*Proc. Natl. Acad. Sci. USA*
Vol. 95, pp. 7264–7269, June 1998
Physics

# Broken symmetry and strangeness of the semiconductor impurity band metal-insulator transition

**(pseudogap/filament/multiple phase transitions)**

J. C. PHILLIPS†

Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974-0636

*Contributed by J. C. Phillips, March 6, 1998*

**ABSTRACT** **The filamentary model of the metal-insulator transition in randomly doped semiconductor impurity bands is geometrically equivalent to similar models for continuous transitions in dilute antiferromagnets and even to the λ transition in liquid He, but the critical behaviors are different. The origin of these differences lies in two factors: quantum statistics and the presence of long range Coulomb forces on both sides of the transition in the electrical case. In the latter case, in addition to the main transition, there are two satellite transitions associated with disappearance of the filamentary structure in both insulating and metallic phases. These two satellite transitions were first identified by Fritzsche in 1958, and their physical origin is explained here in geometrical and topological terms that facilitate calculation of critical exponents.**

The topological structure of the quantum metal-insulator transition (MIT) in the limit $T \rightarrow 0$ in a disordered system (such as a semiconductor impurity band) differs from a classical random resistor network (1, 2) in two fundamental ways. In the quantum case, there are strong Fermion interference effects, and the long range Coulomb interactions cannot be neglected. Both of these effects alter the topological dynamics of the quantum system relative to the classical one. Earlier papers (3, 4) on the metallic critical region showed that proper treatment of the quantum case produces excellent agreement with experiment, notably on Si:P at very low temperatures both very close to, and quite far from, the transition (5) and on neutron-transmutation doped, isotopically pure Ge:Ga (6), in which the uncompensated dopants are truly randomly distributed. This theory is successful because it emphasizes the filamentary aspects of the phase transition in this dilute system, instead of making the effective medium approximation (EMA). The latter apparently has been responsible for many earlier failures (7, 8) to describe observations of critical metallic behavior and for a long time has been thought (9) to be inadequate for treating critical quantum behavior of dilute systems.

The filamentary theory just described is radically different from the overwhelmingly popular theories (7, 8) based on the EMA. Although most workers in this field still hope that the latter can be repaired somehow, one should note that more than 40 years have passed since research began on this subject, and the problem remains unsolved. Thus, all theories based on the EMA are unable to explain why, at the electrical MIT there is apparently no thermal transition, and they are unable to explain why the metallic side of the electrical transition is so sharp. During these 4 decades many different approaches have been tried. Most of them have some merit, at least for fitting curves derived from experiment, and one is reminded of the Indian fable of the seven blind men and the elephant, except that here the number of viewpoints that is partially correct is probably closer to 70 than it is to seven.

Why is this problem so complex? I believe that the problem is not only multi-faceted, as implied by analogy with the Indian fable, but also multi-layered. Because of disorder and long range Coulomb interactions and because the problem concerns charge transport rather than just thermal equilibrium properties, it contains new kinds of broken symmetries that cannot be described only in spectral (equilibrium) EMA terms. Moreover, although some aspects of the problem can be treated with the techniques of continuum mathematics, such as Fermi liquid theory, still others (especially the new broken symmetries needed to explain the sharpness of the electrical transition and the existence of separate thermal and transport transitions) require the methods of discrete mathematics. The interplay between these two layers of complexity, continuous and discrete, produces the multifaceted, and often seemingly contradictory (from the viewpoint of continuum models only), phenomena that are observed experimentally.

To gain an overview of this problem, one can begin by discussing the multiple transition model of Fritzsche (10, 11), which contains, in addition to (II) the central MIT, satellite transitions on both the insulating (I) and metallic (III) sides. This heuristic spectral model was formulated within the EMA, but it turns out to have much more general significance and, despite its antiquity, still holds the keys to many more recent puzzles. Because emphasis in more recent theories and experiments has been placed on treating the MIT as if it were analogous to single critical magnetic transitions, it is useful to review what is known about the latter, especially in the dilute case with spatial disorder (12). The new electric broken symmetry is best understood through analogy with the mechanical stiffness transition of a disordered harmonic oscillator network (13, 14). Given this background, the reader easily should follow several finite scaling calculations, which are apparently similar to conventional scaling theory but which contain several surprises.

**Fritzsche's Model.** According to Fritzsche's heuristic spectral model (10, 11), the conventional MIT (II) occurs at dopant density $n = n_c$, a satellite insulating transition (I) occurs at $n = n_h$, and a satellite metallic transition (III) occurs at $n = n_{cb}$, as shown in Fig. 1. This paper develops the theme that for $n_h < n < n_{cb}$, the dominant topological structure is filamentary, which some would call strange. At $n = n_c$, the Fermi energy $E_F$ coincides with Mott's mobility edge at $E = E_C$, and $E_C$ lies below the extrapolated conduction band edge at $E = E_O$. One can think of $E_O$ as the conduction band edge in the EMA. For $n_h < n < n_c$, we have a strange insulating phase where conduction takes place through a mixture of metallic conduction in clusters and many-electron Coulomb hopping or tun-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



Abbreviations: MIT, metal-insulator transition; EMA, effective medium approximation.

