Energy scale behind the metallic behaviors in low-density Si MOSFETs

Geneviève Fleury[1] and Xavier Waintal[1,2]
[1]*SPEC-IRAMIS, CEA Saclay, F-91191 Gif-sur-Yvette Cedex, France*
[2]*SPSMS-INAC, CEA, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France*
(Received 24 February 2010; published 29 April 2010)

We show that the unexpected metallic behavior (the so-called two-dimensional metal-insulator transition) observed in low-density silicon metal-oxide-semiconductor field-effect transistors is controlled by a unique characteristic energy scale, the polarization energy. On one hand, we perform quantum Monte Carlo calculations of the energy needed to polarize the two-dimensional electron gas at zero temperature, taking into account Coulomb interactions, valley degeneracy, and electronic mobility (disorder). On the other hand, we identify the characteristic energy scale controlling the physics in eight different sets of experiments. We find that our *ab initio* polarization energies (obtained without any adjustable parameters) are in perfect agreement with the observed characteristic energies for all available data, both for the magnetic field and temperature dependence of the resistivities. Our results put strong constraints on possible mechanisms responsible for the metallic behavior. In particular, there are strong indications that the system would eventually become insulating at low enough temperature.

DOI: 10.1103/PhysRevB.81.165117 PACS number(s): 71.30.+h, 02.70.Uu, 71.10.Ca

I. INTRODUCTION

Until 1994, two-dimensional (2D) systems were widely believed to be ultimately insulators[1] as the presence of even a tiny disorder is enough to make the one-body electronic wave functions localized. In fact, until then, the experimental corpus was widely consistent with this paradigm:[2,3] the electrical resistance was found to increase as one lowered the temperature, indicating an eventual divergence at zero temperature. The observation of metallic behaviors in low-density silicon metal-oxide-semiconductor field-effect transistors (Si MOSFETs) (Ref. 4) hence came as a surprise and have been puzzling the community ever since.[5,6] The experiments were first acknowledged with skepticism but they were soon repeated and extended by several groups,[7–11] including in other materials,[12–19] so that it became clear that some intrinsic physics of the two-dimensional electron gas was probed in those experiments.

On the theory side, many different scenarios were proposed. In the weak disorder regime, perturbative calculations performed in 1980s (Refs. 20–22) showed the crucial role of spin and valley degeneracy and more recently, calculations in the limit of an infinite number of valleys suggested that electronic interactions could stabilize a non-Fermi-liquid metallic phase.[23] Other theoretical works proposed that the metal-insulator transition could be due to an emulsion between a Wigner crystal and a Fermi liquid[24] or to some sort of a Wigner-Mott transition.[25] It has also been argued that the metallic behaviors could be the signature of a superconducting phase.[26–28] In contrast to these rather "extreme" scenarios other works proposed a more conservative picture. This includes temperature-dependent screening,[29,30] temperature-dependent disorder,[31,32] or temperature-dependent scattering between two different spin subbands split by a strong spin-orbit coupling.[33–35] Percolation phenomena were also claimed to play a role at the metal-insulator transition itself.[36–41] The problem was also approached using numerical techniques and a delocalization effect by interactions was observed repeatedly.[42–47]

The large spread in the above proposals, together with rather intense controversies between the conservative and the less conservative scenarios naturally lead to some confusion in the field so that no scientific consensus has emerged so far. The goal of this paper is twofold. First, we will perform a critical analysis of the existing experimental data and show that extremely strong constraints can be put on the possible theoretical scenarios. Second, following our recent proposal,[48] we will show that a simple model with a minimum number of (well-established) ingredients is enough to capture all the salient experimental facts.

In Sec. II, we summarize the main experimental facts with a focus on the characteristic energies that can be extracted from the data. In Sec. III, we introduce our minimum model to describe Si MOSFETs, which is then used in Sec. IV to calculate the polarization energy E_p in presence of both disorder and electron-electron interactions. In Sec. V, we come back to the experimental data and show that our calculated E_p agrees quantitatively (with no adjustable parameter) with the characteristic energies that can be extracted from either the magnetoresistance data or the resistance versus temperature data. Finally, we discuss the implications of the above findings for our own scenario[48] (Sec. VI) as well as others (Sec. VII).

II. BASIC EXPERIMENTAL FACTS

A. Main observations

Let us quickly review the basic experimental facts that we want to understand. The first set of measurements that actually started the interest for these high-mobility MOSFETs is the behavior of the resistivity $\rho(T)$ as a function of temperature (at temperatures lower than a few kelvin where phonon scattering no longer comes into play). The $\rho(T)$ behavior is well understood in two limiting cases. At high density, the resistivity is rather small and depends only weakly on temperature, except for a weak negative $\partial\rho/\partial T$ due to weak localization. At low density, the resistivity is much larger and one eventually reaches a clear insulating behavior where ρ depends very strongly (exponentially) on temperature ($\partial\rho/\partial T<0$). Before the report made in Ref. 4, the common

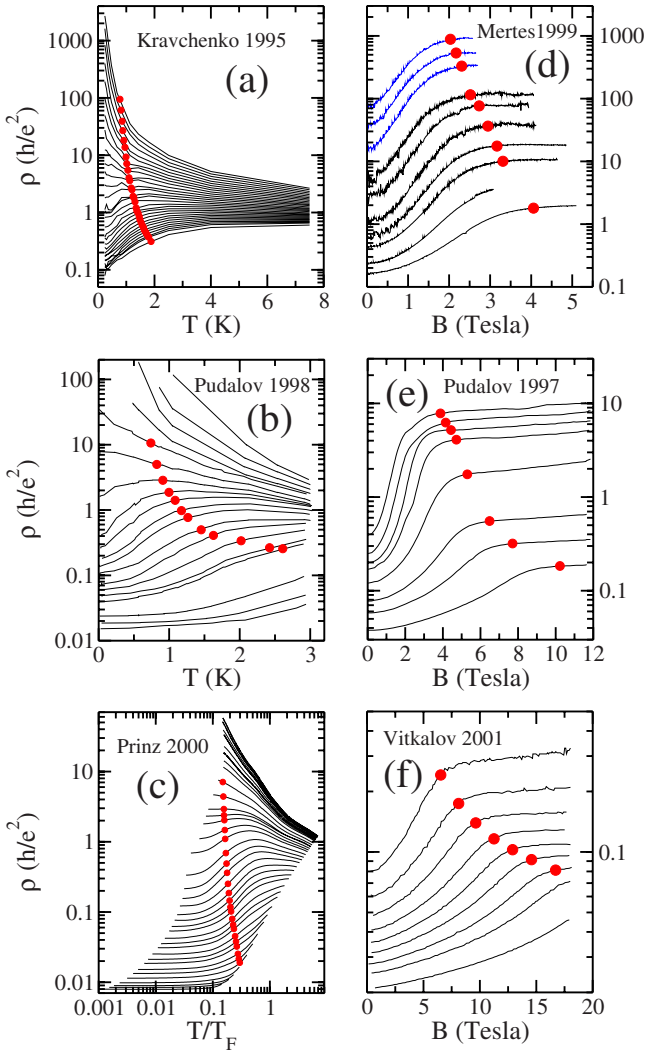

FIG. 1. (Color online) ρ as a function of T (left panels) and as a function of B (right panels), for different n_s, in various Si MOSFETs [(a) from Ref. 7, (b) from Ref. 10, (c) from Ref. 49, (d) from Ref. 8, (e) from Ref. 50, and (f) from Ref. 51]. The red circles are the corresponding polarization temperatures T_p and polarization magnetic fields B_p obtained from the numerical simulations.

