The 2D metal-insulator transition as a strong localization induced crossover phenomenon

S. Das Sarma$^1$ and E. H. Hwang$^{1,2}$

$^1$Condensed Matter Theory Center, Department of Physics,
University of Maryland, College Park, Maryland 20742-4111

$^2$SKKU Advanced Institute of Nanotechnology and Department of Physics,
Sungkyunkwan University, Suwon, 440-746, Korea

(Dated: January 21, 2014)

Low-disorder and high-mobility two-dimensional (2D) electron (or hole) systems confined in semiconductor heterostructures undergo an apparent metal-insulator-transition (MIT) at low temperatures as the carrier density ($n$) is varied. In some situations, the 2D MIT can be caused at a fixed low carrier density by changing an externally applied in-plane magnetic field parallel to the 2D layer. The effective metallic (insulating) phase is characterized by the sign of the temperature dependence of the low-temperature resistivity $\rho(n,T)$ with $d\rho/dT > 0$ ($< 0$) operationally defining the apparent metallic (insulating) phase. Thus, $d\rho/dT < 0$ at low temperature changes its sign at some nonuniversal sample-dependent critical carrier density $n_c$ separating an effective 2D metal ($d\rho/dT > 0$) for $n > n_c$ from an effective 2D insulator ($d\rho/dT < 0$) for $n < n_c$. We study the 2D MIT phenomenon theoretically in the current work as a possible strong localization induced crossover process controlled by the Ioffe-Regel criterion, $k_Fl = 1$, where $k_F(n)$ is the 2D Fermi wave vector and $l(n,T)$ is the disorder-limited quantum mean free path on the metallic side. Calculating the quantum mean free path in the effective metallic phase from a realistic transport theory including disorder scattering effects, we solve the integral equation (with $l$ depending on $n$ and $T$ through multidimensional integrals) defined by the Ioffe-Regel criterion to obtain the nonuniversal critical density $n_c$ as a function of the applicable physical experimental parameters including disorder strength, in-plane magnetic field, spin and valley degeneracy, background dielectric constant and carrier effective mass, and temperature. The key physics underlying the nonuniversal parameter dependence of the critical density is the temperature and density dependence of the Coulomb disorder. Our calculated results for the crossover critical density $n_c$ (and the associated effective metal/insulator or density/disorder phase diagram) appear to be in qualitative and semi-quantitative agreement with the available experimental data in different 2D semiconductor systems lending credence to the possibility that the apparent 2D MIT signals the onset of the strong localization crossover in disordered 2D systems. We provide some results for graphene where a low-temperature 2D MIT becomes possible in the presence of inter valley scattering. We also provide an extensive comparison with the theoretical results obtained on the basis of 2D MIT being considered as a percolation transition.

I. INTRODUCTION

Carrier transport in 2D semiconductor structures [e.g., Si inversion layers in MOSFETs, 2D electron gas (2DEG) or 2D hole gas (2DHG) GaAs-AlGaAs heterojunctions and quantum wells or in Si-SiGe quantum wells] is a strong function of carrier density ($n$) and temperature ($T$)\cite{1-2}. This is true in graphene as well, which is in some sense the ultimate 2D carrier system, as has been discussed in depth in ref.\cite{8}. We provide some limited results for graphene in this paper, but our main focus is 2D electron (or hole) transport in Si or GaAs based semiconductor heterostructures where 2D MIT has been discussed extensively\cite{11-12} in contrast to graphene where the 2D MIT phenomenon has been discussed only occasionally in a few publications in the literature\cite{13-14}. The study of 2D electronic (which include holes also in 2D p-doped structures) transport at low temperatures divides itself naturally into two distinct areas: effectively “metallic” transport at high carrier density manifesting a weak or moderately positive $d\rho/dT \gtrsim 0$ and effective insulating transport dependence at low carrier density manifesting a large negative $d\rho/dT < 0$ at low temperatures, $\rho(T)$ is the temperature dependent 2D resistivity. The current work aims at a theoretical understanding of the density-tuned crossover behavior between this high-density effective metallic and the low-density effectively insulating transport behavior at low-temperatures. Consistent with the widely used terminology, we refer to this density tuned low-temperature phenomenon (i.e., a change of sign in $d\rho/dT$) as the 2D metal-insulator transition (“2D MIT”), but it is more likely to be a crossover behavior rather than an actual quantum phase transition, and indeed in the current work we treat the 2D MIT phenomenon manifestly as a crossover behavior (as already made explicit in the title of this paper) characterized by the Ioffe-Regel criterion. We emphasize right here in the beginning that ours is not a critical theory for a quantum phase transition.

