

More on the S-I transition: normal-metallic phase at $T = 0$

As noted at the end of the last lecture, a semiquantitative understanding of much of the data on the superconductor-insulator transition in thin metallic films can be obtained by regarding it as a zero-temperature QPT. However, while the data in nonzero magnetic field certainly seem to be well explained by taking the field as the tuning parameter (and the same applies to electric fields), what is the tuning parameter when the external field is zero? Experimentally it certainly seems to be the film thickness, but it is difficult to see why this should as such have any fundamental significance, and the natural assumption would seem to be that it is the sheet resistance R_{\square} which is crucial. It is then striking that at least some of the data indicate that the transition occurs very close to the value $R_{\square} = R_Q \equiv h/4e^2 \approx 6.45 \text{ k}\Omega$. Although it is nowadays not generally believed to be the whole truth, there is a simple model¹ which accounts for this feature; it is essentially a generalization of the “JJ array” model explained in lecture 13 to take into account the effects of dissipation.

Let’s start by considering briefly a celebrated model of dissipation in QM, the so-called spin-boson model.² Formally, this describes a spin-1/2 particle coupled (in a particular way) to a bath of simple harmonic oscillators; in the limit of interest to us the Hamiltonian is

$$\hat{H}_{\text{SB}} = -\frac{1}{2}\hbar\Delta\hat{\sigma}_x + \frac{1}{2}q_0\hat{\sigma}_z \sum_{\alpha} C_{\alpha}x_{\alpha} + \hat{H}_{\text{SHO}} \quad (1)$$

where the $\hat{\sigma}_i$ are Pauli matrices and \hat{H}_{SHO} is the Hamiltonian of a set of simple harmonic oscillators, with densely spaced frequencies up to some cut off $\omega_c \gg \Delta$, whose coordinates are $\{x_{\alpha}\}$. In the present context it is helpful to think of (1) as a description of a particle tunneling between two spatially distinct states (wells) labeled by the eigenstates of $\hat{\sigma}_z$ and separated by q_0 , subject to “measurement” of $\hat{\sigma}_z$ by an “environment” modeled by the oscillator bath; then, in the absence of the system-bath coupling described by the second term in (1), the rate of tunneling between the wells would be Δ . All the effects of the “environment” turn out to be encapsulated in the coupling spectral density $J(\omega)$ defined by

$$J(\omega) \equiv \frac{\pi}{2} \sum_{\alpha} \left(\frac{C_{\alpha}^2}{m\omega_{\alpha}} \right) \delta(\omega - \omega_{\alpha}) \quad (2)$$

The case of interest in the present context is when $J(\omega)$ has the so-called Ohmic form

$$J(\omega) = \eta\omega f(\omega/\omega_c) \quad (3)$$

where $f(x)$ is some smooth function (e.g. $\exp -x$) which tends to 1 as $x \rightarrow 0$ and to 0 as $x \rightarrow \infty$; thus for $\omega \lesssim \Delta$ we have simply $J(\omega) = \eta\omega$. By working out the consequences of the coupling for the *classical* motion of the original extended system (see AJL et al.,

¹Chakravary et al., PRL **56**, 2303 (1986); M.P.A. Fisher, *ibid.* **57**, 885 (1986).

²see e.g. AJL et al., RMP **59**, 1 (1987).

ref. cit., section 2) we identify η as the friction coefficient which appears in the classical motion. For subsequent convenience we define the dimensionless parameter

$$\alpha \equiv \eta q_0^2 / 2\pi\hbar \quad (4)$$

The crucial result that we shall need concerning the spin-boson problem, which can be obtained by any one of a variety of arguments, is that for $\Delta \ll \omega_c$ (a condition which we shall always implicitly assume) the point $\alpha = 1$ corresponds to a qualitative change in the thermodynamic (and also the dynamical) behavior: For $\alpha < 1$ the groundstate of the system is still, as for the original uncoupled system, a coherent quantum superposition of the states in the left and right wells, corresponding to $\sigma_z \pm 1$ respectively, although the tunneling splitting which separates this groundstate from the odd-parity excited state may be much reduced (by a Frank-Condon factor) from its original value $\hbar\Delta$. For $\alpha > 1$, on the other hand, coherence is entirely lost and the groundstate is twofold degenerate, corresponding to localization in one well or the other. It is important to observe that provided only that $\Delta \ll \omega_c$, the value of α on the separatrix, namely 1, is *independent of* Δ . This radical change in behavior at $\alpha = 1$ is sometime viewed as a “zero-dimensional phase transition”.