behavior observed in Si MOSFETs as well as in other two-dimensional electron gases was a simple crossover between the high- and the low-density limits. Hence, the observation of an intermediate density regime with $\partial\rho/\partial T>0$ (hereafter referred as property **P1**) came as a large surprise. In the left panels of Fig. 1, we reproduced the data of three different experiments, including the data of the original Kravchenko *et al.*. One can observe a rather pronounced $\partial\rho/\partial T>0$ as the resistivity increases by a factor 10 between 100 mK and 4 K. It was naturally surmised that in this regime, the resistivity could be extrapolated to a finite value at zero temperature, so that the system was in a metallic state, in contradiction with the prediction of scaling theory of localization.[1]

The second important raw experimental feature is the dependence of the resistivity versus an *in-plane* magnetic field B: in presence of a magnetic field, the resistivity increases quickly before saturating (property **P2**). This is illustrated in the right panels of Fig. 1 for three different experiments. The increase in resistivity is comparable to the increase that one gets with increasing temperature and $\partial\rho/\partial T$ quickly becomes negative (insulating like).

In our opinion, **P1** and **P2** are the two main experimental features that require an explanation. There have been an important effort to establish scaling laws and critical exponents close to the metal-insulator transition but we will not discuss this aspect. Indeed, once the metallic behavior **P1** is understood, there is no question that, as the system must be eventually insulating at low density, some sort of transition must occur. While this transition might be of interest, it is not fundamentally puzzling in itself.

B. First analysis

One of the questions that many physicists had in mind when the first experiments[4] came up was, what's new in those devices? Why have we not seen this behavior before? In fact, there was nothing qualitatively new about those MOSFETs. The real novelty lied in the fact that the samples were of extremely high mobility (for Si MOSFETs) which allowed the experiments to be carried out at much lower densities than previously possible. Working at lower densities has three consequences.

Property P3. The ratio r_s of interaction over kinetic energy is large. This ratio $r_s=m^*e^2/(4\pi\epsilon_0\epsilon\hbar^2\sqrt{\pi n_s})$ (where n_s is the electronic density, $m^*=0.19m_e$ the effective mass, e the electron charge, ϵ_0 the dielectric constant, and $\epsilon=7.7$ the relative dielectric constant) takes values between 5 and 10 for the metallic region, hence the system is intrinsically correlated.

Property P4. The effective disorder seen by the electrons, as parametrized by $1/k_Fl$ (k_F Fermi momentum and l mean-free path) is paradoxically not smaller: indeed the high mobility of the samples is compensated by the fact that the density is lower and therefore $k_Fl\propto n_s$ remains on the order of a few units. Note that contrary to n_s (hence r_s) which can be almost directly measured with Shubnikov-de Haas measurements, there is no generic way to measure the disorder. At high density, $r_s\ll 1$ and the interaction effects can be neglected so that k_Fl is simply related to the conductance g of the system, $g=(2e^2/h)k_Fl$. In this limit, we estimate the product $\eta\equiv r_s\sqrt{k_Fl}$ which should remain constant as one lowers the density (as $r_s\propto 1/\sqrt{n_s}$ and $k_Fl\propto n_s$). Hence, η is estimated at large density by the noninteracting formula $\eta=\sqrt{\mu e^{3/2}m^*/(4\pi\hbar^{3/2}\epsilon_0\epsilon)}$. For real samples where disorder have intrinsic characteristic lengths, the mobility μ may have a density dependence. Typical values of η are on the order of 20. In Fig. 2, we have collected the trajectories in the r_s versus $1/\sqrt{k_Fl}$ phase diagram for ten different samples: as one lowers the density, the samples follow a straight line $1/\sqrt{k_Fl}=r_s/\eta$. The regions where the samples have a metallic behavior are plotted in blue while the insulating behaviors appear in dashed red. When comparing the samples where the metallic behavior is observed with the older samples that did not show such a behavior (the two on the left), one indeed observes that the metallic behavior is observed at larger r_s (**P3**) but in the same regime of disorder (**P4**).

Property P5. The Fermi energy E_F decreases so that the

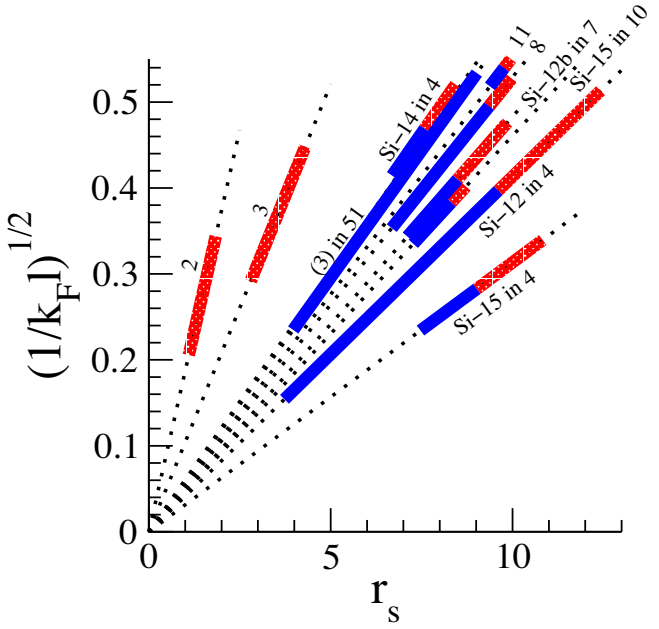

FIG. 2. (Color online) Experimental r_s versus $1/\sqrt{k_F l}$ phase diagram of the Si MOSFET, constructed from ten samples studied in the literature. Each sample corresponds to a straight line in the diagram (dotted lines). Blue (dashed red) refers to the density range where metallic $\partial\rho/\partial T>0$ (insulating $\partial\rho/\partial T<0$) behaviors are observed. This diagram should be compared with Fig. 4 of Ref. 48.

ratio T/T_F increases (where $E_F=kT_F=\pi\hbar^2 n_s/2m^*$ and T_F is the Fermi temperature). Typical experimental values are quite large, $T/T_F\approx 20\%$ around the inflexion point of the $\rho(T)$ curves as can be seen in Fig. 1(c). This last point, raised in Ref. 52 has been overlooked by many authors and is however crucial: the metallic behavior happens for temperatures which are of the same order of magnitude as the Fermi temperature. Hence, the usual "low energy" paradigm where one only looks at electrons close to the Fermi surface does not hold in those devices.