Prior to 1979 – 80, when the scaling theory of localization\cite{15} came into being, the 2D MIT phenomenon in Si inversion layers (based on Si-SiO$_2$ MOSFET structures) was universally considered to be an example of the Anderson localization transition\cite{16-17}, where decreasing 2D carrier density drives the system from being a 2D metal at high density to being a 2D insulator at lower
density as the Fermi level moves through a mobility edge. The scaling theory of localization established two to be the lower critical dimensionality in the non-interacting localization problem, and thus all disordered 2D electronic states are now (i.e., after 1979) thought to be strictly localized although the localization length is exponentially long for weak disorder, making it essentially impossible to distinguish a true extended metallic state from a weakly localized insulating state in finite samples (or at finite temperatures) with the localization length exceeding the sample size (or the temperature-dependent inelastic phase breaking length). For high disorder, however, the 2D system becomes exponentially localized (“strong localization”) and the crossover from “weak localization” (which is a logarithmically weak effect rather difficult to observe experimentally in zero magnetic field) to “strong localization” (which is true exponential Anderson localization with the single-particle wavefunction falling off exponentially with distance in contrast to weak localization with exponentially long localization length) is the subject matter of the current work. With no loss of generality, the weakly localized higher-density phase will be referred to as the “metallic” phase in this paper and the lower density strongly localized phase will be referred to as the “insulating” phase. Experimentally, the two phases (i.e., high-density weakly-localized or “metallic” and low-density strongly localized or “insulating”) are distinguishable by the temperature dependence of their respective low-temperature resistivity with the strongly localized low-density phase manifesting the exponential insulating behavior with a strongly exponentially increasing resistivity with decreasing temperature whereas the metallic high-density phase either exhibiting almost no temperature dependence in its resistivity (at very high density) or the resistivity decreasing (as an effective power law) with decreasing temperature (at not very high density). In principle, the “metallic” phase should manifest a weak logarithmic insulating temperature dependence of low enough temperatures because of the weak localization property of the 2D metallic phase as predicted by the scaling theory, but the observation of such a weak logarithmic insulating temperature dependence is challenging since it is easily overwhelmed by any other power-law temperature dependence in the system (except perhaps at extremely low temperatures) arising, for example, from phonon or screening effects. At very low temperatures, the logarithmic insulating temperature dependence should dominate even in high-density samples, but such low electronic temperatures are often impossible to reach because of carrier heating problems invariably present in semiconductors. In fact, there are only a few reported observations\textsuperscript{14,15} of the weak logarithmic insulating temperature dependence in 2D systems (and in all these reports, the logarithmic correction is small and saturates at sufficiently low temperatures), and we ignore the 2D weak localization complications in the rest of this paper, referring to the higher-density phase as an effective 2D metal (and the lower-density strongly localized phase as a 2D insulator).

Our main goal in this paper is to theoretically study the weak to strong localization transition (or the transition from the effective metal to the effective insulator) as a function of carrier density at a fixed low temperature. It is known that this transition, which is technically a crossover, happens around \(k_Fl = 1\) as defined by the so-called Ioffe-Regel criterion, where \(k_F\) is the 2D Fermi wave vector and \(l\) is the disorder-induced elastic quantum mean free path. We note that \(k_F \propto \sqrt{n}\) in 2D systems, but \(l\) itself is a complicated functional of \(n\) defined through a complex integral function \(l(n)\). Thus, the condition \(k_Fl = 1\) would define an effective crossover “critical density” \(n_c\) with \(n > n_c\) being the 2D metallic phase (where \(l > k_F^{-1}\)) and \(n < n_c\) being the (strongly localized) 2D insulating phase (where \(l < k_F^{-1}\) in our theory). We emphasize that the \(n > n_c\) metallic phase is only an effective metal in finite samples. We note that sometimes the Ioffe-Regel criterion is written as \(l = \lambda_F = 2\pi/k_F\) so that \(k_Fl = 2\pi\) defines the crossover MIT critical density. Sometimes the criterion is even written as \(k_Fl = 1/2\pi\), but all of these different conditions imply exactly the same qualitative dependence of \(n_c\) on disorder and system parameters (but obviously with differing quantitative magnitudes of \(n_c\)). We will use \(k_Fl = 1\) to define \(n_c\) throughout this paper with no loss of generality since the precise quantitative value of \(n_c\) cannot really be compared with experiments any way as the precise disorder configurations in the 2D samples are typically unknown and vary strongly from sample to sample.

