2\overline{W}\delta)^2}{2\delta W^2}\right\}. \tag{B11}$$

As we will see soon, for typical δ and w a hierarchy $\overline{W}\delta \ll \delta W \ll w$ holds. Therefore, only the first two terms in the exponent should be kept, while the last term is much less than unity and can be neglected. As a result, we can rewrite (B11) in a form

$$\tilde{n}(\delta) \approx \frac{4\beta' T}{T'_M} \left(\frac{\xi}{2}\right)^{d+1} \frac{1}{\sqrt{2\pi}} \frac{2\delta W}{w} \exp\left\{-\frac{w^2}{8\delta W^2} - \frac{w\overline{W}\delta}{\delta W^2}\right\}. \tag{B12}$$

After the renormalization of variables

$$u_k \equiv \frac{w\overline{W}}{\delta W^2}\delta_k, \qquad \mathbf{y}_k = \frac{w\overline{W}}{\delta W^2}\mathbf{x}_k, \tag{B13}$$

we arrive at the universal percolation problem, where the sites of color u are randomly distributed with the density

$$\tilde{n}(u) \approx A e^{-u}, \tag{B14}$$

while the percolation criterion reads as

$$\tilde{\xi}_{kj} \equiv |\mathbf{y}_k - \mathbf{y}_j| - u_k - u_j < 0. \tag{B15}$$

Note that this problem is characterized by single constant

$$A \equiv \left(\frac{w\overline{W}}{\delta W^2}\right)^{-(d+1)} \tilde{n}(\delta = 0)$$
$$= \frac{8\beta' T}{T'_M} \frac{1}{\sqrt{2\pi}} \frac{\delta W}{w} \left(\frac{wT}{\delta W^2}\right)^{-(d+1)} \exp\left\{-\frac{w^2}{8\delta W^2}\right\} \tag{B16}$$

and therefore the percolation should be established at

$$A = A_d, \tag{B17}$$

where $A_d \sim 1$ is some universal constant, depending only on the space dimensionality d.

So, the dependence of the effective barrier $W_{\text{eff}}(T) \equiv 2\overline{W} - w$ on the parameters may be found from the equation for w:

$$1 = \frac{8\beta' T}{A_d T'_M} \frac{1}{\sqrt{2\pi}} \frac{\delta W}{w} \left(\frac{wT}{\delta W^2}\right)^{-(d+1)} \exp\left\{-\frac{w^2}{8\delta W^2}\right\}. \tag{B18}$$

In the leading logarithmic approximation, the solution of (B18) reads as

$$\frac{w}{\delta W} \approx \sqrt{8d\,\ln(T_{c2}/T)}, \quad T_{c2} \sim \delta W \left(\frac{\delta W}{T'_M}\right)^{1/d}. \tag{B19}$$

Now, we can estimate $\overline{\varepsilon}_\varepsilon$ and $\overline{\varepsilon}_W$. As it follows from (B13)

$$\overline{\varepsilon}_\varepsilon \sim \overline{\varepsilon}_W \sim \frac{\delta W^2}{w} \sim \frac{\delta W}{\sqrt{8d\,\ln(T_{c2}/T)}} \ll \delta W, \qquad \text{(B20)}$$

which justifies our conjecture (B1).

In the presence of the long-range interaction, resulting in the Coulomb gap in the density of states, one should write $n(\epsilon,W) \propto \epsilon^{d-1}P(W)$, which leads to the following:

(1) Replacement of T'_M by T_0 (and, consequently, to replacement of T_{c2} by T^*_{c2}).

(2) Appearance of an additional factor $(\delta W/w)^{d-1}$ on the right-hand side of (B18). This factor is, however, only a logarithmic one [as follows from (B19)], so it does not change the final result in the leading logarithmic approximation.

[1] N. F. Mott and E. A. Davis, *Electron Processes in Noncrystalline Materials* (Clarendon, Oxford, 1979).

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

[3] N. F. Mott, J. Non-Cryst. Solids **1**, 1 (1968).

[4] A. L. Efros and B. I. Shklovskii, J. Phys. C: Solid State Phys. **8**, L49 (1975).

[5] B. Abeles, P. Sheng, M. D. Coutts, and Y. Arie, Adv. Phys. **24**, 407 (1975).

[6] I. S. Beloborodov, A. V. Lopatin, V. M. Vinokur, and K. B. Efetov, Rev. Mod. Phys. **79**, 469 (2007).

[7] T. Chui, G. Deutscher, P. Lindenfeld, and W. L. McLean, Phys. Rev. B **23**, 6172 (1981); C. J. Adkins, in *Proceedings of the Sixth International Conference on Hopping and Related Phenomena*, edited by O. Millo and Z. Ovadyahu (Racah Institute of Physics, Jerusalem, 1995), p. 245.

[8] D. Yu, C. Wang, B. L. Wehrenberg, and P. Guyot-Sionnest, Phys. Rev. Lett. **92**, 216802 (2004); M. A. Rafiq *et al.*, J. Appl. Phys. **100**, 014303 (2006); M. Salvato, M. Lucci, I. Ottaviani, M. Cirillo, E. Tamburri, S. Orlanducci, M. L. Terranova, M. Notarianni, C. C. Young, N. Behabtu, and M. Pasquali, Phys. Rev. B **86**, 115117 (2012); Mitsuru Inada *et al.*, APEX **8**, 105001 (2015).