It is actually the purpose of this paper to identify the characteristic energy that controls the behavior of $\rho(T)$ and $\rho(B)$. As we shall see, both are controlled by a unique energy scale, the polarization energy E_p of the Si MOSFET.

III. A MINIMUM MODEL FOR SILICON MOSFETs

We now introduce the model that we will use for our calculations. It includes Coulomb repulsion, disorder, spin, and a twofold valley degeneracy (present in Si MOSFETs but not in their GaAs counterparts, silicon being an indirect-gap semiconductor). This model is minimum in the sense that the first three ingredients are already known to be relevant experimentally, as discussed in the above section while the fourth one is known to be present and we will show that it plays an important role for quantitative predictions. The system contains $N\gg 1$ particles in a lattice made of $L_x\times L_y$ sites, the particles being equally split up into the two degenerate valleys, with $N_\uparrow$ spin up and $N_\downarrow$ spin down. The spin configuration defines the system polarization,

$$p=\frac{N_\uparrow-N_\downarrow}{N_\uparrow+N_\downarrow}. \tag{1}$$

The Hamiltonian for the four-component plasma reads

$$H=-t\sum_{\langle\vec{r}\vec{r}'\rangle\sigma} c^\dagger_{\vec{r}\sigma}c_{\vec{r}'\sigma}+\sum_{\vec{r}\sigma} v_{\vec{r}}n_{\vec{r}\sigma}+\frac{U}{2}\sum_{\vec{r}\sigma\neq\vec{r}'\sigma'} V_{\vec{r}-\vec{r}'}n_{\vec{r}\sigma}n_{\vec{r}'\sigma'}, \tag{2}$$

where $c^\dagger_{\vec{r}\sigma}$ and $c_{\vec{r}\sigma}$ are the usual creation and annihilation operators of one electron on site $\vec{r}$ with inner degree of freedom σ, the sum $\Sigma_{\langle\vec{r}\vec{r}'\rangle\sigma}$ is restricted to nearest neighbors, and $n_{\vec{r}\sigma}=c^\dagger_{\vec{r}\sigma}c_{\vec{r}\sigma}$ is the density operator. The internal degree of freedom $\sigma=1,\ldots,4$ corresponds to the spin and valley degeneracy degrees of freedom. The disorder potential $v_{\vec{r}}$ is uniformly distributed inside $[-W/2,W/2]$. t is the hopping parameter and U is the effective strength of the two-body interaction $V_{\vec{r}}$. To reduce finite-size effects, $V_{\vec{r}}$ is obtained[53] from the bare Coulomb interaction using the Ewald summation technique,

$$V_{\vec{r}}=\sum_{\vec{L}}\frac{1}{|\vec{r}+\vec{L}|}\mathrm{Erfc}(k_c|\vec{r}+\vec{L}|)+\frac{2\pi}{L_xL_y}\sum_{\vec{K}\neq\vec{0}}\frac{1}{|\vec{K}|}\mathrm{Erfc}[|\vec{K}|/(2k_c)]\cos(\vec{K}\cdot\vec{r}). \tag{3}$$

In the previous equation, k_c is a (irrelevant) cutoff. The vector $\vec{L}$ takes discrete values $\vec{L}=(n_xL_x,n_yL_y)$ with n_x and n_y integer numbers. The vector $\vec{K}$ also takes discrete values, $\vec{K}=(\frac{2\pi}{L_x}n_x,\frac{2\pi}{L_y}n_y)$ and $(n_x,n_y)\neq(0,0)$. The complementary error function is defined as $\mathrm{Erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$. We work at small filling factor $\nu\equiv N/(L_xL_y)\ll 1$, where we recover the continuum limit. The two dimensionless parameters that control the physics read

$$r_s=U/(2t\sqrt{\pi\nu}),\quad k_Fl=48\pi\nu t^2/W^2. \tag{4}$$

IV. QUANTUM MONTE CARLO CALCULATION OF THE POLARIZATION ENERGY

This section is devoted to the numerical calculation of the polarization energy of the four-component plasma, as a function of r_s and k_Fl. We compute the ground-state energy per particle $E(p)$ of the system with the Green's-function quantum Monte Carlo (QMC) technique in the fixed node approximation (see Ref. 53 for the numerical method). Particular care was given to the extrapolation to the thermodynamics limit ($N\gg 1$) as well as the continuum limit (the algorithm being implemented on a lattice). All the fits given below, as well as data points, are given with a precision better than $\pm 0.02E_F$ (for $k_Fl\geq 1.5$) and $\pm 0.04E_F$ (for $0.3<k_Fl<1.5$). The data have been averaged on 50 to 200 samples for the strongest disorders.

At very high density, i.e., no disorder $(1/k_Fl=0)$ and interaction $(r_s=0)$, it is straightforward to show that the energy depends on the polarization p in a quadratic manner: $E(p)=E_F(1+p^2)/2$. In fact, we find that such a quadratic depen-

