ELECTRONIC TRANSPORT IN GRANULAR METAL FILMS

PING SHENG
Exxon Research & Engineering Company
Route 22 East
Annandale, NJ 08801 U.S.A.

ABSTRACT: Granular metals are composites consisting of a random mixture of nanometer-sized metal and insulator grains. As a function of metal volume fraction, the structure and electrical properties of the granular metals can be divided into two regimes, separated by the percolation threshold. In the metal-rich regime, metal grains form a connected network, and electrical conduction is by electron percolation through the metallic channels. The physical basis and formulation of an effective-medium theory are described for the calculation of this percolative aspect of the electrical transport. The role of microstructure in the effective medium property is particularly emphasized. In the insulator-rich regimes, metal grains are dispersed in the matrix of the insulator. Electrical conduction in this dielectric regime is via the hopping mechanism, which is a term coined to denote the combination of thermal activation and tunneling. A critical-path method is presented to show that activation and tunneling, nominally independent, can actually couple to give a $-\ell n\ \sigma \propto T^{-\alpha}$ type of temperature dependence for the conductivity σ. For granular metals, the widely observed $-\ell n\ \sigma \propto 1/\sqrt{T}$ behavior is ascribed to be an interpolation between the high-temperature activated behavior and the low-temperature $\alpha = 1/4$ behavior. The alternative mechanism of the correlation gap is also discussed.

1. Introduction

Granular metals are metal-insulator composites consisting of interdispersed metallic and insulating grains on the order of 100Å in size. They are mostly in the form of thin films <1 μm in thickness and are formed by co-sputtering or co-evaporation. Over the past two decades, the electrical transport properties of granular metal films have been a topic of active experimental and theoretical study [Abeles, 1976; Abeles, Sheng, Coutts, and Arie, 1975]. Depending on the metal volume fraction p, there are two regimes of electronic conduction. When p is large, the metallic grains form a connected metallic network so that the electrons can percolate directly through connected metallic channels. This is denoted the metallic regime. When p is small, the metal grains form isolated dispersions in an insulating matrix. Electrical conduction in this dielectric regime is by the hopping mechanism in which the charge carriers are transported from one metal grain to another via thermally-activated tunneling. The transition between the metallic and the dielectric regimes is characterized by a percolation threshold at which the metallic network first

G. C. Hadjipanayis and R. W. Siegel (eds.), Nanophase Materials, 381–398.

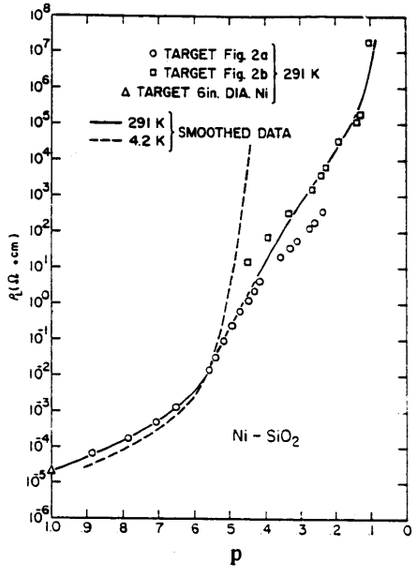


Figure 1. Resistivity as a function of volume fraction p of Ni, in sputtered Ni-SO_2films. The resistivities were measured in the plane of the film at 291 K and 4.2 K. The full and dashed curves are smoothed values of the experimental data.

becomes disconnected and the resistivity increases dramatically as p decreases further. This behavior is shown in Fig. 1. In general, the percolation threshold occurs at $0.4 \le p \le 0.6$.

It is the purpose of these lectures to give the audience a feel for the present theoretical understanding of electronic transport behavior in the two regimes. For brevity, we omit the description on the development of this interesting field. Interested readers are referred to the Abeles review article [1976] for historical perspectives. In what follows, part two is on the metallic regime. Part three is on the dielectric regime, with discussions on hopping conductivity [Sheng, Abeles, and Arie, 1973] presented through the critical path approach [Sheng and Klafter, 1983], numerical simulation of the hopping conductivity, and a discussion of the nature of the Coulomb gap [Klafter and Sheng, 1984] in granular metals.

2. Conductivity in the Metallic Regime

In Fig. 1, two sets of data are shown for the same samples: the room-temperature data and the liquid-helium temperature (4.2 K) data. It is seen that the temperature dependencies of the two regimes are opposite to each other. This is because the conductivity

$$\sigma = e\, n\, \mu ,$$

where e is the electronic charge, n the density of charge carriers, and μ their mobility. In the metallic regime n is a temperature-independent constant, but μ decreases with temperature T as collisions with phonons become more frequent. As a result, σ decreases as T increases. In the dielectric regime, n is strongly temperature-dependent because the available charge carriers are

thermally-activated, so $n \sim \exp(-\Delta E/kT)$, where ΔE is the activation energy and k the Boltzmann constant. Since n increases with temperature, σ increases also.

At the percolation threshold, the granular metal conductivity is nearly T-independent and yet can be very resistive. This is very unusual, since normally resistive materials have a large $d\sigma/dT$ so that thermal runaway can become a problem. Here, however, the novel material can act as a resistive voltage divider without the risk of thermal runaway or burnout.

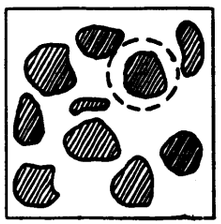
A few words on the applications of granular metals besides the one just mentioned could be interesting to the audience. The strongly non-linear I-V characteristics (at high-electric field) of the composite have been the basis for a varistor device to protect electrical circuits from voltage surges. Composites with semiconducting particles instead of metallic particles can be used as solar-thermal absorbers which absorbs sunlight efficiently but which has a low infrared emissivity. Specialized granular metals have also been used as radar absorbers, related to the Stealth technology. However, overall the granular metals are more rich in novel phenomena than in applications, at least at present. Let us now turn our attention to the metallic conductivity problem.

Since the metallic regime is defined structurally as the composition range where the metal grains form a percolating network, the electrical conductivity in this regime is a direct result of electron transport through metallic channels. The microgeometry of the metallic channels therefore plays an important role in the actual value of the conductivity. The effective medium theory [Sheng, 1980] is used to calculate such conductivity, which is described below. However, an important deviation from the metallic behavior should be mentioned before we begin. That is, at low temperatures the conductivity of all the granular metals has been observed to decrease as temperature decreases [Beamish, Patterson, and Unruh, 1990]. Such non-metallic behavior can be explained qualitatively in terms of the electron multiple scattering and localization effects. I am not going to describe the localization effect, but the audience should be aware of such phenomenon at $T < 10K$.