In addition to obtaining the qualitative dependence of the crossover density \(n_c\) for 2D MIT on the 2D materials parameters (e.g., effective mass, valley degeneracy, dielectric constant, etc.), our main goal would be to ascertain the qualitative dependence of \(n_c\) on disorder. Since disorder determines the mean free path \(l\), the crossover ‘critical’ density \(n_c\), as obtained from the condition \(k_Fl = 1\), would depend crucially on disorder parameters. More disordered the system, the higher would be \(n_c\) since the effective mean free path would be smaller for larger disorder. Since the effective disorder (and therefore the mean free path itself) is temperature dependent, \(n_c\) would manifest an implicit temperature dependence through the temperature dependence of the 2D metallic conductivity, which we would also study theoretically. Another aspect of experimental interest we study theoretically in this paper is the effect of a parallel magnetic field which tends to spin-polarize the system leading to suppressed screening and hence enhanced effective disorder (and thus a suppressed metallic conductivity and a reduced mean free path). Thus, an in-plane applied magnetic field increases \(n_c\) (since it decreases the effective mean free path) consistent with experimental observations. In fact, spin polarization through the application of an in-plane magnetic field would also effectively enhance \(k_F\), but this effect is usually more than compensated by the reduced screening effect, but, in principle, it...
is possible for $n_c$ to increase in some situations where the quantitative effect of the reduction of $k_F$ by the applied field is more important than the suppression of screening. Here, the dependence of the crossover critical density $n_c$ on disorder, temperature, and magnetic field is the main focus of our current theoretical work.

It is important to emphasize that our work is purely phenomenological in nature where the Ioffe-Regel criterion, $k_Fl = 1$, plays the central role in determining the crossover density for 2D MIT. Whether this specific description for 2D MIT is valid or not can only be decided 
\textit{a posteriori} by comparing between our calculated theoretical results for $n_c$ (as a function of disorder, magnetic field, and temperature) and the corresponding experimental results. We should mention right now that our calculated $n_c$ appears to be in reasonable qualitative and semi-quantitative agreement with the available experimental results (as would later be discussed in this article) on 2D MIT in the literature although detailed quantitative comparisons are difficult since independent experimental information on the applicable disorder in the relevant 2D systems is unavailable. We assume in this work that the dominant disorder in the semiconductor structures undergoing 2D MIT arises from uncorrelated random quenched charged impurities in the background.

There is considerable experimental evidence\textsuperscript{15}–\textsuperscript{23} that unintentional random charged impurities in the background and at the interface as well as the remote charged donors (e.g., in modulation doped heterostructures and quantum wells) are the main resistive scattering sources in 2D semiconductor systems at low temperatures (and at low carrier densities where 2D MIT occurs). In Si-MOSFETs the short-range disorder associated with Si-SiO\textsubscript{2} interface certainly plays a role in resistive scattering at higher carrier density (when the 2D electron gas is pushed very close to the interface by the self-consistent electric field generated by the electrons themselves), but for $n \gtrsim n_c$ the main scattering source is the charged impurity disorder even for Si-MOSFETs provided that $n_c$ is not too large\textsuperscript{24}. We therefore neglect all disorder mechanisms other than random charged impurities in the system, which we parameterize using only two parameters: $n_i$ and $d$ [a 2D density of charged impurity centers of concentration $n_i$, distributed randomly in a plane a distance $d'$ away from the semiconductor-insulator interface where the 2D carrier system is localized – we assume the charged impurities to be of random positive and negative sign of unit strength (i.e., of magnitude $e$, the electron charge) with a net charge of zero]. We emphasize that it is straightforward to include in the theory additional (as well as more realistic) types of disorder (e.g., interface roughness, alloy disorder, bulk 3D distribution of random charged impurities) and/or a more general model for charged impurity disorder where some correlations among impurity positions are included in the theory\textsuperscript{25}.