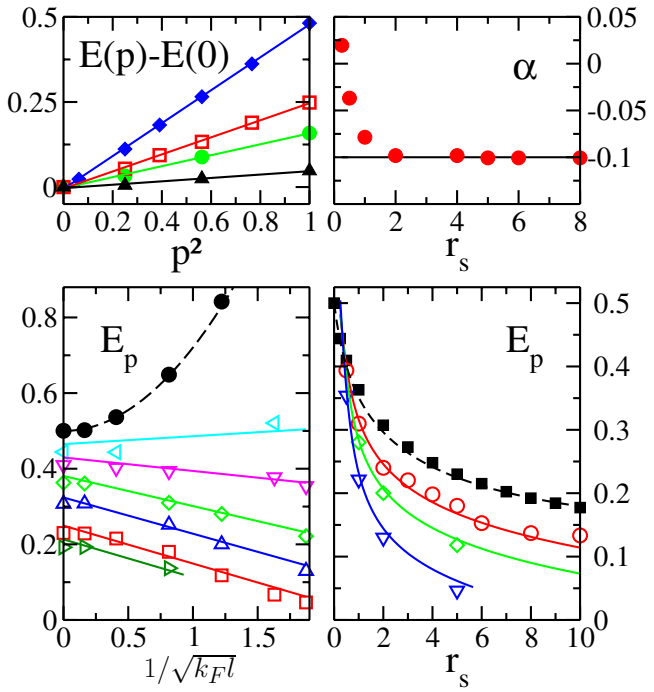

FIG. 3. (Color online) Polarization energy. Upper left panel: $E(p)$ as a function of p^2, for $1/k_Fl=0$, $r_s=0$ (diamonds), $1/k_Fl=0$, $r_s=10$ (circles), $1/k_Fl=0.66$, $r_s=2$ (squares), and $1/k_Fl=2.65$, $r_s=6$ (triangles). Straight lines are linear fits. Upper right panel: α as defined in Eq. (8) as a function of r_s. Lower left panel: E_p as a function of disorder $1/\sqrt{k_Fl}$, for $r_s=0$ (circles), $r_s=0.25$ (left triangles), $r_s=0.5$ (down triangles), $r_s=1$ (diamonds), $r_s=2$ (up triangles), $r_s=5$ (squares), and $r_s=8$ (right triangles). Dashed line stands for $E_p/E_F=1/2+\log 2/(\pi k_Fl)$. Lower right panel: E_p as a function of r_s for $1/k_Fl=0$ (squares), $1/k_Fl=0.66$ (circles), $1/k_Fl=1.49$ (diamonds), and $1/k_Fl=3.51$ (triangles). Dashed line is the result of Ref. 54 without disorder. In both lower panels, the lines are fits given by Eq. (8) for $r_s\geq 0.25$. All energies are in unit of E_F.

dence is verified with good precision for a wide range of parameters $0\leq r_s\leq 10$ and $k_Fl>1$, corresponding to the disorder regime of the experiments, as shown in the upper left panel of Fig. 3. We found (small) deviations from the quadratic behavior for rather large disorder in the localized regime (see triangles at $1/k_Fl=2.65$ and $r_s=6$). Hence we find with good precision that

$$E(p)=E(0)+E_pp^2, \tag{5}$$

where the coefficient $E_p(r_s,k_Fl)$ is the polarization energy of the system and is proportional to the Fermi energy. Without electron-electron interaction and disorder $E_p=E_F/2$, and as interaction is turned on, E_p decreases (indeed a ferromagnetic state tends to minimize Coulomb energy as antisymmetric orbital wave functions have a lower probability for two electrons to be close to each other than symmetric ones).

In presence of an *in-plane* (i.e., purely Zeeman) magnetic field B, Eq. (5) turns into $E(p)=E(0)+E_pp^2-g\mu_BBp/2$ ($g=2$ Lande factor and μ_B Bohr magneton). Introducing $B_p=4E_p/(g\mu_B)$, the minimum of energy is found for a polarization p^* given by

TABLE I. Parameters A_i of Eqs. (8) and (9).

i	0	1	2	3	4	5
A_i	27.93	9.83	56.5	46	1.77	0.019

$$p^*=B/B_p \quad B<B_p, \tag{6}$$

$$p^*=1 \quad B>B_p. \tag{7}$$

Hence, the spin susceptibility $\chi=n_s/2(dp/dB)_{B=0}$ is directly related to the field B_p at which the polarization saturates. Defining the noninteracting susceptibility $\chi_0=g\mu_Bm^*/(2\pi\hbar^2)$, one finds $E_F/E_p=2\chi/\chi_0$. In addition to the characteristic magnetic field B_p, E_p is also associated with the characteristic temperature $kT_p=E_p$ at which the polarized excited states get significantly populated.

In the lower panels of Fig. 3, we plotted our E_p data as a function of disorder $1/\sqrt{k_Fl}$ (left) and interaction r_s (right). Without disorder (squares in the lower right panel), our results are in close agreement with previous diffusion Monte Carlo calculations performed in Ref. 54 (dashed line). As expected, the polarization energy E_p decreases with interaction r_s. However, contrary to the single-valley case where E_p becomes extremely small and could even become negative (ferromagnetic instability), we find that E_p is always positive up to $r_s=20$, where $E_p=0.13E_F$.

We find that disorder tends to increase E_p for very small interactions $r_s\leq 0.25$. At $r_s=0$, the disorder correction to E_p is very well described by the second-order perturbative correction $E_p/E_F=1/2+\log 2/(\pi k_Fl)$ (dashed line in lower left panel of Fig. 3). However, as soon as small interactions are switched on ($r_s>0.3$), disorder tends to decrease E_p. We also find that the second-order perturbative correction $\propto 1/k_Fl$ is valid only for very tiny disorder $1/k_Fl<0.04$ above which the correction becomes proportional to $1/\sqrt{k_Fl}$. In order to make our set of data easily retrievable, we use the following fit, valid for our entire data set for $0.25<r_s<10$,

$$E_p(k_Fl,r_s)=E_p^{\mathrm{cl}}(r_s)+\frac{\alpha(r_s)}{\sqrt{k_Fl}}E_F+A_5E_F, \tag{8}$$

where E_p^{cl} is the polarization energy of the clean system fitted (for $0<r_s<20$) with the Pade formula,

$$E_p^{\mathrm{cl}}(r_s)=\frac{A_0+A_1r_s}{A_2+A_3r_s+A_4r_s^2}E_F. \tag{9}$$

The values of fitting parameters A_i are listed in Table I. The parameter α is plotted in the upper right part of Fig. 3 and is equal with good precision for $r_s\geq 2$ to $\alpha=-0.1$. The small parameter A_5 roughly accounts for the fact that at very low disorder, the disorder-correction to E_p/E_F is not linear but quadratic in $1/\sqrt{k_Fl}$. We note that Eqs. (8) and (9) are merely a compact way of describing our numerical data.

V. COMPARISON OF THE POLARIZATION ENERGY E_p WITH THE EXPERIMENTAL CHARACTERISTIC ENERGY SCALE

Figure 1 shows raw experimental data $\rho(B)$ and $\rho(T)$ on low-density Si MOSFETs from six different experiments, including the original set of data from Kravchenko *et al.* For each sample, we estimate the parameter η from the peak mobility μ^{peak} obtained at high density (hence probably overestimating η) which sets how $\sqrt{k_F l}=\eta/r_s$ evolves with density. Then, for each density, we calculate the polarization energy E_p according to Eqs. (8) and (9). Disorder corrections to E_p are rather small for those high-mobility samples so that our results are dominated by $E_p^{\rm cl}(r_s)$ and not very sensitive to our estimate of η. For each curve $\rho(B)[\rho(T)]$, we draw a circle at the calculated saturation field B_p (characteristic temperature T_p). The calculations are performed without any adjustable parameters and we find an extremely good match between $B_p(T_p)$ and the field (temperature) at which the resistivity saturates (has its inflexion point).