2.1. EFFECTIVE MEDIUM THEORY AND THE MICROSTRUCTURE

When a heterogeneous material is probed by an electromagnetic wave, the resolution limit of the probing wave is set by $\lambda/2$, where λ is the wavelength. In addition, the Rayleigh scattering strength is known to vary as λ^{-4}. It follows that in the long wavelength limit, where $\lambda >> a$, a being the scale of the inhomogeneities, the heterogeneous material would appear homogeneous to the probing wave, with the material properties characterized by effective parameters. As a matter of fact, the physical concept of a "homogeneous medium" is inherently based on the effective medium idea because all materials are discrete if probed with sufficient resolution: molecules are composed of atoms, which in turn are composed of electrons and nuclei, which in turn are composed of protons and neutrons, which in turn are composed of quarks, etc. Therefore, it is important to realize that effective medium is not just a mathematical tool but is rather based on physical experience. An effective medium theory has as its goal the calculation of the effective parameters of the "homogenized" material in terms of the materials and geometric properties of its constituent phases. For the effective-medium calculation of the composite electrical conductivity, the input material properties of the components are summarized by their respective complex dielectric constants $\epsilon_i = \kappa_i + i4\pi\, \sigma_i/\omega$, where κ_i is the real dielectric constant of component i (note that the i in front of 4π means $\sqrt{-1}$), relevant to the polarizability of the component i; σ_i

is the electrical conductivity, and ω is the (angular) frequency. Here we will assume the material to be non-magnetic, so the magnetic permeability is taken to be 1. It should be mentioned that since the criterion for the validity of the effective medium description is the small ratio a/λ, if a is on the order of nanometers, λ can be in the optical range and the effective medium theory would still be valid. The point here is that ω need not necessarily be small. The other input to the effective medium theory is the geometric properties of the constituent components: the relative volume fraction and the shapes and arrangement of the grains. In presenting the effective medium theories, I wish to adopt an approach in which the role of microstructure can be most easily delineated. Basically, the approach is based on considering the composite as made up of elementary structural units. For example, in Figure 2a, we show that in a composite consisting of dispersed inclusions in a matrix the basic unit may be taken as a coated grain. If the inclusions are allowed to touch, however, then the two phases should be considered on an equal basis. This implies that a grain of constituent 1 and a grain of constituent 2 are the two basic units. The resulting structure is schematically illustrated in Figure 2b. A cluster of grains may also be considered as a basic structural unit, but then the theory would lose its calculational simplicity. Once the basic units are chosen, the next step is the embedding of each individual unit in a homogeneous effective medium characterized by a yet undetermined effective complex dielectric constant $\bar{\epsilon}$. To calculate $\bar{\epsilon}$, the forward scattering amplitude, $f_i(0)$, for an incident electromagnetic plane wave by the ith unit is calculated in the long wavelength limit. The scattering amplitude $f_i(\theta)$, where θ means the angle relative to the incident wave direction, is a quantity that can be easily obtained from the solution of a boundary value problem. The effective medium condition is then [Lax, 1951]



(a)

(b)

Figure 2. (a) Schematic depiction of the dispersed-inclusion microstructure. The dashed line delineates a basic structural unit for the composite. (b) Schematic depiction of the symmetric microstructure. Depending on the relative volume fraction of the two components, either component can be the matrix phase.

$$\sum_i p_i f_i(0) = 0, \tag{1}$$

where p_i is the volume fraction of the ith unit. Since $f_i(0)$ is an implicit function of $\bar{\epsilon}$, Eq. (1) represents a condition for its determination.

The condition that the forward scattering amplitudes must be zero on average can be justified heuristically as follows. Since the wavelength of the probing wave is much longer than the scale of basic structural units, the medium should appear homogeneous to the probing wave. Therefore, if one looks in the propagating direction of a plane wave there should be, on the average, no net scattering out of the beam. Application of this approach for deriving the effective medium theories shows that the input structural units determine the type of theory obtained. For example, we can have two types of random composites with the microstructures depicted schematically in Figures 2a and 2b. By using for the basic structural unit a coated sphere with a coating layer thickness determined by the overall composition, we get the equation for the effective dielectric constant $\bar{\epsilon}$ of the dispersed inclusion microstructure:

$$\frac{\bar{\epsilon}-\epsilon_m}{\bar{\epsilon}+2\epsilon_m} + p\frac{\epsilon_m-\epsilon_i}{\epsilon_i+2\epsilon_m} = 0 , \tag{2}$$

where ϵ_m is the matrix dielectric constant, ϵ_i the inclusion dielectric constant, and p the inclusion volume fraction. Equation (2) is known as the Maxwell-Garnett theory (MGT) [Maxwell-Garnett, 1904]. For the symmetric microgeometry (Figure 2b), on the other hand, there are two structural units (spheres of two components) as mentioned before. The resulting equation is

$$p\frac{\bar{\epsilon}-\epsilon_1}{2\bar{\epsilon}+\epsilon_1} + (1-p)\frac{\bar{\epsilon}-\epsilon_2}{2\bar{\epsilon}+\epsilon_2} = 0 , \tag{3}$$

which is known in the literature as Bruggeman's symmetric effective medium theory (EMT) [Bruggeman, 1935]. The predictions of Eqs. (2) and (3) are very different. For dc conductivity of a metal-insulator composite, Eq.(3) predicts a percolation threshold at one-third volume fraction of metal (i.e., conductivity = 0 for metal fraction less than one-third), whereas Eq.(2) tells us that for insulating inclusions, the dc conductivity vanishes only when metal volume fraction approaches zero. This difference is easily linked with the differing microstructures. For the dispersed inclusion microstructure, the insulator can never fully block off the conducting matrix unless it is at p=1. However, for the symmetric microgeometry the two constituents can undergo a matrix inversion as their relative volume fraction is varied. This accounts for the percolation threshold. At the optical frequency regime (but wavelength still >> scale of inhomogeneities) the effective dielectric constants calculated from Eqs. (2) and (3) display equally different behaviors. Whereas the MGT predicts the existence of the optical dielectric anomaly observed in granular metal films [Sheng, 1980], the EMT does not yield a dielectric anomaly. However, EMT does give a percolation threshold as observed, although the predicted threshold value is low compared with the experimental result. It follows that for granular metals neither MGT nor EMT is adequate, and a new effective medium theory is necessary.