Now let us take the original continuous variable q which distinguishes between the two wells to be the phase drop $\Delta\phi$ across a single Josephson junction. We model the latter by the so-called resistively shunted junction (RSJ) model, in which the junction itself is shunted by both a capacitance C and a parallel resistance (typically of unknown origin) R_n . Since the dissipation, which in the case of mechanical particle would be $\eta\dot{q}^2$, is in this case $(\hbar/2e)^2\Delta\dot{\phi}^2/R_n$, we identify η as $(\hbar/2e)^2R_n^{-1}$. Consider now the case where the two wells are separated in phase by 2π , ignoring for the moment any questions about whether those two states are really different. From eqn. (4), the value of α for this case is

$$\alpha = \frac{\hbar}{4e^2} \frac{1}{R_n} \equiv R_Q/R_n \quad (5)$$

Hence, if we take this application of the spin-boson model seriously, the phase drop $\Delta\phi$ across the junction in the groundstate is fixed at the value 0 (or 2π) for $R_n < R_Q$, while for $R_n > R_Q$ it is a superposition of two states, suggesting that it prefers to make the conjugate variable N rather than $\Delta\phi$ definite. Thus we might guess that $R_n < R_Q$ would correspond to the superconducting state and $R_n > R_Q$ to an insulating state.

The argument, while suggestive, is of course far too naive: it is not even clear that we can make sense of a “superposition” of the states $\Delta\phi = 0$ and $\Delta\phi = 2\pi$. However, remarkably, a more sophisticated calculation (see the cited references) for a system of grains connected by Josephson junctions reproduces the result that exactly at $\alpha = 1$ (independently of both the Josephson coupling E_J and the capacitance energy E_C of the grains) the system undergoes a qualitative transition in its behavior: for $\alpha < 1$ (large R_n) the relative phase of neighboring grains is able to “diffuse” over many multiples of 2π , and there is no LRO; the system is thus predicted to be insulating at $T = 0$. For $\alpha > 1$ (small R_n) on the other hand, the dissipation ($\propto R_n^{-1}$) pins the relative

phase on neighboring grains and hence the system possesses LRO and behaves as a superconductor. Thus, counter-intuitively, the presence of dissipation (in the motion of the relative phase) is actually *helpful* to superconductivity!

This is still by no means the whole story, and over the last 20 years there have been a number of calculations along lines similar to the above, in some cases modeling the system not by an array of Josephson junctions but by a “dirty boson” picture. All those calculations have in common that they attribute the S-I transition to quantum phase fluctuations of the Cooper pairs and regard it as a QPT, and that they predict that the value of R_{\square} at the transition should be of order R_Q (though they differ in the degree of universality, if any, predicted for the ratio R_n/R_Q).

Thus, we have two apparently quite different theories of the S-I transition: a “fermionic” mechanism, in which the important effect of disorder is primarily to modify the *single-electron* spectrum (so that no Cooper pairs are present in the insulating phase), and a “bosonic” scenario in which Cooper pairs are formed just as in the pure metal, but may or may not be localized due to the effects of quantum fluctuations of their phase. Thus the obvious question is, what would be the experimental signatures of each of these two mechanisms?

A very spectacular prediction of the bosonic scenario is the scaling relations characteristic of the behavior near a QCP; that such scaling behavior is observed, as a function of both thickness and magnetic field, would seem to be a strong argument in favor of this scenario. On the other hand, the universal value of R_n predicted by most (though not all) versions of the dirty-boson theory is not always seen experimentally. As to the fermionic scenario, a major prediction is that the energy gap, as measured in tunneling experiments, should scale with T_c and vanish in the insulating state; by contrast, the bosonic theory predicts that since the gap is a function only of the amplitude of the order parameter and is insensitive to its phase, it should be finite even in the insulating phase (unless the interaction of the single-electron spectrum with order parameter phase fluctuations somehow depresses it to zero). In this respect experiment seems to favor the fermionic picture: the tunneling gap indeed scales with T_c and vanishes in the insulating phase. Thus it appears that both mechanisms may play a role.

