Electron Tunneling Study of Coulomb Correlations across the Metal-Isulator Transition in Si:B

J. G. Massey and Mark Lee
Department of Physics, University of Virginia, Charlottesville, Virginia 22903
(Received 11 June 1996)

Electron tunneling gives an empirical description of how Coulomb correlations evolve across the metal-insulator transition (MIT) in Si:B. For clearly insulating or metallic samples, the tunneling conductance displays a parabolic or a square-root shape, respectively. Just below the MIT, the conductance shows some metallic features before the samples become truly metallic. Very close to the MIT, the conductance spectra are unusually broad and frequency dependent. We interpret this last feature as a result of long-wavelength screening arising from a divergence of the dielectric constant. [S0031-9007(96)01403-2]

PACS numbers: 71.30.+h, 73.20.Fz, 73.40.Gk

The electronic properties of disordered solids near a metal-insulator transition (MIT) are a common and important manifestation of a general problem in modern physics, an interacting many-particle system. Because disorder localizes electron wave functions, the charges cannot screen their Coulomb interactions as effectively as in ordinary metals. Thus in all real systems the effects of disorder and Coulomb correlations are intertwined. Scaling theories give an essentially complete understanding of the MIT in the noninteracting limit, and Coulomb interactions have been treated with some success as a perturbation, particularly on the metal side [1]. However, a description of the complete transition from insulator to metal incorporating both disorder and Coulomb correlations on equal footings remains a fundamental unsolved problem in the physics of many-particle systems. Such an understanding is impeded by the failure near the MIT of many Fermi liquid simplifications. New theoretical approaches are being constructured to confront this issue directly [2], but these have yet to yield testable results and are hampered by a lack of adequate empirical guidance.

Quite generally, Coulomb interactions among electrons in a disordered potential lead to a depletion in the single-particle density of states $N(\varepsilon)$ near the Fermi energy ε_F, relative to the noninteracting case [3]. Explicit calculations of $N(\varepsilon)$ near ε_F exist for disordered solids that are clearly on either the insulating or the metallic sides of the MIT. On the metallic side, Altshuler and Aronov (AA) [4] and McMillan [5] showed that, in three dimensions, exchange correlations lead to a square-root singularity: $N(\varepsilon) = N(0)[1 + |\varepsilon/\Delta|^{1/2}]$, where the correlation energy parameter $\Delta \approx \hbar D/\ell^2$, D being the diffusion constant and ℓ the mean free path. Here and throughout this paper ε_F defines the zero of energy. On the insulating side, Efros and Shklovskii (ES) [6] argued that the Hartree part of the Coulomb interaction in a hopping conductor leads to a parabolic gap at 0 K: $N(\varepsilon) = (3/\pi)(\kappa/e^2)^3\varepsilon^2$, with a gap width $\Delta_c = e^3(N_0/\kappa^3)^{1/2}$, where κ is the dielectric constant and N_0 is the noninteracting density of states. This is a "soft" gap because $N(\varepsilon)$ vanishes only at ε_F. These and related Coulomb effects have been observed in disordered alloys [7–10] granular metals [11], and doped semiconductors [12].

While these two theories describe the effects of Coulomb correlations in the limiting cases of insulator and metal, they do not resolve the question of how Coulomb effects evolve as the carrier density increases and the system crosses from localized insulator to disordered metal. The ES result can be extended towards the metallic state by inserting the scaling-form divergence of the dielectric function [13] as the dopant density n approaches the critical density n_c: $\kappa = \kappa_0(1 - n/n_c)^{-\eta}$, where $\eta \approx 1.1$ for Si:B. This extrapolation makes Δ_c go to zero at n_c so that $N(\varepsilon)$ is finite through ε_F on the metal side. However, because exchange interactions are neglected, it fails to predict the cusp that appears on the metal side. A similar incompatibility arises if one extrapolates the AA result by letting Δ scale to 0 as $n \to n_c$ from above. A gap in $N(\varepsilon)$ results, but the quadratic energy dependence arising from the Hartree term is not explicitly realized. The inability of these models to join through n_c reflects the incomplete understanding of Coulomb interactions across the MIT.