A. Magnetoresistance experiments

Describing the full $\rho(B)$ dependence is a complicated task. However, as the only effect of an *in-plane* magnetic field is to polarize the sample through Zeemann coupling (the orbital effect of an *in-plane* field can be neglected with good approximation), one expects that when the polarization saturates, the resistivity also saturates[55] so that $\rho(B)$ provides a direct measurement of B_p.

The experimental data of resistivity as a function of in-plane magnetic field B (at low temperature) shown in panels (d)–(f) of Fig. 1 correspond to three samples with mobility $\mu^{peak}=25\,000$ cm^2/V s,[8] 41 000 cm^2/V s,[50] and 20 000 cm^2/V s.[51] We find that the magnetic field B_{sat} at which the resistivity saturates is in extremely good agreement with our calculated B_p. This agreement between the experimental B_{sat} and the calculated B_p can be considered as a validation of the minimum model that we have used as well as of the accuracy of the QMC method.

Let us now discuss a publication[56] which claimed to observe a divergence of the spin susceptibility χ. Such a finding would contradict our calculation as $\chi \propto 1/B_p$ and although B_p decreases as one lowers the density, we did not observe any divergence in the numerical calculations (one should however keep in mind that in the lower right panel of Fig. 3, the energies are in unit of E_F which itself is proportional to the electronic density n_s). The corresponding data are shown in the right panel of Fig. 4 (our B_p: full line and experimental B_{sat}: dashed line). The experimental B_{sat} were obtained in Ref. 56 from rescaling the in-plane magnetoresistivity data ($\mu^{peak}=30\,000$ cm^2/V s) at small field. Here, we find a good, but not very good, agreement between B_p and B_{sat}. To understand the discrepancy, let us come back to the raw $\rho(B)$ experimental data for two values of the density corresponding to the blue squares (experiments) and red circles (our numerics). Those raw data are shown on the left panel of Fig. 4 where we also reproduce our predictions as well as the value of B_{sat} obtained from the rescaling procedure. Clearly, the rescaling procedure underestimates B_{sat} as the resistivity has not saturated yet at this value while our calculated B_p matches precisely the end of the resistivity rise. We find that the agreement between the experiments and the numerics is actually extremely good and that the rescaling procedure only give approximate estimates of B_{sat}. As the divergence of χ claimed in Ref. 56 is extremely sensitive to this procedure, we conclude that there is probably no divergence of the susceptibility as defended in Refs. 57 and 58 but in contradiction with Ref. 56. New thermodynamic measurements are in qualitative[60] and quantitative[59] agreement with our numerics (hence we disagree with the conclusions of Ref. 59 which claim that there is a divergence of the spin susceptibility).

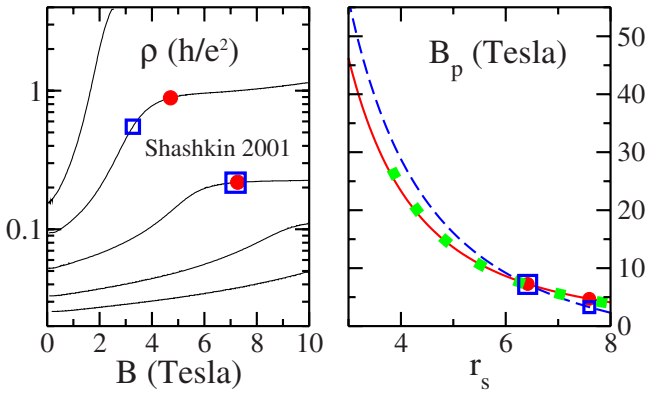

FIG. 4. (Color online) Left panel: ρ as a function of B at different n_s, from Ref. 56. Right panel: experimental B_{sat} from Ref. 56 (dashed line), QMC B_p (full line), and experimental B_p from Ref. 55 (green squares) as a function of r_s. In both panels, red points correspond to QMC B_p and blue squares to B_{sat} from Ref. 56, at two fixed densities.

Finally, we add a last data set that corresponds to the measures made in Ref. 55 (green squares in right panel of Fig. 4) from the analysis of Shubnikov-de Haas oscillations, at a similar mobility ($\mu^{peak}=24\,000$ cm^2/V s). We find that they agree perfectly with our calculations. Overall, we have a quantitative agreement between B_p and B_{sat} for five different experiments indicating that our minimum model and QMC calculations adequately describe the experimental situation.

B. Resistance versus temperature experiments

We now turn to the original puzzle, the resistivity versus temperature $\rho(T)$ experiments. A first remark that was originally made in Ref. 52 is that the large increase in resistance takes place at large temperatures compare to usual transport experiments. This can be seen for instance in Fig. 1(c) where temperature has been rescaled with respect to the Fermi temperature T_F. One finds that the inflexion point of $\rho(T)$ takes place for $T/T_F=0.2$. Refs. 52 and 61 attributed this large energy scale to some semiclassical effect. Here however, we argue that this characteristic scale is the polarization temperature T_p. On a qualitative level, it is not surprising: this energy scale is equal to a fraction of the Fermi temperature ($0.5T_F$ without interaction and disorder and less in their presence) and we have already showed that it controls the magnetoresistance behavior.

More quantitatively, in panels (a)–(c) of Fig. 1, we plot various experimental data of resistivity ρ as a function of temperature T, at zero magnetic field and at various electronic densities, for three Si MOSFETs with, respectively, $\mu^{peak}=30\,000$ cm^2/V s (Ref. 7), 41 000 cm^2/V s (Ref. 10), and 19 600 cm^2/V s (Ref. 49). We have superimposed the circles corresponding to quantum Monte Carlo polarization temperatures T_p. The typical temperature where one observes the change in resistivity associated with the metallic behavior matches our calculated T_p. This is a strong indication that the polarization energy is the unique energy scale that controls the physics of the metallic behavior.

VI. A MINIMUM SCENARIO FOR THE 2D METAL-INSULATOR TRANSITION

We now have a good understanding of what happens on the horizontal axis of the $\rho(T)$ and $\rho(B)$ curves. To gain a full understanding of properties **P1** and **P2**, we need to understand what happens on the vertical axis, i.e., why does the resistivity increase and not decrease upon increasing temperature or magnetic field. In other words, in order to account for both **P1** and **P2**, it is necessary and sufficient to show that *polarized states have a higher resistivity than nonpolarized one* (hereafter referred as property **P6**). Indeed, it is necessary since the $\rho(B)$ curves show that the polarized system (high field) has a higher resistivity than the nonpolarized one (zero field). It is also sufficient as, once **P6** is established, it naturally follows that when one will increase temperature in a range around T_p, the highly resistive excited states will be significantly populated and the overall resistivity will increase.