Is it possible to test directly for the presence of Cooper pairs in the insulating phase? One indirect piece of evidence is that the magnetoresistance is direction-dependent, and thus presumably has an orbital component, with a sign that would be consistent with vortex-flow mechanism. More recently, the group of Valles has reported a striking result: They examined the magnetoresistance of a film of amorphous Bi with a “nanohoneycomb” network of holes drilled in it. The experimental (activated) resistivity, when plotted as a function of field at constant temperature (and thickness) showed dramatic AB-type oscillations, with a period corresponding to one superconducting flux quantum ($h/2e$) through the hole. Unless one interprets this behavior as a variant of the Sharvin effect (see lecture 6), it seems clear evidence that Cooper pairs exist in the insulating phase and moreover that their phase is coherent over a length at least of the order of the circumference of the holes (~ 100 nm). This is puzzling, to say the least...

Let's now turn to a different though somewhat related topic, namely the existence or not of a normal-metallic state at $T = 0$. We shall thus assume that the system of interest does not become superconducting. We recall from lecture 7 the following conclusions:

- (1) If the phase-breaking time is measured, e.g. by measuring the crossover field in the magnetoresistance as a function of T , it should tend to ∞ as $T \rightarrow 0$ (as some negative power of T) except for the effect of static magnetic impurities³ [In particular, el-el interactions should give $\tau_\phi^{-1} \sim T$ in 2D and $T^{3/2}$ in 1D: see LR section III.2.]
- (2) If we simply add the contributions to the correction $\Delta\sigma(T)$ to the Boltzmann conductivity in 2D from WL and from interactions (as seems reasonable when both are small) we find that apart from an uninteresting (and unmeasurable!) constant the total correction is given by the expression

$$\Delta\sigma(T) = -(e^2/\pi h)(\alpha p + 1 - \frac{3}{4}\tilde{F}_\sigma) \ln(1/T\tau_\phi) \quad (6)$$

where p is the power of T with which τ_ϕ^{-1} tends to zero as $T \rightarrow 0$, and α is a constant of order 1. Although values of the "screening function" \tilde{F}_σ up to 3.5 have been measured in Si MOSFET'S, to make the factor in (6) negative would require $\alpha p < 1.6$, which seems unlikely though not impossible. Unless this happens, or the situation changes qualitatively when $\Delta\sigma(T)$ becomes comparable to σ_0 , one would expect that a *normal-metallic state cannot exist in 2D as $T \rightarrow 0$* . This was indeed the accepted wisdom until around 1997.

In the last dozen years two sets experiments have been carried out which severely challenge both the above general conclusions. The first set, associated primarily with names of P. Mohanty and R.A. Webb, claim to refute the assertion that $\tau_\phi(T)$ always $\rightarrow \infty$ as $T \rightarrow 0$. The second set, originally conducted by Kravchenko, Sarachik and co-workers but subsequently confirmed by several other groups, tend to suggest that a normal-metal state can indeed exist at zero temperature in a 2D system even in the absence of a magnetic field and of magnetic impurities. Although there is obviously a possible connection between the two sets of experiments, it makes sense to discuss them independently, and I start with the Mohanty-Webb experiments.

In their original paper,⁴ Mohanty et al. start from the observation that despite the confident theoretical prediction that $\tau_\phi(T)$ should tend to ∞ as $T \rightarrow 0$, all existing experiments on magnetoresistance, whether on 1D systems (wires) or 2D systems (films), if interpreted according to the standard prescription, show a flattening-off of $\tau_\phi(T)$ at

³It is essential, here, to remember that the "anomalous" effects of spin-orbit scattering discussed in lecture 6 require that the scattering be effectively 3D. In Si MOSFET's or GaAs heterostructures, it is effectively 2D and the spin-orbit scattering simply adds to the dephasing terms.

⁴PRL **78**, 3366 (1997).

temperatures ranging from as high as 10 K (for some 1D GaAs devices) to ~ 20 mK (for some 2D Au films). In view of the uncertainties concerning noise and magnetic impurities in those experiments, they then present results of their own measurements of $\tau_\phi(T)$ (as again inferred in the standard way from magnetoresistance curves) in six quasi-1D Au wires (width 25-250 nm, length 50-200). In each case the inferred $\tau_\phi(T)$ flattens off at a temperature ~ 1 K, to a “zero- T ” value which ranges from ~ 0.3 –3 nsec. They find that at low T all their data show an excellent fit to the formulae (which let us for the moment treat as semi-empirical)

$$\tau_\phi(T) = \tau_0 \tanh(\alpha\pi^2 \sqrt{\hbar/\tau_0 k_B T}) \quad (7)$$

where τ_0 is the $T \rightarrow 0$ saturation value which is read off directly from experiment; thus the only fitting parameter is the dimensionless constant α , which they find to vary from 0.6 to 1.1 depending on the sample.