2.2. EFFECTIVE MEDIUM THEORY FOR GRANULAR METALS

As the starting point of a theory appropriate for granular metals, it is necessary to first take into account the formation process for the grains by diffusion and coalescence of the sputtered molecules on the substrate. Consider a spherical region with the dimension of a diffusion length inside the material. Within such a region a fraction p of the volume is taken by the molecules of component 1 and the rest by component 2. Here p is the macroscopic volume composition of component 1. When a grain is formed inside this region by diffusion and coalescence from the sputtered molecules, there are two possible outcomes: Component 1 forms the grain and component 2 the coating, which we denote as a type-1 unit; or component 2 may form the grain and component 1 the coating, which is denoted as a type-2 unit. The relative probability of occurrence for the two cases can be estimated by counting the number of equally possible final configurations (corresponding to different positions of the grain inside the region). By assuming the grain to be spherical, it is clear that, in the case of type 1, the number of configurations is proportional to $\upsilon_1 = (1-p^{1/3})^3$, the free volume (up to a consistent multiplicative factor) accessible by the grain center of mass inside the spherical region. By the same reasoning the number of configurations for the alternative case is proportional to $\upsilon_2 = [1-(1-p)^{1/3}]^3$. It follows that, at any value of p, the relative probability of occurrence for type 1 unit is $f = \upsilon_1/(\upsilon_1+\upsilon_2)$, and that for type 2 unit is 1-f. This simplified picture, though it neglects the interaction between neighboring regions in the grain-forming process, nevertheless does capture the essential structural variation in granular composites as a function of composition p. For $0 \le p \le 0.25$, the value of f decreases slowly from 1 at p=0 to ~0.92 at p=0.35. The preponderance of type-1 unit in this range is in accord with the observed structure of grains of component 1 embedded in the matrix of component 2. As p is increased from 0.35 to 0.65, f drops sharply, reaching ~0.08 at p=0.65. This composition region, with the two kinds of structural units competing for dominance, clearly corresponds to the structural transition regime seen in electron micrographs [Abeles et al., 1975] at similar values of p. Beyond p = 0.65, the matrix inversion is basically complete, and we have grains of component 2 dispersed in the matrix of component 1.

To calculate the dielectric constant of the composite, let us consider the embedding of a type-1 or a type-2 unit in a uniform effective medium of dielectric constant $\bar{\epsilon}$. Upon the application of a uniform electric field, the inclusion would produce a dipole moment $D_{1,2}(i)$, where the subscript corresponds to the type of unit and i denotes the configuration. If we approximate $D_{1,2}(i)$ with the dipole moment of the concentric configuration, $D_{1,2}$, the effective-medium condition of zero average dipole moment [Landauer, 1952] results in the equation

$$fD_1 + (1-f)D_2 = 0. \tag{4}$$

It may be noted that Eq.(4) reduces to the MGT if f is set equal to 1. Although the discussion to this point has been confined to the case of spherical units, Eq.(4) is generally applicable to particles of spheroidal shapes as well, since the preceding arguments remain unchanged if one considers a spheroidal particle enclosed in a similar-shaped region. In that general case, however, $D_{1,2}$ stands for the orientationally averaged dipole moment of coated confocal spheroidal particles embedded in an effective medium $\bar{\epsilon}$:

$$D_1 = \frac{2}{3}D(\bar{\epsilon},\epsilon_1,\epsilon_2,p,A(\alpha,u),B(\alpha)) + \frac{1}{3}D(\bar{\epsilon},\epsilon_1,\epsilon_2,p,3-2A(\alpha,u),3-2B(\alpha)), \tag{5a}$$

$$D_2 = \frac{2}{3}D(\bar{\epsilon},\epsilon_2,\epsilon_1,1-p,A(\beta,\upsilon),B(\beta)) + \frac{1}{3}D(\bar{\epsilon},\epsilon_2,\epsilon_1,1-p,3-2A(\beta,\upsilon),3-2B(\beta)), \tag{5b}$$

where α is the ratio between the minor (major) and major (minor) axes of the elliptic cross section for the type-1 oblate (prolate) spheriodal unit, β is the similar quantity for the type-2 unit, $u = (p/\alpha)^{1/3}$,

$$D(\bar{\epsilon},x,y,u,A,B) = \frac{[A\bar{\epsilon}+(3-A)y](y-x)u+[Bx+(3-B)y](\bar{\epsilon}-y)}{A(3-A)(\bar{\epsilon}-y)(y-x)u+[Bx+(3-B)y][Ay+(3-A)\bar{\epsilon}]}, \tag{6a}$$

$$A(\gamma,\zeta) = \frac{3}{2}\frac{1}{(1-\gamma^2)\zeta^3}\left[\frac{1}{(1-\gamma^2)^{1/2}}\tan^{-1}\frac{(1-\gamma^2)^{1/2}\zeta}{(s^2+\gamma^2\zeta^2)^{1/2}} - \zeta(s^2+\gamma^2\zeta^2)\right], \tag{6b}$$

$\beta(\gamma) = A(\gamma, 1/\gamma^{1/3})$ evaluated at s=0, s being the solution of the equation $(s^2+\zeta^2)^2(s^2+\gamma^2\zeta^2)=1$. The predictions of Eqs. (4)-(6) have some similarities to both the dispersed-inclusion and symmetric microstructure cases. As shown in Figure 3, the dc conductivity exhibits a percolation threshold as in the symmetric case. However, the annealed sample and the as-prepared sample is seen to exhibit opposite curvatures in the $\bar{\sigma}(p)$ variation. This characteristic trait is remarkably reproduced by the theory with two different values of β. The good agreement between theory and experiment therefore suggests that annealing modifies the geometry of the insulator particles from platelike before annealing to more spherical after annealing. The resulting increase in $\bar{\sigma}(p)$ upon annealing can then be simply explained by the fact that, for the same amount of insulator, platelets are expected to be more effective than spheres in impeding current flow. The second feature of the figure is the generally high value of percolation threshold compared with predictions of EMT and the random-percolation model [Scher and Zallen, 1970]. The result of the present theory shows that the structural constraint imposed by the condition of local composition homogeneity is sufficient to raise the percolation threshold to the observed range of $p_c \simeq 0.35$-0.5.