†To whom reprint requests should be addressed. e-mail: jcp@physics.lucent.com.

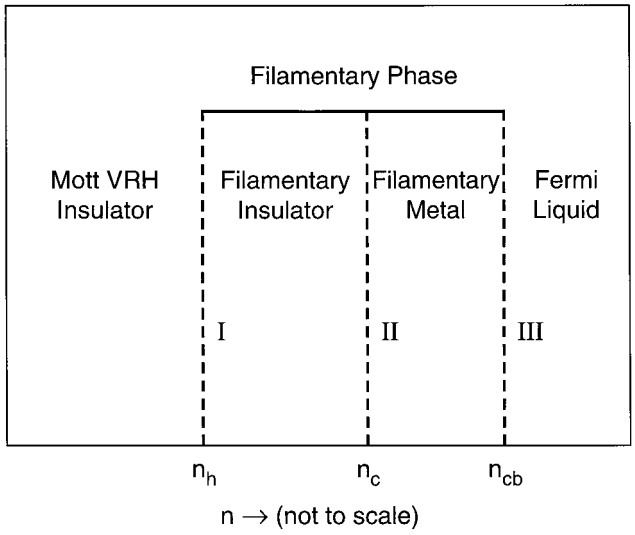


FIG. 1. A schematic topological phase diagram for the semiconductor impurity band MIT in the limit $T \to 0$, based on Fritzsche's heuristic spectral phase diagram (10, 11).

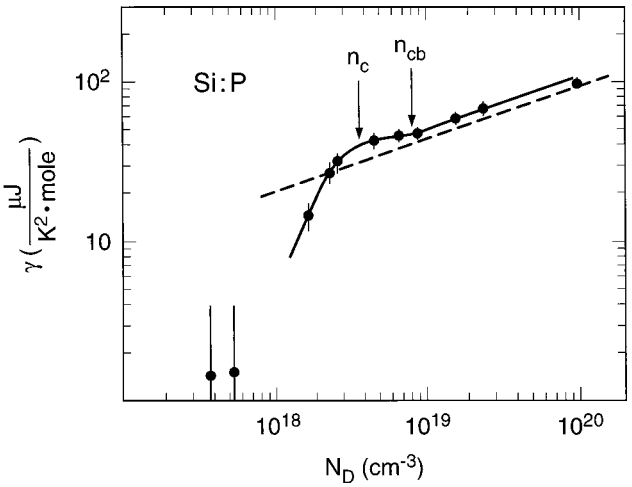


FIG. 2. The electronic-specific heat coefficient $\gamma(n)$ in Si:P. The experimental data are from ref. 16, as is the dashed line, representing noninteracting electrons in a rigid band with the same effective mass value as in the pure crystal. The solid line is a smooth curve drawn through the experimental points, except that a break in slope has been placed at $n = n_{cb}$ for the reasons discussed in the text. Note that for $n > n_{cb}$, the observed specific heat is parallel to, but larger than, the expected crystalline value, which corresponds to mass enhancement, which can be attributed to electron–electron interactions that drive the metallic carriers in extended states to blob and filament internal surfaces, where the internal electric fields are least screened. The plateau between $n_c$ and $n_{cb}$ has a quasi-spinodal character that represents the latent first-order nature of the phase transition. This first-order effect has been retained in the continuous phase transition because of its fractal nature (3).

neling between clusters, whereas for $n_c < n < n_{cb}$, we have the strange (non-Fermi liquid) metallic phase mentioned in the title of this paper. Both $E_C$ and $E_O$ are now regarded by many field and scaling theorists as archaic heuristic constructs without a firm microscopic basis, but $n_c$ and $n_{cb}$ are revived here and defined rigorously in terms of the filamentary model and quantum percolation theory; as we will see, in the new context, $E_C$ and $E_O$ retain much of the physical content envisioned by Fritzsche and Mott. Just as a transport phase transition occurs at (II) $n = n_c$, so a thermal phase transition occurs at (III) $n = n_{cb}$; II is sharp, and in the limit $T \to 0$, its asymptotic (finite size scaling) critical behavior is only slightly affected by the presence of background (oxygen) impurities, whereas III corresponds to a local order–disorder transition that is broadened greatly by random dopant density fluctuations, which have no effect on the critical exponents of II.