Experimentally, doped semiconductors have long served as an excellent resource to study both localization and Coulomb effects. The quality of available materials is very high, and the doping level, which sets the carrier density and the strength of the interactions, can be varied systematically across the MIT. Electron tunneling provides a powerful experimental tool which can be used on doped semiconductors to supply empirical guidance in understanding interaction effects. In the standard tunneling model [14], when one electrode is an ordinary metal, the tunneling conductance $G(V,T) = \partial I/\partial V$ is a convolution of $N(\varepsilon,T)$ in the interacting electron material at temperature T with the thermal broadening function:

$$\frac{G(V,T)}{G_0} = \int \frac{N(\varepsilon,T)}{N_0}\left[-\frac{\partial f(\varepsilon - eV,T)}{\partial(eV)}\right]d\varepsilon, \quad (1)$$

where f is the Fermi function and G_0 is taken to be the conductance at a high enough voltage where interaction effects are negligible. McMillan and Mochel [7] were the

first to use tunneling to study Coulomb correlations near the MIT, using amorphous AuGe films. They succeeded in observing the square-root cusp in $N(\varepsilon)$ in metallic samples, and also noted a gap opening in insulating samples. However, quantitative comparisons with the ES model were not made due to complications associated with tunneling in insulating thin films (discussed below). Later more detailed tunneling studies were done by Hertel *et al.* [8] on NbSi films and White *et al.* [11] on granular Al films and wires. In these cases a systematic effort was made to examine the tunneling conductance as a function of increasing sample resistivity, approaching the MIT from the metal side. While much was learned about correlation effects in disordered metals, the insulating side remained neglected. This is because tunneling data on insulating films are complicated by a significant, possibly nonlinear, voltage drop across the film itself, which is difficult to distinguish from the voltage across the junction. This problem is particularly acute in low-dimensional structures because of inherently unfavorable geometric constraints on the current paths.

In this Letter, we report on tunneling measurements of Coulomb correlation effects across the MIT in a three-dimensional doped semiconductor Si:B. The use of bulk, flat crystals of Si:B allowed us to exploit a favorable current path geometry and reliably measure tunneling conductances well into the insulating side at relatively low temperature. The samples used were Si:B chips, $1/4'' \times 1/4'' \times 0.010''$, cut from a set of waters with varying boron concentrations. Ohmic contacts were made to the back faces of the Si:B chips through four annealed aluminum stripes. The resistance ratio $R(4.2\ \mathrm{K})/R(300\ \mathrm{K})$ was then measured, and n/n_c was determined for each chip using the calibration of Dai *et al.* [15]. Data for samples with n/n_c between 85% and 110% are shown in Fig. 1, where a ratio of 3 marks the MIT, $n/n_c = 100\%$. On the front face of each chip, four Si:B-SiO_2-metal tunnel junctions were fabricated using thermal and chemical oxide growth and lithographic patterning. Explicit details of the junction fabrication were described previously [12]. Both Al and Pb were used as the metal electrodes with equivalent results, with the superconductivity in the Pb suppressed by a modest (2 kG) magnetic field. About one junction in five was functional, with the remainder either having too high a resistance (indicating the SiO_2 was too thick) or showing highly asymmetric Schottky diode characteristics (indicating the SiO_2 was too thin or impure).

To take tunneling conductance spectra, the junctions were suspended stress-free and immersed in pumped liquid ^{4}He at 1.15 K, so that self-heating and temperature drift were negligible. Tunneling conductance measurements were done on a minimum of five good junctions on at least two separate chips at each doping level. We used both current and voltage bias and both digital and standard analog ac methods to measure $\partial I/\partial V$. While the current through the Si:B itself contributed a series voltage drop to the measurement, the use of backside contacts gave the current a large cross-sectional area and a short path length through the Si:B to reach the junction, minimizing this effect. Based on the known geometry and resistivity, we calculate that the resistance contribution from the Si:B itself should be $\sim 5\ \Omega$ at 1.1 K in the most insulating sample. We confirmed by measuring two-point resistances across all pairs of Ohmic backside contacts to the samples that the Si:B resistance was $<20\ \Omega$ worst case. All junction resistances were 5 to 50 kΩ in the bias ranges of interest, so at least 99% of the measured voltage drop occurred across the junction, as desired. Nevertheless, the rapid divergence of $R(T)$ and dR/dT in insulating samples as $T \to 0$ prevented reliable measurements of their tunneling characteristics at significantly lower temperatures. Therefore, for meaningful comparison, data presented in Fig. 2 for both insulating and metallic samples were taken under identical conditions at 1.15 K.