[9] P. Sheng, B. Abeles, and Y. Arie, Phys. Rev. Lett. **31**, 44 (1973); C. J. Simanek, Solid State Commun. **40**, 1021 (1981); E. Z. Meilikhov, J. Exp. Theor. Phys. **88**, 819 (1999).

[10] M. Pollak and C. J. Adkins, Philos. Mag. B **65**, 855 (1992).

[11] J. Zhang and B. I. Shklovskii, Phys. Rev. B **70**, 115317 (2004).

[12] I. P. Zviagin and R. Keiper, Philos. Mag. B **81**, 997 (2001); V. I. Kozub, V. M. Kozhevin, D. A. Yavsin, and S. A. Gurevich, JETP Lett. **81**, 226 (2005).

[13] M. V. Feigel'man and A. S. Ioselevich, JETP Lett. **81**, 277 (2005).

[14] D. V. Averin and Y. V. Nazarov, Phys. Rev. Lett. **65**, 2446 (1990).

[15] Y. V. Nazarov and Y. M. Blanter, *Quantum Transport, Introduction to Nanoscience* (Cambridge University Press, Cambridge, 2009).

[16] A. V. Aleshin, A. N. Ionov, R. V. Parfeniev, I. S. Shlimak, A. Heinrich, I. Schumann, and D. Elefant, Fiz. Tverd. Tela. (Leningrad) **30**, 696 (1988) [Sov. Phys. Solid State **30**, 398 (1988)]; H. Vinzelberg, A. Heinrich, G. Gladun, and D. Elefant, Philos. Mag. B **65**, 651 (1992); R. W. van der Heijden, G. Chen, A. T. A. M. de Waele, H. M. Gijsman, and F. P. B. Tielen, Solid State Commun. **78**, 5 (1991); V. Voegele, S. Kalbitzer, and K. Bohringer, Philos. Mag. B **52**, 153 (1985); I. Terry, T. Penney, S. von Molnar, and P. Becla, Phys. Rev. Lett. **69**, 1800 (1992); P. Dai, Y. Zhang, and M. P. Sarachik, *ibid.* **69**, 1804 (1992); J. J. Kim and H. J. Lee, *ibid.* **70**, 2798 (1993).

[17] A. S. Ioselevich, Phys. Rev. Lett. **71**, 1067 (1993).

[18] S. D. Baranovskii and V. G. Karpov, Fiz. Tekh. Poluprovodn. **20**, 1811 (1986) [Sov. Phys.–Semicond. **20**, 1811 (1986)].

[19] A. S. Ioselevich, Pis'ma Zh. Eksp. Teor. Fiz. **43**, 148 (1986) [JETP Lett. **43**, 188 (1986)].

[20] M. Foygel, R. D. Morris, and A. G. Petukhov, Phys. Rev. B **67**, 134205 (2003).

[21] A. I. Yakimov, A. V. Dvurechenskii, G. M. Min'kov, A. A. Sherstobitiov, A. I. Nikiforova, and A. A. Bloshkina, JETP **100**, 722 (2005).

[22] H. Xing, W. Kong, C. Kim, S. Peng, S. Sun, Z.-A. Xu, and H. Zeng, J. Appl. Phys. **105**, 063920 (2009).

[23] A. J. Houtepen, D. Kockmann, and D. Vanmaekelbergh, Nano Lett. **8**, 3516 (2008).

[24] B. Skinner, T. Chen, and B. I. Shklovskii, Phys. Rev. B **85**, 205316 (2012).

[25] L. I. Glazman and M. Pustilnik, *Nanophysics: Coherence and Transport*, edited by H. Bouchiat *et al.* (Elsevier, 2005), pp. 427–478.

[26] I. S. Beloborodov, A. V. Lopatin, and V. M. Vinokur, Phys. Rev. B **72**, 125121 (2005).

[27] V. I. Mel'nikov, Phys. Lett. A **176**, 267 (1993); see also C. Wasshuber, *Computational Single-Electronics* (Springer, Vienna, 2001).

[28] V. Ambegaokar, B. I. Halperin, and J. S. Langer, Phys. Rev. B **4**, 2612 (1971).

[29] A. S. Ioselevich, Phys. Rev. Lett. **74**, 1411 (1995).

[30] L. V. Keldysh, Pis'ma Zh. Eksp. Teor. Fiz. **29**, 716 (1979) [JETP Lett. **29**, 658 (1979)]; J. E. Mooij and G. Schön, Single charges in 2-dimensional junction arrays, in *Single Charge Tunneling*, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1996), pp. 275–310; A. S. Ioselevich and V. V. Sivak, Phys. Rev. B **92**, 224203 (2015).

[31] W. A. Phillips, J. Low Temp. Phys. **7**, 351 (1972).

[32] P. W. Anderson, B. I. Halperin, and C. M. Varma, Philos. Mag. **25**, 1 (1972).

[33] A. Shnirman, Y. Makhlin, and G. Schön, Phys. Scr. **T102**, 147 (2002).