At optical frequencies (but the wavelength still >> the scale of the inhomogeneities) the predictions of the present theory have also been shown to correctly predict the overall frequency and composition dependencies of the measured transmission data [MacMillan, Devaty, and Mantese, 1991].

3. Conductivity in the Dielectric Regime

When the volume fraction of metal falls below the percolation threshold, there can no longer be a connected metallic network. However, electrical conduction can still occur through the tunneling of charge carriers from one metallic grain to the next along the applied field direction. An important concept for the understanding of electrical conductivity behavior in this regime is the charging energy E_c, [Neugebauer and Webb, 1962] which arises because the grain size is usually small (~100Å), and as a result the energy required to remove from or add an electron to the metal

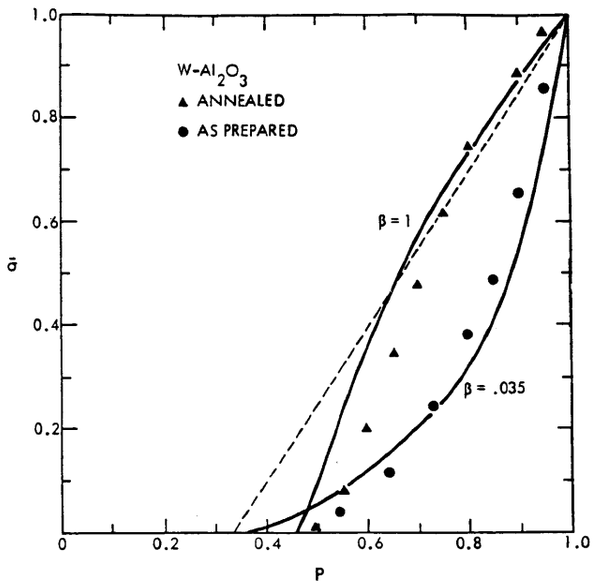


Figure 3. Normalized effective conductivity $\bar{\sigma}$ plotted as a function of metal volume fraction p for samples of W-Al_2O_3 cermets. The data are from Abeles, Sheng, Coutts, and Arie [1975]. Solid lines are calculated from theory. Dashed line denotes Bruggeman's effective medium prediction.

grain, $e^2/\kappa C$—where C~d, the diameter of a grain, is the capacitance of a grain and κ is an effective dielectric constant—can be on the order of 30 meV for $\kappa \simeq 10$. The existence of the charging energy in granular metals raises a number of interesting considerations which are detailed below.

3.1. HOPPING CONDUCTIVITY

Hopping conduction is a term coined to denote the electrical conduction mechanism which is the result of two combined processes: thermal activation and tunneling [Gudden and Schottky, 1935]. Mott was the first one to show that these two nominally independent processes can actually couple to result in non-activation type of behavior [Mott, 1968]. This coupling may be made clear through the critical path method (CPM) presented below, which was first invented by Ambegaokar, Halperin and Langer [1971].

Consider the conductance G_{ij} between two grains, i and j. The G_{ij} may be expressed as the product of the probabilities for thermal activation and tunneling [Miller and Abrahams, 1960]:

$$G_{ij} = G_0 \exp(-2\chi S_{ij} - E_{ij}/kT), \tag{7}$$

where χ is the WKB tunneling constant, generally on the order of 1Å^{-1} in granular metals, S_{ij} denotes the tunneling distance between the two grains (which is not the center-to-center separation

of the grains, but rather the distance of closest approach between the grain surfaces) k denotes Boltzmann's constant, and

$$E_{ij} = \frac{1}{2}(|E_i - E_j| + E_i + E_j). \tag{8}$$

The expression for E_{ij} is the result of considering both the probability of having a charge carrier at grain i and the additional activation, if required, for tunneling to gain j. E_i denotes the changing energy of grain i. The indices i and j are not required to denote only the nearest-neighbor grains. Rather, they span all pairs of grains. How does one arrive at the expression for E_{ij}? That requires some explanation. Suppose grain i is bigger than grain j. For a charge on grain i to move to grain j, it has to overcome an energy barrier that is the difference of the two charging energies. However, for a charge to hop from j to i there is no such barrier. Therefore from this point of view the conductance is asymmetrical. However, what has been said ignored the probability that a charge carrier will be on grain i or j in the first place, before they hop. If we take into account that probability, which in thermal equilibrium is proportional to $\exp(-E_i/kT)$ or $\exp(-E_j/kT)$, then the conductance becomes symmetrical because in going from i to j means

$$\exp(-E_i/kT)\exp[-(E_j - E_i)/kT] = \exp(-E_j/kT)$$

and in going from j to i is the same. Therefore,

$$E_{ij} = \max[E_i, E_j],$$

and the expression (8) precisely gives the larger of E_i, E_j in a mathematical way. An additional comment about G_{ij} is that they are correlated and not independent from each other. For example, G_{ij} and G_{ik} share the same grain i. If that grain becomes smaller so its charging energy E_c becomes larger, then both G_{ij} and G_{ik} would become smaller in a correlated way. This correlation is a very important character of the hopping conductivity, and it prevents the problem from being simply soluble by a variety of methods which requires independent variation of each G_{ij}.

To apply the CPM, let us consider the following rule: pick a G and consider the pair ij connected if $G_{ij} > G$, and disconnected otherwise. If we pick a value of G that is sufficiently large, then only a few G_{ij}'s can satisfy the criterion of connectedness and as a result the connected sites form only disjoint clusters. Now let us lower the value of G and apply the rule again. The connected clusters are expected to increase in size until at $G = G_c$, the percolation conductance, an infinite network of connected sites is formed. This value of G_c may be regarded as the conductance of the macroscopic sample for the following reasons.

First, since at $G > G_c$ the network is not percolating, the last added conductance with value G_c must be in series with the rest of the network. That means $1/G_c$ must be added together with the resistance of the rest of the network. Since Eq.(7) gives an exponential form for G_{ij}, distributions in the values of S_{ij} and E_{ij} translate into an exponentially broad distribution for G_{ij}. That means most resistances in the percolating network may be regarded as having zero resistances when compared with $1/G_c$, and only an exponentially small number of resistances are comparable with $1/G_c$, so that the overall network resistance should be proportional to $1/G_c$. Second, the

conductances with $G_{ij}<G_c$ can certainly add additional parallel conducting channels; however, here again the exponentially broad distribution of G_{ij} means these added parallel channels will mostly have conductances exponentially smaller than G_c, and only very few will be comparable in magnitude to G_c.