One of the key assumptions of scaling theories (8, 15) of the MIT is that $n_c = n_{cb}$ and that there is only one phase transition; this assumption implies that the thermodynamic coherence length $\xi$ should be identified with the transport mean free path l. However, in the limit $T \to 0$, by analogy with conventional Boltzmann transport theory (or Fermi liquid theory), one would expect that, in a true metal, the residual resistance and l would be determined by scattering from background impurities or nonrandom density fluctuations in the dopant impurities, with an average spacing $L_b$, not from random density fluctuations in the dopant impurities or the electronic charge density, which determine $\xi$. Thus, there is no reason to suppose that $n_c = n_{cb}$, and experimentally, the electronic specific heat coefficient $\gamma(n)$ appears to be a smooth function of n at $n = n_c$ (16), as shown in Fig. 2. However, the data do show an unambiguous plateau in $\gamma(n)$ for $n \approx (1–2)\ n_c$, which can be identified with $n_{cb} = 2\ n_c$. As we shall see later, the filling factor associated with the strange percolative metallic phase at $n = n_c$ is small, and this factor flattens $\gamma(n)$ in the region $n_c < n < n_{cb}$ so much that it is difficult to determine $n_{cb}$ from the small change in slope of the specific heat at $n_c = n_{cb}$. Fritzsche was well aware of this point, and he therefore attempted to estimate $n_{cb}$ from the (also broad, but not quite flat) maximum in the Hall mobility, which gave $n_{cb} \approx 6n_c$. In any event, the plateau (16) in $\gamma(n)$ for $n > n_c$ is fundamentally inconsistent with the scaling assumption (8, 15) that the thermal and transport phase transitions are the same. Finally, the difference in screening of internal electric fields in the filamentary and Fermi liquid phases is apparently the origin of the 2% first-order jump in the Debye $\Theta$ at $n = n_{cb}$; this point is discussed below in connection with the phase diagram.

**Critical Metallic Conductivity Exponent.** Although neither Fermi liquid theory (7) nor scaling theory (8, 15) has been able to describe simultaneously the transport and thermal properties of the MIT, they are well established despite these failings because they are accompanied by elaborate (but not internally consistent) mathematical formulae that can be, and often are, used for curve fitting of experimental data, thereby creating respectable facades. In earlier papers (3, 4), I focused my attention primarily on demonstrating that a filamentary model based on quantum percolation can explain why the transport MIT is continuous, rather than first-order, as it must be in the EMA (compare the Wigner transition of electrons in a box), with a critical metallic conductivity $\sigma_M(x, T)$ described by $\sigma_M(x,0) \propto x^{\mu}$, where $x = n/n_c - 1$ and $\mu = \frac{1}{2}$. This value of $\mu$ is surprisingly small; in liquid He, which has Bose statistics and short range interactions, it is 2/3 at the thermal $\lambda$ transition, and this value represents the lowest value acceptable to finite-size scaling theories that require $\mu \geq 2/3 = 2/d$, so that the random fluctuations in a volume of order $\xi^d$ can be internally consistent (17, 15). Some readers have requested more justification for the assumptions made in the derivation (3) of $\mu = \frac{1}{2}$, and this justification will be given here.

First, we note that the essentially percolative filamentary character of dilute phase transitions is well established for critical magnetic transitions (12), so that there is nothing strange or radical about analyzing the MIT from this viewpoint; indeed, quite the opposite is true, as in the EMA the MIT should be first-order, whereas percolative transitions are continuous by construction. The differences from the dilute, short range Bose antiferromagnetic case [where also (12) $\mu = 2/3$] must arise from either or, more likely, both many-electron Fermi statistics and long range Coulomb interactions. To analyze these differences quantitatively, one must categorize and separate localized from extended states. In ref. 3, it was

shown how this can be done in principle in the limit $T \rightarrow 0$; it then becomes possible to make a small amplitude approximation to the quasi-particle excited states.

Some readers have suggested that it is not possible in principle to identify extended states in the presence of spatial noise because no algorithm is known for doing so. This is a very general argument, which many theorists associate with the existence of special high symmetry modes (called Goldstone modes by mathematical physicists) obtained by group-theoretic arguments. Those arguments are, in fact, too general: They are valid for arbitrarily large amplitudes of some general displacements, in other words, arbitrarily high temperatures, where thermal mixing of excited states eventually destroys phase coherence in the ground state. In the quantum case this is true, but in the classical case of coupled harmonic oscillators (atomic vibrations), Thorpe *et al.* (13, 14) have shown how one can identify and count cyclical modes ($\omega = 0$) near the mechanical stiffness transition of a random network that is fully connected geometrically but not mechanically; these modes (first studied by Maxwell) have no recognizable symmetries. For the present transport problem, one should consider these modes analogously to be extended states with complex frequencies $\omega + i\Gamma = 0$, corresponding to quasiparticle states at the Fermi energy $\omega = 0$, and a scattering rate $\Gamma = 0$ in the absence of background or residual resisitivity scattering. In this way we recover the residual resistivity concept that is essential to avoid having infinite conductivity in the limit $T \rightarrow 0$ and $L_b \rightarrow \infty$. Note that in switching from a thermal equilibrium quantity (the elastic constants of the mechanical random system) to a transport property (the $T = 0$ conductivity of the random metal), we have replaced the real part of the self-energy ($\omega$) by the complex self energy $\omega + i\Gamma$. Distinguishing between $\omega$ and $\Gamma$ is similar to distinguishing the coherence length $\xi$ from the mean free path $l$, which scaling theory (8, 15) has not done so far, and it is this that will enable us to distinguish $n_{cb}$ from $n_c$.