We neglect all these non-essential details simply because including them would require many more additional unknown parameters to characterize the system disorder compared with our minimal model of disorder which is characterized by only two parameters $n_i$ and $d$. Since our minimal model (characterized by only two parameters $n_i$ and $d$) already captures the essential physics of 2D MIT as well as that of 2D metallic transport, we believe that this minimal model should suffice for our theoretical purpose. We note in this context that even this minimal model contains two (in general unknown) parameters ($n_i$ and $d$) defining the semiconductor disorder which implies that the sample mobility (for example, at some high designated density $n \gg n_c$) by itself cannot uniquely define the disorder. Thus, in general, $n_c$ cannot be a unique function of the sample mobility (at high density) although it will be a unique function of $n_i$ and $d$ for a given system. This immediately leads to the conclusion that two different samples with the same maximum mobility (at some high fiduciary density $n \gg n_c$) may very well (but not necessarily) have different values of $n_c$ depending on their distinct values of $(n_i, d)$. Of course, if $d'$ is fixed (e.g., if the impurity locations are similar), then $n_c$ would become a unique function of mobility.

The last point to emphasize in the Introduction is that our phenomenological theory for 2D MIT (i.e., crossover) based on the Ioffe-Regel criterion is manifestly a high-density theory approaching the critical crossover density $n_c$ from above ($n > n_c$) by lowering the carrier density and using a transport theory for the effective metallic phase to calculate the mean free path, and then using the Ioffe-regel criterion $k_Fl = 1$ (to obtain $n_c$) explicitly in the effective metallic phases where $k_F$ and $l$ are well-defined quantities. The reverse theory, i.e., approaching $n_c$ from below ($n < n_c$) in the insulating phase and increasing the carrier density in the effective strongly localized phase to reach $n_c$, would be desirable, but is out of scope for the current work. In principle, such an approach to 2D MIT starting in the effective insulating phase ($n < n_c$) is possible, by using the so-called ‘infrared catastrophe’ theory where one uses the fact that the static long-wavelength carrier dielectric function $\epsilon(q)$ diverges in the long wavelength limit ($q = 0$) in the metallic phase, thus defining $n_c$ through the formula $1/\epsilon(q = 0) = 0$ as $n \rightarrow n_c^-$. We do not, however, believe that this infrared catastrophe approach is practical for calculating $n_c$ (starting from the localized insulating phases) because there is no reliable theoretical method available for calculating $\epsilon(q)$ in the localized phase which would enable the proper $q \rightarrow 0$ limit to be approached for $n \leq n_c$ leading to a practical use of the condition $1/\epsilon(q = 0) = 0$. In our theory we manifestly approach the crossover density from above ($n > n_c$) staying always in the effective metallic phases.

\section*{II. BACKGROUND}

In this section we discuss the issue of the experimentally observed strong metallic temperature dependence $(dp/dT > 0)$ for $n > n_c$ in high-mobility 2D samples.
(or the lack of such strong metallic temperature dependence for \( n > n_c \) in low-mobility samples) in the context of 2D MIT phenomena. We basically address the question why the current interest in 2D MIT phenomena started only around 1995 although 2D semiconductor transport was studied very extensively in the 1970s and 1980s. Although this issue of the temperature dependence of the resistivity in the effective metallic phase \(( n > n_c )\) has been extensively studied theoretically in the literature\textsuperscript{25–32} and is only of indirect significance to the current work where our focus is the dependence of \( n_c \) itself on system parameters (and not the properties of the metallic phase), it is highly relevant for us to discuss this topic to provide a background and a context for why the 2D MIT crossover characterized by the critical density \( n_c \) (separating an effective metallic phase with \( d\rho/dT \geq 0 \) for \( n > n_c \) from an effective insulating phase with \( d\rho/dT < 0 \) for \( n < n_c \)) has become a topic of great current interest (post 1995) in spite of the subject of 2D Anderson localization and metal-insulator transition being an active area of research already during 1970–1980.\textsuperscript{4}