Thus, with only small modifications, we expect G_c to be proportional to the macroscopic conductance of the sample. Let us now examine the temperature dependence of this G_c, which basically arises from the inequality $G_{ij}\geq G_c$. By taking the logarithm of both sides of this inequality and substituting Eqs.(7) and (8) for G_{ij}, one gets

$$\frac{S_{ij}}{(1/2\chi)\ell n(G_o/G_c)}+\frac{1}{2}\frac{E_i+E_j+|E_i-E_j|}{kT\ell n(G_o/G_c)}\leq 1. \qquad (9)$$

Equation (9) immediately tells us that there is indeed a correlation between the tunneling and the activation processes. This correlation has its origin in the global optimization of the hopping conduction paths based on the concept of percolation. Now, since the inequality, Eq.(9), requires that $S_{ij}\leq(1/2\chi)\ell n(G_o/G_c)$ and $E_i,E_j\leq kT\ell n(G_o/G_c)$, one can define $S_m=(1/2\chi)\ell n(G_o/G_c)$ and $E_m=kT\ell n(G_o/G_c)$ as the maximum upper bounds for S_{ij} and E_i, respectively.

To make further progress requires the consideration of the formation requirement for a percolation network. One such criterion is the average number of bonds emanating from a site must exceed a critical number b_c before percolation is possible [Shante, 1973]. Since b_c is a dimensionless number, it must be proportional to $\rho(0)E_mS_m^3$, where $\rho(0)$ is the density of states in the vicinity of the Fermi level. In fact, a more detailed calculation has shown this to be the case [Sheng and Klafter, 1983]. By writing

$$b_c = f\rho(0)\,E_mS_m^3, \qquad (10)$$

where f is a proportionality constant and $\rho(0)$ is treated as a constant, then one immediately obtains

$$\frac{kT}{(2\chi)^3}[\ell n(G_o/G_c)]^4=\frac{b_c}{f\rho(0)}, \qquad (11)$$

or

$$\ell n\,G_c \propto -T^{-/4}\quad, \qquad (12)$$

which is Mott's famous $T^{-1/4}$ law widely observed in amorphous materials [Mott, 1968].

Physically, the fractional temperature dependence comes from the continuously shifting of the globally optimal conduction paths as the temperature is varied. This can best be described in terms of Mott's argument. At high temperatures, activation is easy, so the limiting factor is the tunneling distance. The optimal path is thus the one with the smallest tunneling distances between the successive sites, and the temperature dependence is expected to be activated, with the activation energy given by the largest E_i of the conduction path sites (a more detailed calculation shows a behavior of $-\ell nG_c\propto T^{-\alpha}$, with α less than one but greater than half [Sheng and Klafter, 1983]). As temperature is lowered, activation becomes more of a limiting factor, and the tradeoff between a lower activation energy and a larger tunneling distance becomes favorable to the latter. Another way of looking at the tradeoff is that as T is lowered, the number of accessible energy states (to within kT of the starting energy of the hopping charge carrier) becomes smaller, and it forces some tunneling distances to be larger so that there can still be a percolating network for conduction. Since the tunneling distance can be continuously increased as T decreases, this type of conduction is usually denoted as "variable-range hopping".

3.2. HOPPING CONDUCTION IN GRANULAR METALS

For granular metals, the temperature dependence observed is almost invariably $\ell nG_c\propto -T^{-1/2}$. This is argued to result from interpolating the high-temperature activation behavior and the low-temperature $-\ell nG_c\propto T^{-1/4}$ behavior [Sheng and Klafter, 1983] as can be illustrated by numerical simulation [Chen, Sheng, Abeles and Zhou, 1990] described below.

Since the charging energy of a metal grain, $E_c = e^2/\kappa d$, is inversely proportional to the diameter d of a grain, a distribution of grain sizes will be directly reflected in the distribution of charging energies. In particular, if d is log-normal distributed as has been experimentally measured [Granqvist and Buhrman, 1976], then E_c will have the same distribution:

$$N(E_c)=\frac{1}{\sqrt{2\pi}\,\mu E_c}\exp[-\frac{\ell n^2(\frac{E_c}{E_o})}{2\mu^2}], \qquad (13)$$

where μ is the width of the distribution and E_o is the median charging energy determined by the median grain size d_o. As can be seen from Fig. 4, this distribution has no density at $E_c=0$. Thus there is a natural "Coulomb gap". However, as pointed out by Adkins [1986], there can be "random potentials" in the material which can locally shift the gap position relative to the overall chemical potential, or Fermi level, of the system. Yet for each grain, this shift can not place the energy of the grain more than $\pm E_c$ away from the Fermi level since otherwise it would become advantageous to auto-ionize by one electronic charge and move closer to the Fermi level. If we consider only the energy for adding an electron to a neutral grain, the above argument means that the random potential can place the energy level of a grain anywhere between 0 and $2E_c$, where 0 denotes the Fermi level. Figure 5 shows a plausible distribution function F(E) for the smearing effect of the random potentials. Instead of a uniform distribution, which would involve discontinuities at $\pm E_c$, a more smoothly varying function is chosen that is given by

$$F(E)=N^{-1}\begin{cases}\tanh 10(E/E_c+1) & -E_c\leq E\leq 0\\ \tanh 10(1-E/E_c) & 0<E\leq E_c\\ 0 & \text{otherwise}\end{cases}, \qquad (14)$$

where N is a normalization constant. The actual energy distribution function P(E) is thus a convolution of $N(E_c)$ and F(E), i.e.

$$P(E)=\int_0^\infty N(E_c)\,F(E-E_c)\,dE_c\,. \qquad (15)$$

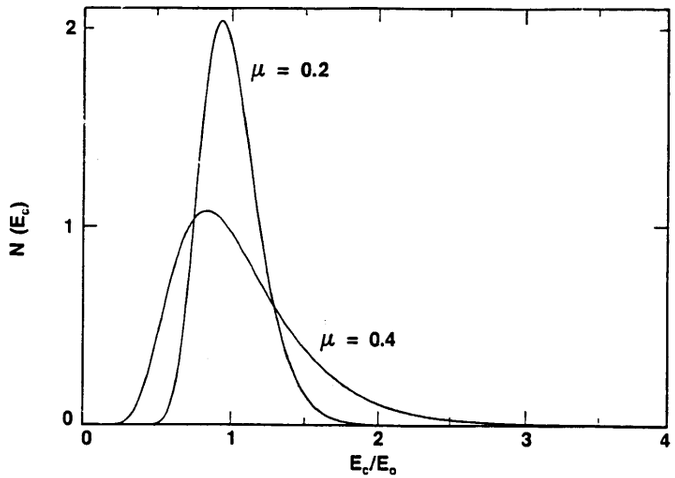


Figure 4. The log-normal distribution for two values of μ, which represents the width of the grain size distribution. Note the absence of particles with $E_c \sim 0$.