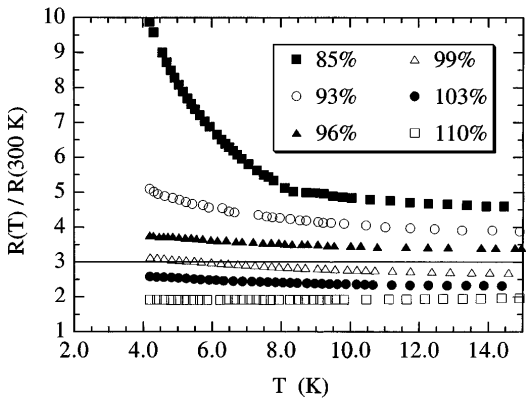

FIG. 1. Resistances of the Si:B chips as a function of temperature down to 4.2 K, normalized to their 300 K value. The values of n/n_c listed were determined by the ratio $R(4.2\ \mathrm{K})/R(300\ \mathrm{K})$ for each sample, using the calibration of Ref. [15]. The horizontal line indicates a ratio $R(4.2\ \mathrm{K})/R(300\ \mathrm{K}) = 3$, which is the MIT threshold, $n/n_c = 100\%$.

Figure 2 shows representative tunneling conductance spectra measured at 10 Hz excitation frequency. Here raw conductance data were digitized and the thermal broadening function in Eq. (1) was deconvoluted using a numerical Fourier transform algorithm [16] in order to bring out the intrinsic characteristics of each trace. It must be emphasized that an interacting $N(\varepsilon, T)$ should have an inherent temperature dependence [17] distinct from the extrinsic Fermi function broadening. Therefore the data in Fig. 2 do not represent $N(\varepsilon, 0\ \mathrm{K})$, but rather $N(\varepsilon, 1.1\ \mathrm{K})$ with ordinary thermal smearing removed. All the traces are nearly symmetric with regard to voltage polarity over bias ranges of a few mV, with the more insulating samples tending to be slightly more asymmetric. This indicates some degree of symmetry in the low energy electron-hole dispersion. The most insulating sample $n/n_c = 85\%$ shows a deep parabolic depletion near zero bias and shoulders near ± 0.5 mV. This is the signature of the ES gap and is discussed more completely in Ref. [12]. The most metallic sample $n/n_c = 110\%$ shows a square-root shape [18] extending out to many mV. This behavior is

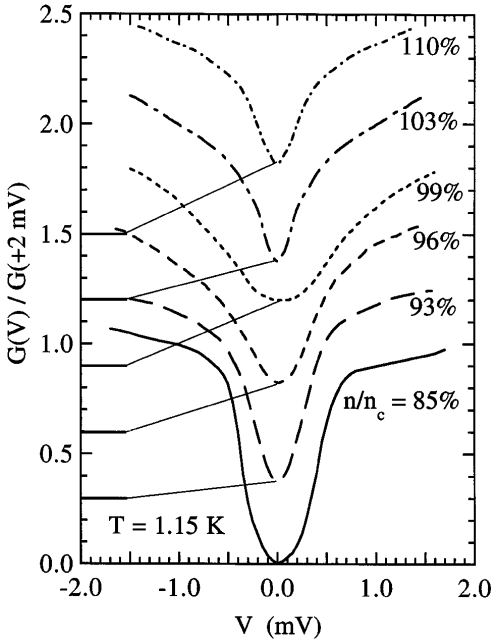

FIG. 2. Representative tunneling conductance traces, normalized to the conductance at +2 mV, from junctions at 1.15 K using the Si:B samples of Fig. 1. The thermal smearing function has been deconvoluted as described in the text. The zero of conductance for each trace has been offset for clarity and is indicated on the left axis.

shown more definitively in Fig. 3, where $G(V)$ is plotted linearly against $V^{1/2}$, with a fit by the AA form for $N(\varepsilon)$ yielding $\Delta \approx 40$ μeV. The cusp shape is reminiscent of what Wolf *et al.* [19] reported 25 years ago in Schottky junctions formed on metallic Si:B.

As n_c is approached from below (i.e., $n/n_c = 85\%$, 93%, and 96%), the curvature of $G(V)$ remains positive at and near $V = 0$, demonstrating a parabolic gap. As $n \to n_c$, the gap closes by narrowing, as expected from the ES form, and the zero-bias conductance rises so that the gap also fills in from the bottom. The gap filling is an expected consequence of the temperature dependence of $N(\varepsilon, T)$ and should occur at any nonzero measurement temperature [17]. Also evident is the disappearance of distinct shoulders as $n \to n_c$ from below. In fact, the higher energy (i.e., $|V| > 1$ mV) parts of the spectra on the 96% and 99% samples can be fit reasonably well by a $|V|^{1/2}$ dependence, which is a signature of metallic state correlations. Such a fit clearly does not work for the 85% and 93% samples. The Si:B resistance diverges on these same 96% and 99% samples at lower temperature, so these samples are still technically insulators despite the higher bias metallic correlation features seen in their tunneling spectra.