The key empirical difference between the early (i.e., the old) era (< 1995) and the later (i.e., the present) era (> 1995) of 2D MIT physics is that the early work involved highly disordered lower-mobility (and hence higher-\( n_c \)) 2D samples whereas the current era involves higher-quality and higher-mobility (hence lower-\( n_c \)) 2D samples. This is crucial difference between the two eras of 2D MIT physics. We will establish in this section that this key difference (higher versus lower \( n_c \), i.e., lower versus higher sample quality) is sufficient to introduce strong (weak) metallic temperature dependence in the resistivity of the effective metallic phase (for \( n > n_c \)). Of course, the strongly localized phase (for \( n < n_c \)) always manifests strong exponential insulating temperature dependence \(( d\rho/dT < 0 \)) with the resistivity being exponentially activated (or showing exponential variable range hopping transport) in both situations, but the metallic phase manifests weak \( T \)-dependence in the older lower-quality samples in contrast to the present era. We argue in this section that having a large (small) value of \( n_c \) completely ordains the weak (strong) temperature dependence of the metallic phase (for \( n > n_c \)) as observed in the early (late) era of 2D MIT physics. This will put our study of the crossover density as a function of system parameters in the full context of the present day phenomenology of 2D MIT physics where the strong temperature dependence of the resistivity in the apparent metallic phase (for \( n > n_c \)) of high quality 2D samples has played a very important role in the recent (post-1995) excitement and interest in the subject. We argue below that this strong temperature dependence of the metallic resistivity for \( n > n_c \) in the higher-mobility 2D samples arises entirely and directly from having a relatively small value of \( n_c \) in low-disorder 2D samples, which immediately leads to a strong temperature dependence in the low density metallic resistivity.

In Fig. 1 we schematically depict the two distinct generic situations for 2D MIT with respect to the behavior of \( \rho(T, n) \) with Fig. 1(a) and (b) respectively showing the resistivity \( \rho(T) \) for various density \(( n )\) in low-mobility (high-disorder) and high-mobility (low-disorder) situations. The only difference between the two situations is that one [Fig. 1(a)] has a “high” value of \( n_c \) (because of stronger disorder) whereas the other [Fig. 1(b)] has a “low” value of \( n_c \) (because of weaker disorder). Thus, Figs. 1(a) and (b) qualitatively show the respective 2D MIT behaviors in the early (< 1995) and the present (> 1995) days as described in the beginning of section II. In Fig. 1(a) and (b) the temperature dependence of \( \rho(n, T) \) is weak and strong respectively for \( n > n_c \). We mention in this context the seminal importance of the work of Kravchenko and collaborators\textsuperscript{33} who first experimentally established the connection between the sample quality and the strong temperature dependence of the 2D resistivity in the metallic \(( n > n_c )\) phase using low-temperature transport studies in high-mobility \(( > 25,000 \text{ cm}^2/\text{Vs})\) Si-MOSFETs. Indeed, it is the seminal 1994–95 work of Kravchenko and collaborators which created the modern subject of 2D MIT, serving as the temporal milestone separating the early days of 2D MIT [i.e., Fig. 1(a)] from the present days [i.e., Fig. 1(b)] of 2D MIT. Below we establish that the key to the strong metallic temperature dependence of the 2D resistivity (for \( n > n_c \)) is having (low-disorder-induced) low values of the crossover density \( n_c \), which makes \( \rho(T) \) manifest somewhat complementary temperature dependence.
\( (d\rho/dT > 0 \text{ for } n \geq n_c \text{ and } d\rho/dT < 0 \text{ for } n < n_c) \) on two sides of the 2D MIT. We emphasize, however, that at very high (low) density both kinds of samples (low and high disorder in Fig. 1) manifest similarly weak (strong) temperature dependence.

To understand how the strong (weak) metallic temperature dependence (for \( n > n_c \)) correlates with low (high) values of \( n_c \), we introduce three independent density dependent temperature scales \( (T_F, T_{BG}, T_D) \) which characterize the temperature dependence of the resistivity in the metallic phase. These are the electron temperature scale defined by the Fermi temperature \( (T_F) \), the phonon temperature scale defined by the Bloch-Grüneisen temperature \( (T_{BG}) \), and the disorder temperature scale defined by the Dingle temperature:

\[
k_B T_F = E_F = \frac{\hbar^2 k_F^2}{2m} = \frac{\hbar^2}{2m} \left( \frac{4\pi n}{g_s g_v} \right) \propto n, \tag{1}
\]

\[
k_B T_{BG} = 2h k_F v_{ph} = 2h v_{ph} \left( \frac{4\pi n}{g_s g_v} \right)^{1/2} \propto n^{1/2}, \tag{2}
\]

\[
k_B T_D = \Gamma = \frac{\hbar}{2} \left( \frac{e}{m\mu} \right) \propto \mu^{-1}. \tag{3}
\]