Two obvious possible explanations of the data (or at least of the qualitative features) are

- (a) that because of external noise the true temperature is not that measured by their thermometer but has saturated, and
- (b) that the flattening is real but is an effect of magnetic impurities (which as we saw in lecture 6 are expected to give a temperature-independent contribution to τ_ϕ).

In an attempt to exclude explanation (a), the authors measured the resistance as a function of T in a field large enough that $L_M \ll L_\phi(T)$, so that the WL effect should have saturated. They find the behavior theoretically expected for interaction effects in 1D, namely $\Delta\rho \sim T^{-1/2}$ (with the constant also as predicted) and infer that the electrons are still in good contact with the thermal bath⁵ (i.e. the “true” T is that measured by the thermometer). To try to exclude explanation (b), they compared the $\tau_\phi(T)$ of their samples before and after ion implantation of up to 2.8 ppm of Fe impurities (which are known to form localized magnetic moments). As expected, the Fe ions decrease τ_ϕ by about an order of magnitude; however, they do *not* appear to lead to saturation, but rather to a $\ln T$ term in the resistivity; this is thought to be due to the Kondo effect, which is known to occur in the AuFe system.

Mohanty et al. go on to argue that the flattening of τ_ϕ is a genuinely intrinsic effect, and is due to “zero-point dephasing” by electron-electron interactions. On the basis of their hypothesis they calculate a value for the τ_0 which appears in their semi-empirical fit formula, namely (1D system only!)

$$\tau_0 = \left(\frac{4\pi\hbar^2 L}{e^2 d^2 R m^2 D^{3/2}} \right)^2 \quad (8)$$

where L is the length of the wire and $d = 2$ or 3 its effective dimensionality with respect to the elastic mean-free path l , D is the diffusion coefficient and R the resistance. In subsequent work⁶ they elaborate this picture further and attempt to relate it to other

⁵This conclusion might of course fail if the thermal contact itself were to depend strongly on magnetic field, but there is no obvious reason to suspect this.

⁶e.g. PRB **55**, 13452 (1997), Physica B **280**, 646 (2000), PRL **88**, 146601 (2002).

experimental anomalies such as the diamagnetism of mesoscopic rings.

Ever since its original publication, this work has been extremely controversial. Basically, there seems to be three possible interpretations:

- (1) The apparent saturation is an experimental artifact, due e.g. to noise in the current leads.
- (2) The saturation is real but is due to effects of an already known type, e.g. two-level tunneling systems.
- (3) The saturation is real and shows need for radical revision of standard ideas concerning decoherence etc.

My own view (based on personal knowledge of the authors!) is that (1) is extremely unlikely. The problem with position (2) is that it appears to be essentially unfalsifiable, since it is very difficult to exclude the presence of a concentration of some unspecified “two-level systems” which is sufficient to give the observed saturation but essentially unobservable in other types of experiments. Position (3) has been strongly advocated not only by the original experimental team but also by Zaikin and Golubev,⁷ who claim to have done an improved diagrammatic calculation which shows that “zero-point dephasing” by el-el interactions is a real effect; it has been equally strongly contested by a number of theorists who believe that the original picture developed in the early 80’s remains completely valid, and who claim to have faulted the Zaikin-Golubev calculation. At the time of writing the jury is still out. . .

Now let’s turn to the experiments⁸ which suggest the existence of a normal metal in 2D at $T = 0$. These experiments are conducted on Si MOSFET’s, AlGaAs heterostructures and similar systems, with mobilities of the order of $\sim 4 \times 10^4 \text{ cm}^2/\text{V sec}$. In previous experiments in such systems, e.g. on the quantum Hall effect, the areal density n is typically $\sim 10^{11}$ - 10^{12} ; in these experiments, it was lower, typically in the range $\sim 10^{10}$ - 10^{11} . It should be noted that the value of the dimensionless parameter $r_s \equiv \frac{m^*}{m\epsilon} \frac{2}{a_0\sqrt{\pi n_s}}$ (a measure of the ration of the typical Coulomb interaction to the Fermi energy) is quite large in these systems, ranging from ~ 4 to ~ 40 . In fact, the Fermi energy is typically $\sim 0.6 \text{ meV}$ while the Coulomb energy is $\sim 10 \text{ meV}$.