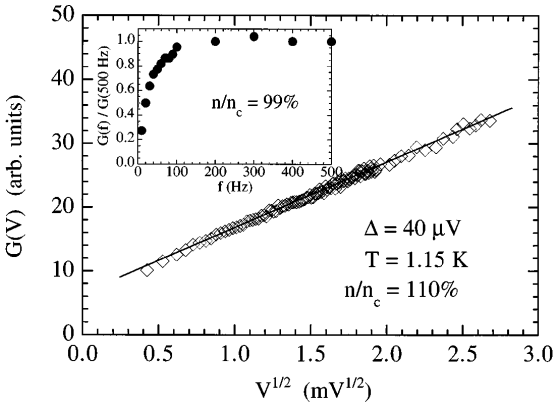

Fig. 3. A plot of the positive bias tunneling conductance from the 110% sample as a function of $V^{1/2}$. The line is a fit by the Altshuler-Aronov and McMillan density of states, yielding a correlation energy parameter $\Delta = 40$ μeV. Inset: Frequency dependence of the zero-bias conductance for a junction made on a 99% Si:B chip.

These metallic correlations away from zero bias may be qualitatively explained by considering the positions of the Fermi level and the mobility edge ε_c. The MIT occurs when n is sufficiently large that ε_F crosses ε_c, so that just below n_c, ε_F should lie very close to ε_c. Given a symmetric electron-hole dispersion near ε_F, a tunnel junction at bias voltage $\pm V$ injects electrons (or holes) into the Si:B at energy $+eV$ (electrons) or $-eV$ (holes) relative to ε_F. Thus tunneling charges injected with $|eV| > 0$ can go into or come from barely metallic states in the Si:B when n is close to n_c. While thermal smearing of the Fermi level was deconvoluted from Fig. 2, thermal and possible intrinsic broadening of ε_c will lead to a smooth transition from insulator to metal correlations in the conductance spectra on barely insulating samples as V is ramped away from zero. In this sense tunneling measurements tend to underestimate n_c compared to resistivity measurements.

Returning to the behavior near $V = 0$ as $n \to n_c$, it is apparent that the low-energy behavior of the 96% and especially the 99% samples appear qualitatively out of place. Both are significantly more smeared and have larger zero-bias conductances than the spectra from samples with either higher or lower n/n_c. We emphasize that this trend is reproducible and is not a thermal smearing effect since $-\partial f/\partial(eV)$ was deconvoluted from Fig. 2. We observed quantitatively similar traces on seven 99% junctions spanning four separate chips and on five 96% junctions on three separate chips. No spectra with sharper characteristics than those shown were observed on junctions with these two concentrations.

This additional smearing and large zero-bias conductance on the insulating samples closest to the MIT present a challenge to explain. The most helpful additional clue is the frequency dependence depicted in the inset of Fig. 3, which shows the zero-bias conductance of a 99% junction, measured dynamically using a conventional ac lock-in technique, as a function of excitation frequency. The conductance increases rapidly with frequency between 10 Hz and 100 Hz, becoming nearly frequency independent above 200 Hz. Junctions on samples of either higher or lower n/n_c had much less pronounced frequency dependence. While the frequency response of a tunnel junction is normally that of a parallel *RC* circuit, the higher frequency conductance saturation in Fig. 3 (inset) can only be fitted by a series *RC* circuit, indicating

that the observed behavior is not a property of the capacitance from the SiO_2 barrier alone. Fitting to a series *RC* circuit model yields a capacitance nearly 2 orders of magnitude larger than the 20 nF appropriate to the geometry of the device and the dielectric constants of SiO_2 and pure Si. This fact strongly suggests that the divergence of κ as $n \rightarrow n_c^-$ plays an important role very close to the MIT.

This behavior allows us to offer one plausible interpretation of the tunneling conductance in the 99% sample that can account for many of the observed features. Below but very close to the MIT, the increase in κ reflects the ease with which long wavelength charge fluctuations can be induced by a tunneling electron. These fluctuations can dynamically screen the Coulomb interaction. A large κ just below n_c, possibly combined with the proximity of a highly conducting metal electrode, could enhance the long wavelength screening enough to suppress the ES gap, with much less effect on the exchange correlations that give rise to the AA form. As a consequence the tunneling charge will feel exchange interactions but may not see the ES gap, at least until the fluctuations have damped, accounting for the general shape of the tunneling spectra in the 99% sample. The ES gap in this sample may not be observable by tunneling because the derivation of Eq. (1) assumes that the tunneling charge relaxes instantaneously into the single-particle density of states. Any long-lived induced charge fluctuations violate this condition, so the $G(V)$ would not be related to $N(\varepsilon)$ in the simple manner of Eq. (1). In the more insulating samples, the ES gap becomes apparent because the smaller κ no longer screens significantly, while the metal electrode itself is not expected to provide sufficient screening to remove the ES gap [12,20].