Here \( E_F, k_F = (4\pi n/g_s g_v)^{1/2} \), \( m \), \( v_{ph} \), and \( \Gamma \) are respectively the 2D Fermi energy, 2D Fermi wave vector, the carrier effective mass, the phonon velocity, and the impurity-scattering induced level broadening (with \( \mu \) as the sample mobility). To keep our considerations general, we assume a carrier valley degeneracy \( g_v \) and a spin degeneracy \( g_s \) so that the total ground state degeneracy is \( g_s g_v = 2 \) in general except in the presence of a strong applied magnetic field which could spin-polarize the system making \( g_s = 1 \) whereas \( g_v = 1 \) in general except in Si(100)-MOSFETs where \( g_v = 2 \) because of the peculiar Si bulk conduction band structure (and even in Si-MOSFETs the valley degeneracy could sometimes be lifted making \( g_v = 1 \)). The Fermi temperature \( T_F \) defines the intrinsic quantum temperature scale for the 2D electrons, and when \( T_F \) is very large (i.e., \( n \) very high since \( T_F \propto n \)), there cannot be any temperature dependence in the metallic resistivity at low temperatures arising from intrinsic electronic effects since \( T/T_F \ll 1 \). Thus, \( n_c \) needs to be relatively low just in order to keep \( T_F \) low so that \( T/T_F \) is not too small before phonon effects become significant. The Bloch-Grüneisen temperature \( T_{BG} (\propto k_F \propto \sqrt{n}) \) defines the characteristic temperature scale for phonon scattering effects to become important in the 2D metallic resistivity. For \( T < T_{BG} \), phonon effects are strongly suppressed, leading to a weak \( T^p \)-type \((p \approx 5 - 7)\) very high power law in the 2D resistivity arising from phonon scattering whereas for \( T > T_{BG} \), the phonon scattering contribution to the 2D resistivity is linear in \( T \) (which is universally observed in all 2D semiconductor systems in the metallic phase for \( T > 1 - 10K \) depending on the carrier density). Thus, the observation of 2D metallic behavior (for \( n > n_c \)) at low temperatures requires \( T < T_{BG} \) since trivial phonon scattering contribution to the resistivity is not the issue here.

This immediately implies that \( T_F < T_{BG} \) is necessary (certainly sufficient) for the manifestation of the strong metallic temperature dependence in the resistivity since otherwise (i.e., for \( T_F \gg T_{BG} \)) the low-temperature (i.e., \( T < T_F \)) resistivity will be already dominated by the \( \rho \sim T \) behavior arising from phonon scattering effects applicable for \( T > T_{BG} \). Since \( T_F \propto n \) and \( T_{BG} \propto \sqrt{n} \), the condition \( T_F < T_{BG} \) necessitates a low carrier density leading to the conclusion that a large \( n_c \) would lead to the temperature dependence of metallic resistivity \((n > n_c)\) being dominated by phonon scattering effects. Thus, simple dimensional considerations of the characteristic electronic \((T_F)\) and phononic \((T_{BG})\) temperature scales in the problem lead to the inevitable conclusion that any strong metallic temperature dependence arising purely from a quantum electronic mechanism necessitates \( T_F < T_{BG} \) (or at least \( T_F \gg T_{BG} \) is not allowed), and hence necessarily a low \( n_c \) (so that \( T_F \) is not too large even for \( n \geq n_c \)). As an aside we mention that in 3D metals \( T_F \sim 10^4K \) and the phonon temperature scale \( T_{BG} \) is replaced by the Deybe temperature \( \Theta_D \sim 10^5K \), so that \( T_F \gg \Theta_D \) always. This means that the metallic temperature dependence in the resistivity arising purely from an electronic mechanism is simply impossible in 3D metals at low temperatures where phonon effects always dominate down to low temperatures. There can be a weak \( T^2 \) contribution to the resistivity in 3D metals arising from electron-electron scattering through umklapp processes which cannot happen in the 2D semiconductor systems since the umklapp scattering involves very large lattice-scale momentum transfer not of interest in semiconductor transport.