**Variational Stability.** Another puzzle concerns the possibility of scattering between localized and extended states, and this puzzle leads directly to the recognition of a new kind of broken symmetry (3). It has been known for a long time that it is inadvisable to solve transport equations (or to attempt to identify extended states) solely through the application of boundary conditions (18), for example, in the electrical case, because this method can generate spurious one-dimensional localization; this localization can be avoided, for example, with a local equilibrium variational condition on the extended states, which assigns them an average drift velocity in the presence of an applied field. The number density of the extended states is $n_e(x)$, which goes to zero as x goes to zero. Indeed $n_e$ is robust and is not affected to first order by resonant scattering from extended to localized states because it has been maximized to yield the maximum conductivity. One can simply picture the $T = 0$ extended states as flowing along dynamically stationary one-dimensional percolative channels between background impurity scattering events.

Once one has accepted this simple physical picture, one easily can calculate $\mu$ for oneself (19); in any d*-dimensional electronic system where the carrier Fermi energy $E_F$ excess $\Delta E_F = E_F(x) - E_F(0)$ is proportional to x or to wave vector **k** squared, and $\delta\sigma(x) = \sigma(x) - \sigma(0) \approx n_e(x) \approx k^{d^*}$, $\mu = d^*/2$. Here, the presence of strong disorder generates $d^* = 1$ filaments for much the same reasons as in the magnetic case (12, 20, 21). Thus, $\mu = d^*/2 = \frac{1}{2}$ and because the filaments have been selected from the set of all many-electron wave functions by a variational condition, this result is exact.

Here many readers will be surprised that the calculation of $\mu$, which has proved impossible in EMA models (7, 8), should be so simple and easy algebraically. This simplicity is explained by two points: (*i*) The algebra leading to a simple fraction, such as $\frac{1}{2}$, must itself be simple; the complexity lies not in the algebra, but in the set-theoretic and variational concepts (3) that justify the algebra. (*ii*) Of course, I am hardly the first person to have considered this $d^* = 1$ mechanism, but it always had been thought before, in the EMA context in which extended states were not separated from localized states and there was no distinction between thermal and transport transitions, that any one-dimensional mechanism would produce a divergence of the specific heat. This question will now be discussed in greater detail than before (3), relying on standard descriptions of percolation in dilute systems and new aspects of transport scaling associated with the residual resistivity and the new scaling length $L_b$.

**Percolative Quantum Transport in the Limit $T \rightarrow 0$.** To treat the residual resistivity correctly, one begins by partitioning the sample volume into Voroni polyhedra (volume $\approx L_b^d$) centered on the background impurities (average spacing $\approx L_b$). How many filamentary paths can the average polyhedron accommodate that cross it from one side to the other? Clearly, that number is limited by the condition, characteristic of the many-Fermion ground state electron wave function, that two such paths should not cross, for if they did, one would have destructive interference. [At the center of the crossing region, the wave function intensity is proportional to $A_1A_2 \exp(i[\varphi_1 - \varphi_2])$, where $A_i$ and $\varphi_i$ represent the amplitude and phase, respectively, of channel i. Averaging over [] gives zero by the random phase approximation.] One may note here that, in the semiclassical theory of the $d = 2$ large integer quantum Hall effect, such crossing of local paths (or ribbons) at saddle points has been discussed extensively (22). In that case, one does not have an MIT in the sense that one does for $d = 3$, but one does have localization–delocalization transitions, and then the crossing points are treated as insulating tunneling volumes. The latter have a substantial effect on the scaling exponents, which agrees well with experiment.

In ref. 3, it was noted that a sufficient condition that precludes filamentary crossing is that there be no more than one filament/polyhedron. Here, this analysis is refined by scaling with respect to $L_b$. In the EMA, the filament would be straight and its length would scale with $L_b^{\kappa}$, where $\kappa = 1$. If the filament corresponded to a random walk, then $\kappa = 2$. In classical percolation theory, near the percolation threshold, the geometry of percolative paths is best described by the blob-links or blob-dendrite model (23). Within this model, $\kappa = d_{min}$ as well as the fractal dimension $d_{min}$ (d) has been calculated (24) for $d = 2$ and 3; quite generally, $d_{min}$ ($d \geq d_c^+$) $= 2$, with $d_c^+ = 6$, and of course $d_{min}$ (1) $= 1$.

The mean number of path pair coincidences per polyhedron is of order $P_c \approx L_b^{2\kappa}/L_b^d$, and the probability of no coincidences is given by the Poisson normalization factor $\exp(-P_c)$. Thus, as $L_b \rightarrow \infty$, one can have one or many coincidences according to whether $d_{min}$ is larger or smaller than $d/2$, respectively. In the latter case, the maximum number of paths $N \approx L_b^{\lambda}$ is given by $\lambda = d/2 - d_{min}$.