In a previous publication,[48] we have performed a systematic study of the interplay between Anderson localization and electron-electron interactions in Si MOSFETs. We found that upon increasing interactions, the localization length of the nonpolarized ground state strongly increases (in absolute value but, in particular, with respect to the polarized excited states) so that one naturally accounts for **P6**. This result is also shown in the right panel of Fig. 5 where we compare as a function of r_s the localization lengths ξ (in unit of the average distance between electrons a) computed at zero temperature for the nonpolarized electron gas and for the fully polarized one. At very weak interaction ($r_s \lesssim 0.5$), the polarized system is less localized than the nonpolarized one. Indeed, polarizing the system raises the Fermi level E_F (by a factor 2), hence also $k_F l$, hence ξ which depends exponentially on the latter. On the contrary, at stronger interaction ($r_s \gtrsim 0.5$), the situation is reversed and the polarized system is more localized than the nonpolarized one. We checked that we also get such an inversion of the localization lengths between the nonpolarized ground state and *partially* polarized states so that we indeed recover **P6**. In Fig. 4 of Ref. 48 we presented the effective phase diagram constructed out of this mechanism which should be compared with the experimental one presented in Fig. 2. The agreement is semiquantitative, meaning that our scenario also captures properties **P3** and **P4** (as well as **P5** discussed above). Hence, we find that the minimum model discussed in this paper is enough to account for all the relevant experimental facts discussed in the beginning of this paper.

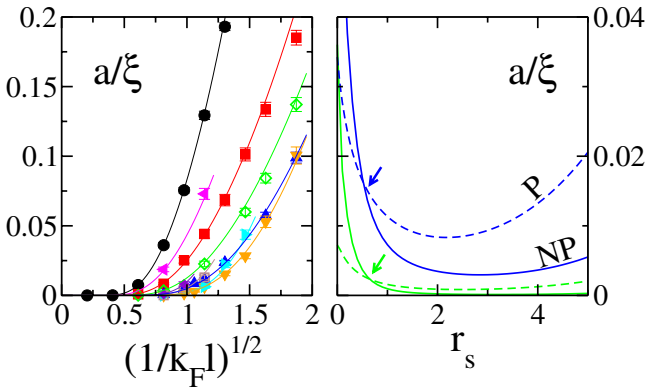

FIG. 5. (Color online) Left panel: a/ξ as a function of $1/\sqrt{k_F l}$ for the nonpolarized Si MOSFET. The symbols correspond to different strengths of interaction: from left to right, $r_s=0$ (circles), $r_s=0.1$ (left triangles), $r_s=0.25$ (full squares), $r_s=0.5$ (empty diamonds), $r_s=1$ (up triangles), then from right to left, $r_s=2$ (down triangles), $r_s=4$ (right triangles), $r_s=6$ (empty squares), and $r_s=10$ (full diamonds). Right panel: a/ξ as a function of r_s for the nonpolarized (NP, full lines) and polarized (P, dashed lines) Si MOSFET, at $k_F l=1.51$ (two lower green curves) and $k_F l=0.89$ (two upper blue curves). The arrows indicate the interaction strength r_s at which the polarized excited state becomes more localized than the nonpolarized ground state. In both panels, the lines are given by Eqs. (10)–(12).

We end this section with a discussion of the existence of "true" metal at zero temperature, i.e., does the localization length diverge in the region where the metallic behavior is observed. We emphasize that in light of the above discussion, this is a rather academic question as what has actually been measured experimentally corresponds to rather high energy physics. In Ref. 48, we have claimed that there probably does not exist a true metal at zero temperature. Our main argument was the absence of any visible deviation to the one-parameter scaling theory[1] in presence of interactions, despite a huge delocalization effect. Here we give another argument. In the left panel of Fig. 5, we plot a/ξ as a function of the disorder strength $1/\sqrt{k_F l}$ for various values of r_s. Without interaction,[62] the localization length of the two-dimensional electron gas depends exponentially on $k_F l$: $\xi/a=(k_F l/B)\exp(Ck_F l)$. We find that in the presence of interactions our entire set of data can be fitted by the same law upon renormalizing the B and C parameters. Hence, while the localization length can become large in the presence of interaction, it always remains finite and the system remains ultimately an insulator. In practice this insulating behavior should appear at extremely low temperature when the phase coherence length becomes larger than the localization length. Such a metal-to-insulator re-entrance as a function of temperature has been observed at high density[63,64] and at low density, in the vicinity of the so-called metal-insulator transition, in a GaAs two-dimensional hole gas[65] and in Si MOSFETs[66].

Below we provide the fitting parameters that account for our localization length data. We have

TABLE II. Parameters b_i and c_i of Eqs. (11) and (12) for the nonpolarized Si MOSFET (b_i^{NP} and c_i^{NP}) and for the polarized one (b_i^{P} and c_i^{P}).

i	0	1	2	3	4	5	6
b_i^{NP}	0.191	−0.325	0.24	−0.0526	0.0035	0	0
b_i^{P}	0.136	−0.09093	−0.03197	0.18232	−0.17749	0.07549	−0.010827
c_i^{NP}	0.84	−0.177	1.094	2.0616	−1.6784	0.3027	0
c_i^{P}	1.68	−0.712	2.418	−1.1874	0.1594	0	0

$$\frac{a}{\xi}(k_F l, r_s) = \frac{B(r_s)}{k_F l}\exp[-C(r_s)k_F l], \tag{10}$$

where two different sets of parameters are needed for the nonpolarized ($B^{\mathrm{NP}}, C^{\mathrm{NP}}$) and the polarized ($B^{\mathrm{P}}, C^{\mathrm{P}}$) cases. In turn, the parameters B and C are well fitted by

$$B(r_s) = b_0 + b_1\sqrt{r_s} + b_2 r_s + b_3 r_s\sqrt{r_s} + b_4 r_s^2 + b_5 r_s^2\sqrt{r_s} + b_6 r_s^3 \tag{11}$$

and

$$C(r_s) = c_0 + c_1\sqrt{r_s} + c_2 r_s + c_3 r_s\sqrt{r_s} + c_4 r_s^2 + c_5 r_s^2\sqrt{r_s} + c_6 r_s^3, \tag{12}$$

where the values of the b_i and c_i fitting parameters are given in Table II for the nonpolarized system (b_i^{NP} and c_i^{NP}) and for the polarized one (b_i^{P} and c_i^{P}), with the convention that $k_F l$ always refers to the Fermi level of the nonpolarized gas. Equation (10) supports the idea that the two-dimensional electron gas behaves as a Fermi liquid (at least for moderate interactions $r_s \leq 10$) and that the effect of electron-electron interactions is to renormalize its effective characteristics.