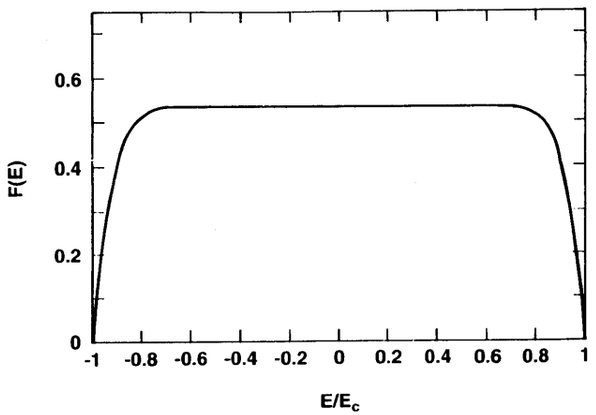


Figure 5. The effect of random potentials is represented by F(E), which vanishes at $E=\pm E_c$ and is nearly constant over $E = -0.9E_c$ to $0.9E_c$.

The result of this operation is shown in Fig. 6. It is seen that there is a small dip in the density of states near the Fermi level whose width is about 1/5-1/10 of E_o. If E_o=30 meV, the width of the dip is about 3-6 meV. Such a gap has indeed been observed through experiments on tunneling from a superconductor to a granular metal [Abeles and Sheng, 1974].

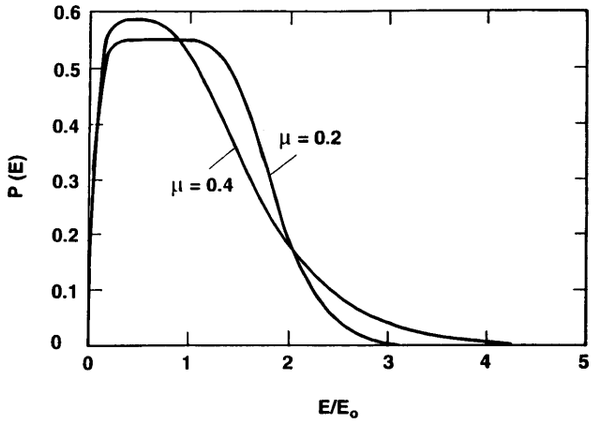


Figure 6. The effective density of states for granular metals for two different values of μ.

For calculational purposes the metal grains are mapped onto a 24x24x24 simple cubic lattice so that they can be indexed. The tunneling distance between the neighboring sites, however, is not a constant. Instead, it is given by

$$\vec{S}_{ij} = \vec{R}_{ij} + \vec{c} \tag{16}$$

where $\vec{R}_{ij}$ is the mean tunneling distance vector between sites i and j on a regular simple cubic lattice with a lattice spacing of 15Å, and $\vec{c}$ is a random vector with a uniform distribution of its magnitude between zero and 15Å. The only other parameter is μ for the energy distribution, which we choose to be 0.2. The E_o is used to renormalize the temperature scale.

By generating random S_{ij} and E_{ij} in accordance with their respective distributions, we get a network of G_{ij}'s. However, instead of connecting any two sites by a bond, we will limit ourselves to only the 26 nearest, next-nearest, and next-next-nearest neighbors for each site so that the calculation becomes manageable. In this way we also avoid those tunnelings that involve going through an intermediate grain. The resulting Kircchoff equations are solved numerically as a function of kT/E_o to obtain the effective conductance in units of G_o. The results are shown in Fig. 7. It is seen clearly that the temperature dependence is indeed accurately $-\ell nG \propto T^{-1/2}$ over more than a decade of temperature and three to four orders of magnitude in conductance [Chen, Sheng, Abeles, and Zhou, 1990]. This range corresponds well with what has been seen experimentally [Abeles et al., 1975]. However, due to the fact that tunneling is limited to only 26 neighbors, there is a finite percolation threshold, and that implies 1/T behavior at very low temperature. Otherwise the $1/\sqrt{T}$ behavior should change over to the $1/T^{1/4}$ behavior as seen experimentally [McAlister, Inglis, and Kroeker, 1984] until at the very lowest temperatures the dip in the density of states P(E) starts to have an effect. This dip is really the remnants of the charging energy gap of individual grains, and this brings us to the topic of the next section.

3.3. THE COULOMB QUASIGAP

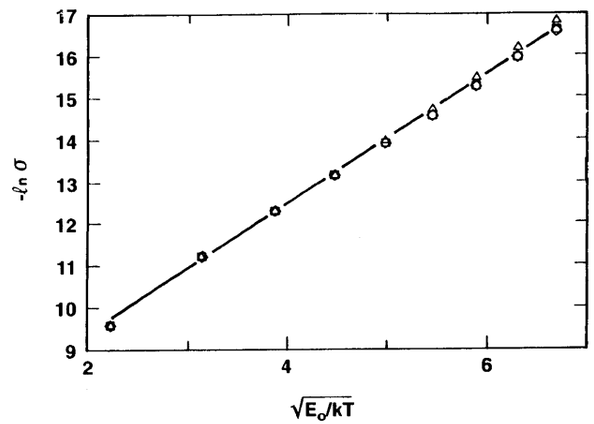


Figure 7. The calculated temperature dependence of $-\ell n\sigma$ for three randomly generated 24x24x24 conductance configurations with the same distributions for S_{ij} and E_{ij}. The value of μ used in the calculation is 0.2. Excellent linear $1/\sqrt{T}$ variation is seen.

The charging energy of a metal grain, arising from the electric field energy associated with a charged grain, is a good example of the Coulomb gap. In recent years, there have been discussions [Eatin-Wohlman, Gefen, and Shapiro, 1983; Klafter and Sheng, 1984] on the role of the Coulomb quasigap in the transport characteristics of granular metals (Due to the fact that for a macroscopic material the Coulomb gap is zero only at the Fermi level, it is more accurate to denote it as a quasigap). In particular, the role of Efros-Shklovskii's (ES) correlation gap [Efros and Shklovskii, 1975] has been the focus of these discussions. Is the Coulomb quasigap observed in tunneling experiments involving granular metals the ES correlation gap? This section is devoted to the clarification of this question.