Now let us calculate the electronic specific heat $\gamma T$ to see whether or not it diverges as $p \rightarrow p_c^+$. Here $\gamma \approx dN/dE$, the electronic density of states, $\approx (dN/dk)/(dE/dk) \approx k^{(d^* - 1)}/k \approx k^{(d^* - 2)}$. The smallest possible current-carrying value of k is $L_b^{-1}$, and with $d^* = 1$, as $L_b \rightarrow \infty$, so apparently does $\gamma$, in accordance with the usual argument mentioned above. One must not forget, however, the noncrossing condition that gives a weighting factor $N/L_b^d = L_b^{\lambda - d}$. Then $\gamma \approx L_b^{\varpi}$, with $\varpi = 2 - d^* - d/2 - d_{min}$. With $d_{min} \geq 1$, even if $d^* = 1$, there is no specific heat divergence for $d \geq 1$, as $\varpi < 0$.

One may wonder how this argument would change in a quantum context. First, one can imagine that, in the blob regions, the connectivity or coherence might not depend only on pair connectedness, but this seems likely only to renormalize $n_c$. A more serious concern is $d_{min} \geq 1$, but note that the numerical value of $d_{min}$ (d) for $d = 3$ was not actually used. Had it been necessary to use it, one could probably obtain it by

interpolation (31) between $d_c^-$ and $d_c^+$, replacing the classical value of $d_c^+ = 6$ with the quantum value $d_c^+ = 4$.

**Conventional Scaling Theory Revisited.** The treatment used so far (3) emphasizes measurement of the electrical conductivity in an applied electrical field $\mathbf{F}_a$. There is an alternative picture that passes to the ohmic limit $\mathbf{F}_a = 0$ and describes the filamentary metal in terms of fluctuating internal electric fields $\mathbf{F}_i(\mathbf{r},t)$. This alternative raises the question of whether such fields are screened in a filamentary metal with broken symmetry as they are in a Fermi liquid with translational symmetry. This question also will be discussed below after the phase diagram has been sketched, but here it is enough to note that the filaments (the backbone in classical percolation theory) fill only a small part $\approx L_b^{d/2}$ of the sample volume occupied by the dopants and that the volatile (apparently nonscalable) blobs (23) fill the remainder $\approx L_b^d$. The average strength of the incoherent Coulomb fluctuation fields that the blobs generate is also $\approx L_b^{d/2}$, but the tail of this distribution will exceed the fixed screening capability of the filaments, so that the internal fields acting on the latter will be unscreened partially. Note that it is just in the links or dendrites that the magnitudes of the internal fields depart most strongly from their average values, so that it is there that screening is expected to break down.

One can now repeat the scaling arguments previously used (15, 17) to derive the result $\mu \geq 2/3$ for disordered three-dimensional systems with short range forces [dilute antiferromagnets (12), liquid He (17)] in the context of the $d$ = 4-dimensional space-time manifold $(d) = (d, d^*) = (\mathbf{r}, \mathbf{r}\cdot\mathbf{F}_i(\mathbf{r},t)) = (\mathbf{r}, \tau(\mathbf{F}_i(\mathbf{r},t))$. Here $\tau$ is a dynamically fluctuating local time that reflects the relative group velocities of electron wave packets accelerated by the tangential component of the internal field $\mathbf{F}_i(\mathbf{r},t))$. In this picture, the JC inequality (15, 17) becomes $\mu \geq 2/d = ½$, which is exactly right because it says that the smallest value of $\mu$ is attained when the dopant distribution is fully random. This picture also explains why there are no renormalization corrections because mean field theory is indeed valid for $d - 4 = \varepsilon = 0$. It has been stated (12) that "Unfortunately, this result cannot be checked experimentally," but this is just what the MIT does. It is remarkable that the ideas of scaling theory apply to complex two-fluid models (localized and extended many-Fermion states) simply by using the hyperspace $(d, d^*) = (d)$ construction.

Both the filamentary counting and the finite-size scaling arguments give $\mu = ½$, but the latter seems to be simpler. There is, in fact, an even simpler derivation of $d = 4$. Coherent currents flow along the paths so that $d_\parallel = 2$ (complex $\mathbf{k}_\parallel$) but are attenuated (imaginary $\mathbf{k}_\perp$) normal to the paths, $d_\perp = d - 1$. Thus $d = d_\parallel + d_\perp = d + d^* = d + 1$; all roads lead to Rome. In this derivation, the internal fields seem to have disappeared, but not so. Without some kind of electric field no coherent current flow is possible in the limit $T \to 0$.

**Phase Diagram.** The three-transition phase diagram is shown in Fig. 1. To compare Fritzsche's spectrally motivated phase diagram (10, 11), based implicitly on the EMA, with more modern data (5, 6), it is necessary to realize that not only were his data taken at temperatures $\approx$ 1K, compared with temperatures $\approx$ mK in ref. 5 but that the effective energy scale in the Ge samples used by Fritzsche and in (6) is 10 times larger than that for Si samples (5). It was shown unambiguously in ref. 5 that, with their chemically doped samples of Si:P, such low temperatures were essential to obtain reliable extrapolations of $\sigma(x, T)$ to $\sigma(x,0)$ and thus reliable values of $\mu$. However, in the neutron transmutation doped, uncompensated (isotopically pure) Ge:Ga samples of ref. 6, critical exponents very similar to those of ref. 5 are reported for $T \approx 50$ mK (equivalent to 500 mK in Si). This result is very surprising, especially when one considers the difference in energy scales between Si and Ge, but it seems to suggest that, with truly random doping, critical behavior extends not only to unexpectedly large values of composition $x \approx 1$, instead of $10^{-2}$, as already shown by ref. 5, but also to much higher values of T ($\approx$20 times larger) as well.