As usual in such systems, the only quantity which is easy to measure is the resistivity (or sheet conductance), and the only things which can be easily varied apart from the areal density are the strength and direction of the electric and magnetic fields. With regard to the magnetic field, it turns out that the clearest pattern of data emerges where it is parallel to the plane (so that there are presumably no orbital effects): where it is (partially or wholly) perpendicular to the plane there are complications connected with the quantum Hall effect (which is in fact seen at sufficiently high values of the field). I will therefore concentrate, here, on the situation in parallel fields.

⁷PRL **81**, 1074 (1998), PRB **59**, 9195 (1999).

⁸The best review is probably Abrahams et al., RMP **73**, 251 (2001).

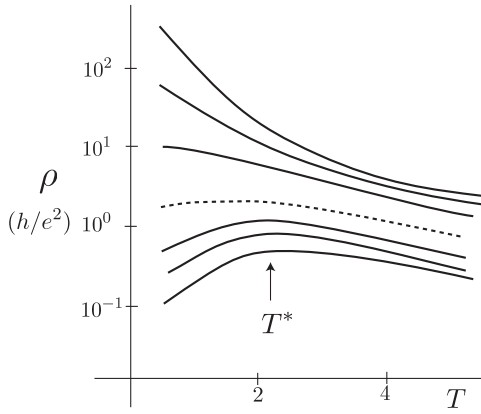


Fig. 1

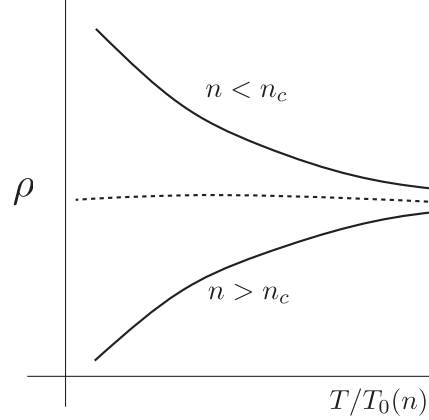


Fig. 2

First let's consider the behavior in zero field, and consider for definiteness the original Si MOSFET system. The linear resistivity behaves as follows: For densities below a critical value n_c (approximately $9 \times 10^{10} \text{ cm}^{-2}$) $\rho(T)$ increases slowly with decreasing temperature and appear to be tending to ∞ for $T \rightarrow 0$ (the lowest temperature obtained in the experiments to date are probably around 4 mK). However, at densities above n_c the behavior is quite different: $\rho(T)$ rises slowly with decreasing T down to a temperature $T^* \sim 2 \text{ K}$, then starts to *decrease*. Thus there exists a “separatrix” along which $\rho(T)$ is essentially flat in the limit $T \rightarrow 0$; this corresponds, as stated, to a critical density $n_c \approx 9 \times 10^{10} \text{ cm}^{-2}$, and the $T \rightarrow 0$ limit of the sheet resistance ($\equiv \rho_c$) is approximately $3h/e^2$. Qualitatively similar results are found for GaAs-AlGaAs heterostructures, with n_c in this case being $\sim 1.5 \times 10^{10} \text{ cm}^{-2}$ and a $\rho_c \sim 1.5 (h/e^2)$. (Note that the approximate factor of 2 difference does *not* reflect the presence of 2 valleys in the Si MOSFET's vs. 1 in GaAs: it is the “wrong way round”!). The picture is qualitatively quite similar to what is found in e.g. granular Al films, except that in that case the resistivity of the curves with $\rho < \rho_c$ drops very sharply below the “crossover” temperature, indicating the onset of superconductivity. In the present case there is no indication that $\rho(T) \rightarrow 0$ as $T \rightarrow 0$, for any value of n .

A surprising and spectacular feature of the data on Si MOSFET's is their scalability: for any given value of n , and $T \lesssim 2 \text{ K}$, one can find a characteristic temperature $T_0(n)$ such that we can write

$$\rho(T, n) = \rho_c f(T/T_0(n)) \quad (9)$$

where f has two branches, one “insulating” ($n < n_c$) and one “metallic” ($n > n_c$). A striking factor is that the function $T_0(n)$ is independent of the sign of $n - n_c$: in fact, we find that if $\delta \equiv (n - n_c)/n$, then

$$T_0(\delta) = |\delta|^b \quad (10)$$

where the exponent b is 1.60 ± 0.1 . Another striking and suggestive feature is that the scaled resistivity $\rho^*(T) \equiv \rho(T)/\rho_c$ satisfies the “duality” property (at not too low T)

$$\rho^*(\delta, T) = 1/\rho^*(-\delta, T) \quad (11)$$

This is the behavior which is theoretically expected near a quantum phase transition. The above scaling properties are found in Si MOSFET's and p-SiGe heterostructures, but *not* in GaAs heterostructures. The actual form of the T -dependence is controversial: on the insulating side it is close to $T^{-1/2}$ (*not* $\ln T$!) while on the metallic side it *may* be fitted by $\rho_0 + \rho_1 \exp -(T_0/T)^\gamma$, $\gamma \sim 1$.