There are two types of Coulomb quasigap. To see this fact, consider two neutral metal grains separated by a distance r_{12} that is >> their diameters. When a negative charge is removed from one grain and placed on the second grain, the energy stored in the resulting electric field can be separated into three components. The first two components are associated with the charging energies of the two grains, each considered individually without the presence of the other. The two charging energies are $e^2/\kappa C_1$ and $e^2/\kappa C_2$, respectively, where C_1 and C_2 are the capacitances of the two grains. The third component of the energy, $-e^2/\kappa r_{12}$, is negative and arises from the interaction of the two oppositely charged grains. It has to be emphasized here that the Coulombic form of the interaction energy is only valid in granular systems where r_{ij} is large. When r_{ij} is comparable to grain separation, the interaction is very much dependent on the geometry of the grains and their closest approach. There is no single, nice form for the interaction energy in that case. Moreover, the Coulombic form is almost certainly wrong for r_{ij}~d since it does not take the screening effect of the finite grains into consideration. Let us now suppose r_{ij}>>d, then the electric-field energy of a collection of N grains may be written as [Chui, 1991]

$$E=\sum_i \epsilon_i n_i^2 + \sum_{i\neq j}\sum t_{ij} n_i n_j, \tag{17}$$

where n_i=0 or $\pm$1 denotes the number of excess or deficit charges on a given grain, $\epsilon_i = e^2/\kappa C_i$ is the charging energy of grain i, and $t_{ij} \simeq e^2/\kappa r_{ij}$ denotes the interaction energy between grain i and grain j. The ϵ_i may be denoted as the site energy, or the diagonal energy. Its characteristic is that it depends only on the n_i and C_i of a single grain. The t_{ij} may be denoted as the correlation energy, or the off-diagonal energy. Its characteristic is that it depends on the separation and the n_i, n_j of two grains.

The Coulomb gap in granular metals arises mainly from the distribution of the site energy ϵ_i, i.e. $P(\epsilon_i)$ is non-uniform and actually exhibits a dip at the Fermi level. Such gap is denoted as the site energy gap, or the diagonal gap. Efros and Shklovskii [1975] have shown that in doped semiconductors, even if the site energy distribution is uniform in the absence of interaction, there can still be a dip in the density of states by turning on the correlation energy t_{ij}. This type of gap is denoted as the correlation gap. To see how this can come about, consider the simplest case of the states near the Fermi level at T=0. The existence of interaction t_{ij} directly implies a stability criterion of

$$\epsilon_i-\epsilon_j-t_{ij} \geq 0, \text{ for } \epsilon_i>\epsilon_j . \tag{18}$$

Since otherwise a lower-energy state can be achieved by moving an electron from j to i, giving rise to a positively changed site and a negatively charged site. Now t_{ij} is recalled to be $e^2/\kappa r_{ij}$; therefore

$$\epsilon_i - \epsilon_j \geq \frac{e^2}{\kappa r_{ij}}. \tag{19}$$

If $\Delta\epsilon = \epsilon_i-\epsilon_j$ is very small, i.e. both ϵ_i and ϵ_j are close to the Fermi level, then the inequality (19) implies that r_{ij} must be very large. Therefore, given an energy state close to the Fermi level, another state similarly close to the Fermi level has to be a large distance away. In other words, the density of states for those energy states close to the Fermi level must be very low. In particular, it has to vanish at the Fermi level since $r_{ij} = \infty$ in that case. We thus see that by turning on the interaction t_{ij}, the energy levels must be re-organized in such a way so as to satisfy the stability criterion, which directly implies a correlation gap in the density of states. Since the density of states ρ is given by $1/(\Delta\epsilon\Delta V)$, then heuristically

$$\rho = \frac{1}{\Delta\epsilon\Delta V} \sim \frac{1}{\Delta\epsilon (r_{ij})^3} < \frac{\kappa^3(\Delta\epsilon)^3}{e^6\Delta\epsilon} = \frac{\kappa^3(\Delta\epsilon)^2}{e^6}. \tag{20}$$

What Eq.(20) tells us is that the final density of states must be bounded by $4\kappa(\epsilon)^2/e^6$ if one takes $\Delta\epsilon \sim 2\epsilon$.

Suppose a system has a quadratic gap (making Eq.(20) an equality). What is the temperature dependence of the conductivity? If we use the variable-range hopping argument heuristically, one should optimize the exponent of G_{ij}, i.e.

$$-\ell n G_{ij} \propto \frac{E_{ij}}{kT} + 2\chi r_{ij} \tag{21}$$

of the hopping conductance should be optimized. Now E_{ij} may be estimated as the average level separation in a sphere of radius r_{ij}, i.e.

$$E_{ij} \simeq \left[\frac{4\pi}{3} r_{ij}^3 \rho\right]^{-1} = \left[\frac{4\pi}{3} r_{ij}^3 \frac{\kappa^3}{e^6} \epsilon^2\right]^{-1}. \tag{22}$$

Since from Eq.(20) $\epsilon = \frac{e^2}{\kappa}(r_{ij})^{-1}$, we get from Eq.(22)

$$E_{ij} \simeq \left[\frac{4\pi}{3} r_{ij} \frac{\kappa}{e^2}\right]^{-1} = \frac{3}{4\pi} \frac{e^2}{\kappa r_{ij}}. \tag{23}$$

Substituting Eq.(23) into the expression (21), and setting the derivative with respect to r_{ij} equal to zero yield the optimal hopping distance $(r_{ij})_o$:

$$(r_{ij})_o = \left(\frac{e^2}{\kappa} \frac{1}{2\chi} \frac{3}{4\pi} \frac{1}{kT}\right)^{1/2}. \tag{24}$$