VII. CONCLUSION

Let us discuss a few theoretical scenarios that have been proposed by other authors. In light of the findings of this paper, we believe that models whose characteristic energies do not involve the polarization energy should not be applied to those experiments. This includes, for example, the percolation scenario proposed in Ref. 36 (although percolation probably plays an important role close to the insulating region in GaAs heterostructures). This also includes the proposal related to Wigner crystal Ref. 24 (which we dismissed in Ref. 53 on other grounds). Reference 25 has some overlap with our scenario (on the role of polarization energy in particular) but relies on the existence of some form of local order (which we did not observe in our simulations) and does not include the presence of disorder. Extrinsic models as the temperature-dependent disorder scenario proposed in Refs. 31 and 32 can also be excluded on the same ground (but these authors were the first to recognize the presence of a high energy scale in the experiments). An important theoretical advance was made in Ref. 23 where the authors studied the interplay between interaction and disorder in the diffusive limit, in the limit of an infinitely large number of valleys. We believe that those authors correctly pointed out the important role of valley degeneracy in the experiments at the origin of the large differences between the behaviors observed in silicon MOSFETs and in other heterostructures. However, the limit of an infinite number of valleys artificially increases the role of electron-electron interactions. In particular, those authors find that even a very weak interaction will drive the system toward a non-Fermi-liquid fixed point. Those results, which imply a complete breakdown of one-parameter scaling theory are in contradiction with our numerical results.[48] More importantly, the corresponding analysis of the experiments[67,68] is in contradiction with the present analysis of the characteristic energy scales involved in these systems.

In summary, we have considered a *minimum* model of Si MOSFETs taking into account Coulomb repulsion, spin, valley degeneracy, and disorder in a nonperturbative way. Our chief result is that the polarization energy E_p calculated for this model is in quantitative agreement with the characteristic energy scale that controls the metallic behavior of those high-mobility Si MOSFETs. Beside spin, Coulomb repulsion plays a crucial role here as it decreases E_p by a factor 2–3 (see Fig. 3). The presence of valley degeneracy is also very important as it also decreases E_p by a factor 2. On the other hand, disorder only gives corrections of 10–30% to E_p for the samples considered in this survey. We note that the polarization temperature is the "crossover temperature" that was forseen in Ref. 52. However, we find that this crossover temperature is an intrinsic property of the electron gas and does not involve any additional extrinsic ingredient. A simple corollary of our chief result is that the nonpolarized system has to be a much better conductor than the polarized one (to explain the sign of $d\rho/dT$ and $d\rho/dB$ in the metallic region). We have argued before[48] that our minimum model accounts for this counterintuitive point (**P6**), as the interplay between Coulomb repulsion and Anderson localization depends strongly on the polarization. These findings imply that our "necessary" model (in the sense that all its ingredient are known to be present and *a priori* relevant) is also "sufficient" to capture the essential features of the metallic behaviors. Another consequence of our model is the apparent failure of one-parameter scaling. *For a given polarization*, we could not find any deviation to the (one-parameter) scaling theory of localization.[48] However, at finite temperature excited states of different localization lengths come into play so that the physics is no longer controlled by a single parameter. In this work, we have focused on Si MOSFETs but at the qualitative level, many statements also apply to other materials such as GaAs heterostructures. In our view, the situation in the latter is made a bit different by the conjunction of three

elements. First the absence of valley degeneracy makes the delocalization effect of the ground state much less effective. Second, the present of the doping layer close to the gas induces some long-range disorder which can mask the local physics. Last, extremely high mobility are available for these systems, making it possible to study almost ballistic samples.

ACKNOWLEDGMENTS

We thank S. De Palo and G. Senatore for interesting discussions as well as M. Reznikov for valuable comments on the manuscript. Support from the CCRT supercomputing facilities is also acknowledged.