This kind of behavior is not what is expected from the EMA, or by comparison with magnetic (boson) critical behavior (12), but it seems to be consistent with Fermion filamentary percolation because electron–electron fluctuation effects in the filaments essentially have disappeared because of the variational nature of the filamentary selection process. If further measurements with the samples of ref. 6 at lower temperatures confirm the existence of a much wider critical temperature range than was found in ref. 5, this would be taken as very strong evidence in favor of the present variational model and against EMA models. The latter then would be seen as suitable only for qualitative discussions of the transport properties of macroscopically inhomogeneously doped samples with locally varying values of $n_c$.

Another important fundamental point can be made without detailed curve-fitting. In the Fritzsche model, the order–disorder transition III occurs at $x_{cb} \approx 2$–6 so that there is indeed ample room in "x space" for a wide filamentary critical range from $x = 0$ to $x = 1$. It was noted earlier that the specific heat transition III is very broad. The qualitative reason for this is that there is a broad distribution of volumes of Voroni polyhedra $i$, with transitions at different values of the local average density $n_i$. The tail of the distribution of Fermi-liquid polyhedra can well extend down to the region near $x = 0$, and it is possible that this tail, with a width dependent on sample dopant homogeneity, is what was responsible for the tail variations of $\sigma(x, T)$ reported in ref. 25.

All of the analysis of the filamentary phase that has been given, of course, did not prove that such a phase exists; this could be done only by calculating the total energy of this phase relative to that of a normal EMA phase, a Fermi liquid phase. As noted earlier, we assumed that the phase between transitions II and III must be a filamentary phase for two theoretical reasons: (*i*) These are what are used to describe dilute magnetic phases, as studied either experimentally (12), or in numerical simulations (20, 21), and (*ii*) in the electric case, I believe that, in the EMA, Coulomb forces always produce a first-order II transition; the observed II transition is continuous.

On the other hand, the thermal transition III is definitely a first-order transition in the enthalpy, as is shown by the gigantic ($\approx$2%) jump (36) in the Debye $\Theta$ at $n = n_{cb}$; the dopant atomic density, relative to the host, here is $2.10^{-4}$, and the dopant charge density, again relative to the host, is four times smaller. Thus, the effectiveness of the donor electrons in enhancing the lattice stiffness is $\approx$25 times larger in the Fermi liquid phase than in the filamentary phase. The origin of this huge factor is the presence of unscreened electric fields in the filamentary phase. That such fields must be present is proved strikingly by the contrast (36) in the composition dependence of the thermal Debye $\Theta$ (which shows the 2% jump) with the elastic Debye $\Theta$ (which does not). This contrast is a direct result of broken symmetry. The measured elastic constants reflect the cubic symmetry of the crystalline space group, whereas the dopant dependence of the thermal Debye $\Theta$ reflects the tetrahedral point symmetry of the P donors. The latter lacks inversion symmetry, and so the internal field can have a first-order stiffening effect in the filamentary phase that is absent in the Fermi liquid. This symmetry analysis directly confirms a key feature of the Fritzsche model, which is that the III transition occurs when $E_C$ crosses $E_0$. One can suppose that the factor 25 reflects the "spread-out" nature of the dopant electrons (binding energy 0.1 eV compared with the conduction band width $\approx$ 2.5 eV).

From what has been said above concerning the wide temperature range of the transport critical transition, I do not believe that measuring the specific heat (not a transport property) at lower temperatures than did the authors of refs.

16 and 36 would yield much new information. Indeed, this was done for a few samples (26), with the result that, in Si:P, localized spin enhancement was evident below 1K not only for $x < 0$ (insulating) but also for $x > 0$, which is interpreted not only in ref. 26 but also here as evidence for the existence of isolated blobs in the filamentary metallic phase, in other words, localized states do persist in the filamentary metallic phase, with a filling factor that becomes smaller as x increases but by no means vanishes for $x > 0$. The actual filling factor is difficult to estimate because presumably the average size of isolated blobs also increases with x, which reduces the temperature at which paramagnetic effects set in.