Of course, other physical interpretations of this data are possible. For example, it is possible that inhomogeneities in the Si:B broaden $G(V)$. Resistance measurements define n_c as the concentration where a metallic path percolates across the macroscopic system. Barely below n_c, $G(V)$ can include contributions from regions of the Si:B that are already metallic but which have not yet percolated, averaged in with insulating gapped regions. This too would tend to smear the tunneling conductance near n_c and yield aspects of metallic correlations in nominally insulating samples, though it is not as obvious how a strong low frequency dependence might develop.

In summary, we used electron tunneling to provide an empirical description of the evolution of Coulomb correlation effects across the MIT, where there is presently no adequate theoretical model. Clearly into either the insulating or metallic regimes, the tunneling conductance shows direct evidence of Coulomb correlation effects expected by existing theory. Approaching the MIT from the insulating side, metallic correlations appear in the higher energy parts of the tunneling conductance before the samples become truly metallic. This is likely due to injected charges crossing the mobility edge when the Fermi level is sufficiently close. Very close to the critical density, the conductance close to zero bias becomes unusually smeared and shows a significant frequency dependence at low frequency. This implicates the excess long wavelength screening caused by the dielectric constant divergence as important to the degree of Coulomb interactions.

We thank H. Baranger, L. Levitov, and R. Bhatt for useful comments. This work was supported by NSF Grant No. DMR-9316803.

[1] A standard review of the large amount of work in this area is given by P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[2] L. S. Levitov, A. V. Shytov, and A. Y. Yakovets, Phys. Rev. Lett. **75**, 370 (1995); see also D. Belitz and T. R. Kirkpatrick, Rev. Mod. Phys. **66**, 261 (1994).
[3] M. Pollak, Discuss. Faraday Soc. **50**, 13 (1970); G. Srinivasan, Phys. Rev. B **4**, 2581 (1971); V. Ambegaokar, B. I. Halperin, and J. S. Lander, Phys. Rev. B **4**, 2612 (1971).
[4] B. L. Altshuler and A. G. Aronov, Solid State Commun. **30**, 115 (1979).
[5] W. L. McMillan, Phys. Rev. B **24**, 2739 (1981).
[6] A. L. Efros and B. I. Shklovskii, J. Phys. C **8**, L49 (1975); B. I. Shklovskii and A. L. Efros, Fiz. Tekh. Poluprovodn **14**, 825 (1980 [Sov. Phys. Semicond. **14**, 487 (1980)].
[7] W. L. McMillan and J. Mochel, Phys. Rev. Lett. **46**, 556 (1981).
[8] G. Hertel *et al.*, Phys. Rev. Lett. **50**, 743 (1983).
[9] J. H. Davies and J. R. Franz, Phys. Rev. Lett. **57**, 475 (1986).
[10] Y. Imry and Z. Ovadyahu, Phys. Rev. Lett. **49**, 841 (1982).
[11] A. E. White, R. C. Dynes, and J. P. Garno, Phys. Rev. B **31**, 1174 (1985); Phys. Rev. Lett. **56**, 532 (1986).
[12] J. G. Massey and M. Lee, Phys. Rev. Lett. **75**, 4266 (1995).
[13] M. Capizzi *et al.*, Phys. Rev. Lett. **44**, 1019 (1980); H. F. Hess *et al.*, Phys. Rev. B **25**, 5578 (1982).
[14] See, for example, E. L. Wolf, *Principles of Electron Tunneling Spectroscopy* (Oxford Univ. Press, Oxford, 1985).
[15] P. Dai, Y. Zhang, and M. Sarachik, Phys. Rev. Lett. **66**, 1914 (1991).
[16] W. H. Press, S. A. Teukolsky, W. T. Vertterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, 1994), pp. 537–544.
[17] E. I. Levin *et al.*, Zh. Eksp. Teor. Fiz **92**, 1499 (1987) Sov. Phys. JETP **65**, 842 (1987).
[18] The conductance through $V = 0$ in the metallic samples is not truly sharp after deconvolution because a finite spatial frequency cutoff must be introduced in the deconvolution procedure, resulting in a slight rounding of any nonanalytic features.
[19] E. L. Wolf *et al.*, Phys. Rev. Lett. **26**, 438 (1971).
[20] E. Cuevas *et al.*, Philos. Mag. B **20**, 1231 (1994).