[1] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).
[2] D. J. Bishop, D. C. Tsui, and R. C. Dynes, Phys. Rev. Lett. **44**, 1153 (1980).
[3] M. J. Uren, R. A. Davies, and M. Pepper, J. Phys. C **13**, L985 (1980).
[4] S. V. Kravchenko, G. V. Kravchenko, J. E. Furneaux, V. M. Pudalov, and M. D'Iorio, Phys. Rev. B **50**, 8039 (1994).
[5] S. V. Kravchenko and M. P. Sarachik, Rep. Prog. Phys. **67**, 1 (2004).
[6] V. M. Pudalov, arXiv:cond-mat/0405315 (unpublished).
[7] S. V. Kravchenko, W. E. Mason, G. E. Bowker, J. E. Furneaux, V. M. Pudalov, and M. D'Iorio, Phys. Rev. B **51**, 7038 (1995).
[8] K. M. Mertes, D. Simonian, M. P. Sarachik, S. V. Kravchenko, and T. M. Klapwijk, Phys. Rev. B **60**, R5093 (1999).
[9] D. Popovic, A. B. Fowler, and S. Washburn, Phys. Rev. Lett. **79**, 1543 (1997).
[10] V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, JETP Lett. **68**, 442 (1998).
[11] S. V. Kravchenko and T. M. Klapwijk, Phys. Rev. Lett. **84**, 2909 (2000).
[12] Y. Hanein, D. Shahar, J. Yoon, C. C. Li, D. C. Tsui, and H. Shtrikman, Phys. Rev. B **58**, R13338 (1998).
[13] E. Ribeiro, R. D. Jäggi, T. Heinzel, K. Ensslin, G. Medeiros-Ribeiro, and P. M. Petroff, Phys. Rev. Lett. **82**, 996 (1999).
[14] S. J. Papadakis and M. Shayegan, Phys. Rev. B **57**, R15068 (1998).
[15] P. T. Coleridge, R. L. Williams, Y. Feng, and P. Zawadzki, Phys. Rev. B **56**, R12764 (1997).
[16] Y. Hanein, U. Meirav, D. Shahar, C. C. Li, D. C. Tsui, and H. Shtrikman, Phys. Rev. Lett. **80**, 1288 (1998).
[17] A. P. Mills, Jr., A. P. Ramirez, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **83**, 2805 (1999).
[18] M. Y. Simmons, A. R. Hamilton, M. Pepper, E. H. Linfield, P. D. Rose, D. A. Ritchie, A. K. Savchenko, and T. G. Griffiths, Phys. Rev. Lett. **80**, 1292 (1998).
[19] K. Lai, W. Pan, D. C. Tsui, S. A. Lyon, M. Mühlberger, and F. Schäffler, Phys. Rev. B **72**, 081313(R) (2005).
[20] B. L. Altshuler, A. G. Aronov, and P. A. Lee, Phys. Rev. Lett. **44**, 1288 (1980).
[21] A. M. Finkelshtein, Sov. Phys. JETP **57**, 97 (1983).
[22] C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. B **30**, 527 (1984).
[23] A. Punnoose and A. M. Finkel'stein, Science **310**, 289 (2005).
[24] B. Spivak, Phys. Rev. B **67**, 125205 (2003).
[25] A. Camjayi, K. Haule, V. Dobrosavljevic, and G. Kotliar, Nat. Phys. **4**, 932 (2008).
[26] D. Belitz and T. R. Kirkpatrick, Phys. Rev. B **58**, 8214 (1998).
[27] P. Phillips, Y. Wan, I. Martin, S. Knysh, and D. Dalidovich, Nature (London) **395**, 253 (1998).
[28] J. S. Thakur and D. Neilson, Phys. Rev. B **58**, 13717 (1998).
[29] S. Das Sarma and E. H. Hwang, Phys. Rev. Lett. **83**, 164 (1999).
[30] T. M. Klapwijk and S. Das Sarma, Solid State Commun. **110**, 581 (1999).
[31] B. L. Altshuler and D. L. Maslov, Phys. Rev. Lett. **82**, 145 (1999).
[32] B. L. Altshuler, D. L. Maslov, and V. M. Pudalov, Phys. Status Solidi B **218**, 193 (2000).
[33] S. J. Papadakis, E. P. De Poortere, H. C. Manoharan, M. Shayegan, and R. Winkler, Science **283**, 2056 (1999).
[34] S. S. Murzin, S. I. Dorozhkin, G. Landwehr, and A. C. Gossard, JETP Lett. **67**, 113 (1998).
[35] Y. Yaish, O. Prus, E. Buchstab, S. Shapira, G. Ben Yoseph, U. Sivan, and A. Stern, Phys. Rev. Lett. **84**, 4954 (2000).
[36] Y. Meir, Phys. Rev. Lett. **83**, 3506 (1999).
[37] J. Shi, S. He, and X. Xie, arXiv:cond-mat/9904393 (unpublished).
[38] S. He and X. C. Xie, Phys. Rev. Lett. **80**, 3324 (1998).
[39] R. Leturcq, D. L'Hote, R. Tourbot, C. J. Mellor, and M. Henini, Phys. Rev. Lett. **90**, 076402 (2003).
[40] S. Das Sarma, M. P. Lilly, E. H. Hwang, L. N. Pfeiffer, K. W. West, and J. L. Reno, Phys. Rev. Lett. **94**, 136401 (2005).
[41] M. J. Manfra, E. H. Hwang, S. Das Sarma, L. N. Pfeiffer, K. W. West, and A. M. Sergent, Phys. Rev. Lett. **99**, 236402 (2007).
[42] G. Benenti, X. Waintal, and J.-L. Pichard, Phys. Rev. Lett. **83**, 1826 (1999).
[43] R. Berkovits and J. W. Kantelhardt, Phys. Rev. B **65**, 125308 (2002).
[44] R. Kotlyar and S. Das Sarma, Phys. Rev. Lett. **86**, 2388 (2001).
[45] T. Vojta, F. Epperlein, and M. Schreiber, Phys. Rev. Lett. **81**, 4212 (1998).
[46] B. Srinivasan, G. Benenti, and D. L. Shepelyansky, Phys. Rev. B **67**, 205112 (2003).
[47] P. J. H. Denteneer, R. T. Scalettar, and N. Trivedi, Phys. Rev. Lett. **83**, 4610 (1999).
[48] G. Fleury and X. Waintal, Phys. Rev. Lett. **101**, 226803 (2008).
[49] A. Prinz, V. M. Pudalov, G. Brunthaler, and G. Bauer, Superlattices Microstruct. **27**, 301 (2000).
[50] V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, JETP Lett. **65**, 932 (1997).
[51] S. A. Vitkalov, H. Zheng, K. M. Mertes, M. P. Sarachik, and T. M. Klapwijk, Phys. Rev. Lett. **87**, 086401 (2001).
[52] B. L. Altshuler, G. W. Martin, D. L. Maslov, V. M. Pudalov, A. Prinz, G. Brunthaler, and G. Bauer, arXiv:cond-mat/0008005 (unpublished).
[53] X. Waintal, Phys. Rev. B **73**, 075417 (2006).
[54] M. Marchi, S. De Palo, S. Moroni, and G. Senatore, Phys. Rev. B **80**, 035103 (2009).

[55] T. Okamoto, K. Hosoya, S. Kawaji, and A. Yagi, Phys. Rev. Lett. **82**, 3875 (1999).
[56] A. A. Shashkin, S. V. Kravchenko, V. T. Dolgopolov, and T. M. Klapwijk, Phys. Rev. Lett. **87**, 086801 (2001).
[57] V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, arXiv:cond-mat/0103087 (unpublished).
[58] O. Prus, Y. Yaish, M. Reznikov, U. Sivan, and V. Pudalov, Phys. Rev. B **67**, 205407 (2003).
[59] A. A. Shashkin, S. Anissimova, M. R. Sakr, S. V. Kravchenko, V. T. Dolgopolov, and T. M. Klapwijk, Phys. Rev. Lett. **96**, 036403 (2006).
[60] N. Teneh, A. Yu. Kuntsevich, D. L. Maslov, V. M. Pudalov, T. M. Klapwijk, and M. Reznikov, arXiv:0910.5724v1 (unpublished).
[61] G. Brunthaler, A. Prinz, G. Bauer, and V. M. Pudalov, Phys. Rev. Lett. **87**, 096802 (2001).
[62] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[63] A. R. Hamilton, M. Y. Simmons, M. Pepper, E. H. Linfield, P. D. Rose, and D. A. Ritchie, Phys. Rev. Lett. **82**, 1542 (1999).
[64] V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, Phys. Rev. B **60**, R2154 (1999).
[65] J. Huang, J. S. Xia, D. C. Tsui, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **98**, 226801 (2007).
[66] O. Prus, M. Reznikov, U. Sivan, and V. Pudalov, Phys. Rev. Lett. **88**, 016801 (2001).
[67] S. Anissimova, S. V. Kravchenko, A. Punnoose, A. M. Finkel'stein, and T. M. Klapwijk, Nat. Phys. **3**, 707 (2007).
[68] D. A. Knyazev, O. E. Omel'yanovskii, V. M. Pudalov, and I. S. Burmistrov, Phys. Rev. Lett. **100**, 046405 (2008).