**Weak Localization.** The term "weak localization" is used to describe the effects in the EMA of bare electron–electron interactions on the temperature dependence of the metallic conductivity, as calculated in ref. 27. With our present phase diagram, the combination of the EMA with unscreened e-e interactions places us on the horns of a dilemma: The EMA assumption is valid only for the Fermi liquid region, $n > n_{cb}$, where the e-e interactions are screened; the interactions are unscreened only for $n < n_{cb}$, where the plane-wave basis states of the EMA, which give the Coulomb interaction the form $\approx q^{-2}$ (which is essential to obtaining $\delta\sigma \approx T^{1/2}$) are no longer correct. In other words, there is no region of the phase diagram of a dilute, randomly disordered system for which weak localization is correct. So great, however, are either (or both) the attachment to the EMA or the fear of the complications of percolative models that this model has been used very widely to discuss experimental data. In particular, it is meaningless to discuss the sign of a term that depends essentially on two mutually contradictory assumptions. It is to be hoped that, with the emergence of a new class of experiments (6), these conditions will change, especially as an alternative, nonperturbative, and internally consistent percolative explanation not only for the functional form but also for the sign changes of $\delta\sigma \approx T^{1/2}$ is available (3, 28).

**Filamentary Insulating Phase.** The filamentary insulating phase occurs between transitions I and II and for densities in the range $n_h < n < n_c$; from the data of refs. 6 and 11 on Ge:Ga, it seems that $n_h \approx n_c/20$. In the insulating phase, it is by now quite well established (6) that the Coulomb pseudogap model (29, 30) gives a good account of $\sigma_I(x, T)$ as a function of T, but not as a function of x. Can this filamentary model explain both functional dependences? Yes, providing that we recognize that the insulating behavior must come, for small x, from breaks in the dendrites or links of a blob-link percolation model (23). Thus, the pseudogap is associated with states with large amplitudes in the gap region, in other words, surface states of the blobs that are mutually overlapping in virtual link regions where the breaks have been formed. These breaks are all essentially equivalent; the break resistance can be minimized by an optimization procedure analogous to Mott's variable range hopping model, except that allowance must be made for the effects of unscreened Coulomb forces $\propto r^{-1}$ on the states near $E_F$ in these regions (29). The temperature dependence will then be given by (6, 29, 30) $\sigma_I(x, T)$ 181 $\exp(-T_0/T)^{1/2}$ with the pseudogap width $T_0$.

In refs. 29 and 30, $T_0$ is not optimized but is simply matched smoothly to a background density of states with the result $T_0 \propto (\kappa(n)\xi(n))^{-1}$, where in the spirit of the EMA the density dependencies of the dielectric constant $\kappa(n)$ and the coherence length $\xi(n)$ are to be taken from experiment. The reader will not be surprised to learn that (once again!) the EMA fails; this procedure yields $T_0 \propto x^{\alpha}$, with $\alpha = 1.5$ for Si:P, for example (6), compared with an experimental value (6) for Ge:Ga of $\alpha = 1.0$. I believe the latter value to be correct. It can be derived very simply from the filamentary model. One dopant atom will be associated with each optimized filamentary break (not zero, no break; not two, because at the weak points one is enough), and in the impurity band because of charge neutrality exactly one orbital state is associated with each missing dopant atom; thus, the number of pseudogap states is proportional to x, but it is also proportional to $T_0$, so $T_0$ is proportional to x and $\alpha = 1$ exactly. This brief yet exact analysis again shows how powerful and precise discrete variational filamentary counting methods (13, 14) can be in the transition regions compared with nonvariational continuum approximations (7, 8) based on the EMA.

**Application to d = 2.** Within the filamentary model of random metals, all of the transport properties of metals familiar to electronic theorists from the basically isotropic Bloch–Sommerfeld theory of uniform crystalline metals are retained save one: This is the d-dimensional isotropy, which is replaced by the local anisotropy of a $d^* = 1$ filament embedded isotropically in a d-dimensional space. As a consequence, the effective dimensionality becomes $d = d + d^* = d + 1$. Not only does this enable us to understand why there are three phase transitions (I-III) in the electrical case, compared with only one in the magnetic case, but it also explains why $\mu(d = 3) = \mu(d = 4) = \frac{1}{2}$ instead of 2/3, as expected from arguments (15, 17) based on short range forces.

In the case $d = 2$, many old arguments (32, 33) suggested that a true MIT was not possible because it was not possible to form metallic states ($d = 2$ is marginal). In very high quality depletion layers, however, indications of an MIT have been observed (34). Because such a transition is not consistent with the old ideas, it was suggested (34) that the unexpected coherence might be caused by superconductivity. A much less exotic explanation is based on the new filamentary broken symmetry discussed here, especially because $d = 3$ is not marginal. The unscreened surface fields $F_a$ induce the formation of metallic states; this formation is possible when $F_a \gg F_r$, where $F_r$ is a surface roughness field associated with dopant nonrandomness.

**Compensated Samples.** In chemically doped samples, one expects significantly nonrandom effects from chemical association of impurities so that to study randomly doped compensation one should use neutron-transmutation doped samples. In that case, the effects of compensation are primarily to disrupt metallic ballistic coherence and replace quantum conductivities by diffusive conductivities, much as one does in uncompensated samples at high temperatures. One also, in effect, greatly increases the level of background impurity densities, so that by the Voroni partitioning construction, the average number of dopants/polyhedron is no longer $\gg 1$. The effect of the crossover from uncompensated to compensated percolation on critical exponents is discussed in ref. 19. The qualitative effects of this crossover on $n_c$ and $n_{cb}$ are discussed in ref. 35.