The effect of large increasing the electric field is qualitatively similar to that of increasing the temperature, with a field of ~ 50 mV/cm playing the role of T^* . Again one finds a scaling behavior: at constant T

$$\rho/\rho_c = f_2(\delta/E^{1/a}) \quad (12)$$

where the exponent a is ~ 2.7 .

Now let's turn to the effect of a magnetic field (in the plane of the surface unless otherwise stated). Again we focus first first on Si MOSFET's at *high* T ($\gtrsim T^*$) such a field has essentially no effect. At low T , a field induces a very large increase in resistivity (positive magnetoresistance!): the effect is up to 4 orders of magnitude in a field of 1 T! The effect is most pronounced on the (originally) "metallic" side of the transition, but is seen for all densities, and even at the highest densities studied ($\sim 8.8 \times 10^{10}$ cm²) a magnetic field as small as 1.4 T can make the system insulating (see Abrahams et al., fig. 10). The magnetic field effect appears to saturate around a field of ~ 3 T; it may be significant that this is roughly the field for which the Zeeman splitting $g\mu_B B$ becomes comparable to $k_B T^*$. Qualitatively similar behavior is seen in GaAs heterostructures, but *not* in one prima facie analogous system, namely p-SiGe heterostructures; in the latter system a parallel magnetic field has virtually no effect at all! A very plausible explanation of this phenomenon is that in this system the spin-orbit coupling is very strong, and thus since the orbital angular momentum must lie perpendicular to the plane the spins are also constrained to lie in that direction, so a parallel field has no (first-order) effect.

Apart from the resistivity, the most significant quantity that has been measured on those systems is the "compressibility" of the electrons or holes, which can be measured by a capacitance technique. This has in fact been measured for p-GaAs/AlGaAs heterostructures, and is found to *change sign* at the critical value of the density.

The explanation of those phenomena is at the time of writing wide open. There are a few things one can say, however, with fair confidence. The first is that it is almost certain that the effect of the magnetic field is overwhelmingly through the Zeeman coupling, and that the degree of metallicity is, qualitatively, inversely correlated to the spin polarization of the system. The second is that the spectacular scaling behavior strongly suggests that the point $n = n_c$, $T = 0$ corresponds to a "quantum phase transition" (i.e., as explained in lecture 13, a phase transition occurring at $T = 0$ as other parameters, in this case n , are varied). And the third is that in considering the possible nature of this transition, the most reasonable approach (as in the case of the QHE, though for a reason different in detail) is to concentrate first on the effects of the Coulomb energy and that the KE is in some sense a perturbation. Beyond this there is no agreement.

One obvious possibility, in view of the (probable) dominance of the Coulomb inter-

action, is that the insulating phase is a sort of “incipient Wigner crystal” – that is, a state in which the Coulomb energy favors localization of the single-electron states. For a translation-invariant system numerical calculations indicate that the WC state is stable for values of r_c above a critical value of ~ 37 ; this is higher than the value of r_c (~ 15) at which the separatrix occurs in Si MOSFET’s, but not by a large margin. This hypothesis of course does not in itself explain why smaller values of r_s correspond to a metallic state, that is why WL does not have the expected effect.

A second possibility⁹ is that the system is an incipient superconductor, which presumably is always above the KT transition in the range of temperatures obtained in experiments so far. Assuming the pairing is of the standard “singlet” type, this would explain qualitatively why a magnetic field has such a drastic effect (because it acts as pair-breaking for the spins). The qualitative picture would then be that in the absence of superconducting-type pairing the system would behave in the traditionally expected way, i.e. localize at $T = 0$ (as indeed it seems to do in a sufficiently high field), while the possibility of pairing destroys the usual argument for localization. It is not clear how far this scenario can explain the data quantitatively. (For further possibilities, see Abrahams et al., loc. cit., section 3).

⁹See e.g. P. Phillips et al., Nature **395**, 253 (1998), Phys. Rev. B **64**, 184511 (2001); *ibid.* 052507 (2001).