Using this $(r_{ij})_o$ in Eqs. (23) and (21) gives

$$-\ell n\sigma \simeq \sqrt{T_o / T}, \tag{25}$$

where

$$T_o = \frac{6\chi}{4\pi} \frac{e^2}{\kappa}. \tag{26}$$

Therefore, a quadratic gap around the Fermi level would imply a $\alpha = 1/2$ behavior for the conductivity, similar to that observed in granular metals. There are, however, some difficulties in applying this correlation-gap concept to explain the temperature dependence of the granular-metal conductivity. First of all, if we put numbers into Eq.(24) to estimate the optimal hopping distance, letting $\kappa = 10$, $\chi = 1\text{Å}^{-1}$, and $k = 1.4 \times 10^{-16}$ erg K^{-1}, we get

$$(r_{ij})_o \simeq \frac{40}{\sqrt{T}} \text{Å}, \tag{27}$$

where T is in K. For T = 10-300K, which is the range where the $\alpha = 1/2$ behavior is most widely observed, we get the optimal tunneling distance of about a nanometer or less. That means the Coulombic form of the interaction energy, which is crucial to the quadratic form of the correlation gap, is not valid due to the finite grain size and its screening effect. Thus, there is an inconsistency. However, if T = 0.01K, then $(r_{ij})_o$ would be on the order of 400Å, a distance range where the Coulombic form of the interaction energy could be valid. Therefore, in order to see the effect of the correlation gap one needs to go to very low temperatures. In the range of 10-300K one is thus forced to the interpolation interpretation, discussed earlier, for the explanation of the $-\ell n\,\sigma \propto 1/\sqrt{T}$ behavior.

What happens if the site energy ϵ_i starts out with a density of states ρ that already satisfies the bound given by Eq.(20)? In that case Eq.(20) will not add anything more since the system is already stabilized by the diagonal Coulomb gap. For the example of two grains, each grain has a sharp site-energy gap at $e^2/\kappa C_{1,2}$. If the Fermi level in each grain coincides with the chemical potential, then the system of two neutral grains is always stable against charge fluctuations, i.e. the neutral grains are always the lowest-energy state because of the stabilization provided by the site energies. A similar statement can be made for the general N-particle system.

For a log-normal distribution of grain sizes, Fig. 4 shows a natural site-energy quasigap that satisfies the bound of Eq.(20). Therefore site-energy quasigap dominates. With the addition of random potential to the granular system, which might arise due to the existence of donors or acceptors in the material, the gap is less pronounced, as seen in Fig. 5. However, as long as the smearing effect of the random potential does not completely wipe out the site-energy gap, the site-energy still provides the dominant stabilization component. As can be seen from the simulation results, in this case the characteristic $-\ell n\,\sigma \propto T^{-1/2}$ temperature dependence arises mainly as an interpolation effect rather than as a manifestation of the shape of the Coulomb correlation gap.

4. Concluding Remarks

From the discussion of the previous sections it is clear that there is no single theory for the explanation of all the transport phenomena observed in granular metals. This is due to the complexity of the material, which is also the source of a rich variety of optical, electrical, and superconducting characteristics that has been utilized for a diverse array of applications [Cody, Geballe, and Sheng, 1990]. While a tremendous amount has been learned about granular metals in the past two decades, yet much more remains to be studied, both in the context of quantum wave nature of the grain properties as well as in the context of macroscopic characteristics. Better experimental control on the grain size and on the microstructure of granular metals will undoubtedly further unlock many new properties now obscured by the broad grain size distribution and the short correlation length of the associated microstructure.

5. References

Abeles, B., 1976, Appl. Solid Science 6, 1.

Abeles, B., Sheng, P., 1974, Low Temperature Physics - LT13, Vol. 3, edit. by Timmerhause, O'Sullivan and Hammel (New York: Plenum Press), p. 378.

Abeles, B., Sheng, P., Coutts, M. D., and Arie, Y., 1975, Adv. Phys. 24, 407.

Adkins, C. J., 1986, J. Phys. C: Solid State Phys. 20, 235.

Ambegaokar, V., Halperin, B. I., and Langer, J. S., 1971, Phys. Rev. B4, 2612.

Beamish, J. R., Patterson, B. M., and Unruh, K. M., 1990, in Physical Phenomena in Granular Materials, MRS Symposium Proceedings vol. 195, Cody, G. D., Geballe, T. H., and Sheng, P., edits, (MRS, Pittsburgh, Pennsylvania, 1990), p. 129.

Bruggeman, D. A. G., 1935, Ann. Phys. (Leipzig) 24, 636.

Chen, L. F., Sheng, P., Abeles, B., and Zhou, M. Y., 1990, in Physical Phenomena in Granular Materials, MRS Symposium Proceedings Vol. 195, Cody, G. D., Geballe, T. H. and Sheng, P., edits., (MRS, Pittsburgh, Pennsylvania 1990), p. 187.

Chui, S. T., 1991, Phys. Rev. B43, to appear.

Efros, A. L., and Shklovskii, B. I., 1975, J. Phys. C8, L49.

Entin-Wohlman, D., Gefen, Y., and Shapiro, Y., 1983, J. Phys. C: Solid State Phys. 16, 1161.

Granqvist, C. G., and Buhrman, R. A., 1976, J. Appl. Phys. 47, 2200.

Gudden, B., and Schottky, W., 1935, Z. Tech. Phys. 16, 323.

Hofstadter, D. R., 1976, Phys. Rev. B14, 2239.

Imry, V., 1986, Europhys. Lett. 1, 249.

Klafter, J., and Sheng, P., 1984, J. Phys. C: Solid State Phy. 17, L93-96.

Landauer, R., 1952, J. Appl. Phys. 23, 779.

Lax, M., 1951, Rev. Mod. Phys. 23, 287.

MacMillan, M. F., Devaty, R. P., and Mantese, J. V., 1991, Phys. Rev. B43, 13838.

Maxwell-Garnett, J. C., 1904, Phil. Trans. Roy. Soc. 203, 385.

McAlister, S. P., Inglis, A. D., and Kroeker, D. R., 1984, J. Phys. C17, L751.

Miller, A., and Abrahams, E., 1960, Phys. Rev. 120, 745.

Mott, N. F., 1968, J. Non-Cryst. Solids 1, 1.

Neugebauer, C. A., and Webb, M. B., 1962, J. Appl. Phys. 33, 74.

Scher, H., and Zallen, R., 1970, J. Chem. Phys. 53, 3759.

Shante, V. K., 1973, Phys. Lett. 43A, 249.

Sheng, P., 1980, Phys. Rev. Lett. 45, 60.

Sheng, P., Abeles, B., and Arie, Y., 1973, Phys. Rev. Lett. 31, 44.

Sheng, P., Klafter, J., 1983, Phys. Rev. B27, 2583.

Sheng, P. and Zhang, Z. Q., 1991, J. Phys.: Cond. Matter 3, 4257.

Zhang, Z. Q. and Sheng, P., 1991, Phys. Rev. B, B124, 3304.