1. Stephen, M. J. & Grest, G. S. (1977) *Phys. Rev. Lett.* **38,** 567–570.
2. Golden, K. M. (1997) *Phys. Rev. Lett.* **78,** 3935–3938.
3. Phillips, J. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10528–10531.
4. Phillips, J. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10532–10535.
5. Thomas, G. A., Paalanen, M. A. & Rosenbaum, T. F. (1983) *Phys. Rev. B* **27,** 3897–3900.
6. Itoh, K. M., Haller, E. E., Beeman, J. W., Hansen, W. L., Emes. J., Reichertz, L. A., Kreysa, E., Shutt, T., Cummings, A., Stockwell, W., *et al.* (1996) *Phys. Rev. Lett.* **77,** 4058–4061.
7. Lee, P. A. & Ramakrishnan, T. V. (1985) *Rev. Mod. Phys.* **57,** 287–337.
8. Belitz, D. & Kirkpatrick, T. R. (1994) *Rev. Mod. Phys.* **66,** 261–380.
9. Landauer, R. (1978) in *Electrical Transport and Optical Properties of Inhomogeneous Media*, eds. Garland, J. C. & Tanner, D. B. (Am. Inst. Phys., New York), Vol. 40, pp. 2–43.
10. Fritzsche, H. (1958) *J. Phys. Chem. Sol.* **6,** 69–80.
11. Fritzsche, H. (1978) in *The Metal-Non-Metal Transition in Disordered Systems,* eds. Friedman, L. R. & Tunstall, D. P. (Scottish Univ. Sum. Sch. Phys., Edinburgh), pp. 193–238.
12. Collins, M. F. (1989) *Magnetic Critical Scattering* (Oxford Univ. Press, Oxford), pp. 14, 23, 29, 37, 60, 167.

13. He, H. & Thorpe, M. F. (1985) *Phys. Rev. Lett.* **54,** 2107–2110.
14. Jacobs, D. J. & Thorpe, M. F. (1996) *Phys. Rev. Lett.* **75,** 4051–4054.
15. Chayes, J. T., Chayes, L., Fisher, D. S. & Spencer, T. (1986) *Phys. Rev. Lett.* **57,** 2999–3002.
16. Kobayashi, N., Ikehata, S., Kobayashi, S. & Sasaki, W. (1977) *Solid State Commun.* **24,** 67–70.
17. Josephson, B. D. (1966) *Physics Lett.* **21,** 608–609.
18. Erdos, P. & Haley, S. B. (1969) *Phys. Rev.* **184,** 951–967.
19. Phillips, J. C. (1992) *Phys. Rev. B* **45,** 5863–5867.
20. Lubensky, T. C. (1977) *Phys. Rev. B* **15,** 311–314.
21. Stanley, H. E., Birgenau, R. J., Reynolds, P. J. & Nicoll, J. E. (1976) *J. Phys.* **C9,** L553–L560.
22. Bratberg, I., Hansen, A. & Hauge, E. H. (1997) *Europhys. Lett.* **37,** 19–24.
23. Herrmann, H. J. & Stanley, H. E. (1984) *Phys. Rev. Lett.* **53,** 1121–1124.
24. Herrmann, H. J. & Stanley, H. E. (1988) *J. Phys. A* **21,** L829–L833.
25. Rosenbaum, T. F., Thomas, G. A. & Paalanen, M. A. (1994) *Phys. Rev. Lett.* **72,** 2121–2121.
26. Bhatt, R. N., Paalanen, M. A. & Sachdev, S. (1989) *J. Physique* **49(C8),** 1179–1184.
27. Altshuler, B. L. & Aronov, A. G. (1981) *Solid State Commun.* **38,** 11–14.
28. Phillips, J. C. (1991) *Europhys. Lett.* **14,** 367–371.
29. Efros, A. L. & Shklovskii, B. I. (1975) *J. Phys. C* **8,** L49–L51.
30. Shklovskii, B. I. & Efros, A. L. (1984) in *Electronic Properties of Doped Semiconductors*, Solid State Series 45.
31. Stanley, H. E. (1996) in *Fractals and Disordered Systems,* eds. Bunde. A. & Havlin, S. (Springer, New York), pp. 1–57.
32. Abrahams, E., Anderson, P. W., Licciardello, D. C. & Ramakrisnan, T. V. (1979) *Phys. Rev. Lett.* **42,** 673–676.
33. Phillips, J. C. (1983) *Solid State Commun.* **47,** 191–193.
34. Simonian, D., Kravchenko, S. V., Sarachik, M. P. & Pudalov, V. M. (1997) *Phys. Rev. Lett.* **79,** 2304–2307.
35. Fritzsche, H. (1980) *Phil. Mag. B* **42,** 835–844.
36. Keyes, R. W. & Kobayashi, N. (1978) *Solid State Commun.* **27,** 63–64.