The role of the disorder-dependent Dingle temperature \( T_D \) in the transport problem is rather subtle and is relevant at the lowest temperatures \( T < T_D \) where \( T_D \) acts as a lower cutoff suppressing the temperature dependence for \( T < T_D \). This is because the strong temperature dependence of carrier screening leading to the metallic temperature dependence is cutoff for \( T < T_D \) by impurity disorder effects parametrized by the Dingle temperature \( T_D \). Thus, \( T_D \) explains why the metallic temperature dependence for \( T < T_F < T_{BG} \) arising from quantum electronic processes does not persist all the way to \( T = 0 \) as it would for \( T_D = 0 \) (and of course, if the electronic temperature can be reduced indefinitely which may be an impossibility). Thus, the temperature dependence of \( \rho(T) \) in the metallic phase \((n > n_c)\) is bounded from above \((by T_{BG})\) and from below \((by T_D)\) with the temperature dependence being strong only in the window \( T_D < T < T_F < T_{BG} \). Since \( T_D \propto \Gamma \propto \mu^{-1} \) where \( \mu \) is the characteristic mobility of the system, a large \( \mu \) (i.e., low disorder) is necessary in order to keep \( T_D \) small so that the temperature dependence of \( \rho(T) \) can show up in an appreciable temperature window satisfying \( T_D < T < T_F < T_{BG} \) with \( T_{BG} > T_F \) guaranteeing
that phonon scattering would not play a role in the 2D MIT physics. Note that if \( T_D > T_{BG} \) (i.e., in highly disordered samples) all metallic temperature dependence will be totally suppressed.

The high mobility low-disorder samples of Kravchenko et al. (and of others since then in the modern era of 2D MIT physics) satisfy the constraint \( T_D < T < T_F < T_{BG} \) enabling the observation of the strong metallic temperature dependence since \( n_c \) and \( T_D \) are both low in these high-mobility samples whereas the older Si-MOSFET samples (before the Kravchenko era), where the 2D MIT phenomenon was studied in the early days, \( n_c \) and \( T_D \) had high disorder (and low mobility) and consequently high \( n_c \) (and \( T_D \)) leading to large \( T_F \) (as well as large \( T_D > T_{BG} \)) in the metallic phase \( (n > n_c) \) so that no metallicity could be observed except for phonon scattering effects for \( T > T_{BG} \). Thus, the amount of disorder in the sample leading to low or high \( n_c \) is the key to seeing or not seeing a strong metallic temperature dependence in \( \rho(T) \) for \( n > n_c \).

To give a quantitative description underlying the qualitative picture discussed above, we borrow (without any derivations) from our earlier-obtained\(^{4,6,30,35}\) theoretical results providing \( \rho(T) \) in the 2D effective metallic phase assuming that the resistive scattering arises from screened Coulomb disorder in the system. The quantitative analytical considerations provided below for \( \rho(T, n) \) in the metallic phase serve three purposes: (1) They re-inforce in a concrete manner the qualitative discussion given above establishing how the consideration of the characteristic temperature scales \( T_F, T_{BG}, \) and \( T_D \) (particularly, their density and mobility dependence) immediately leads to the conclusion that \( n_c \) must be small (i.e., low disorder and high mobility) for the 2D MIT phenomenon to be associated with a strongly metallic temperature dependence in \( \rho(T, n) \) for \( n > n_c \); (2) they provide a quantitative understanding of what low (or high) \( n_c \) actually means in a materials-dependent manner, i.e., tell us how large can \( n_c \) be in a specific system (e.g., Si-MOSFETs) and still manifest a strongly temperature-dependent \( \rho(T) \) for \( n > n_c \) without any phonon effects; and (3) they describe how large or small \( n_c \) should be in going from one 2D system to another (e.g., from 2D Si-MOSFETs to 2D GaAs quantum wells) in order for similar metallic temperature dependence to show up in different 2D systems for \( n > n_c \).

The Boltzmann transport theory gives\(^{4,6,30,35}\) the following analytical results for the semiclassical \( \rho_i(T) \) in 2D electron systems at asymptotically low \( (T \ll T_F) \) and high \( (T \gg T_F) \) temperatures, respectively

\[
\rho_i(T \ll T_F) \approx \rho_0 \left[ 1 + \frac{2x}{1 + x} \frac{T}{T_F} + y \left( \frac{T}{T_F} \right)^{3/2} + \ldots \right],
\]

\[
\rho_i(T \gg T_F) \approx \rho_1 \frac{T_F}{T} \left[ 1 - \frac{3\sqrt{\pi}x}{4} \left( \frac{T_F}{T} \right)^{3/2} + \ldots \right],
\]

where \( x = q_{TF}/2k_F \) and \( y = 2.646[x/(1 + x)]^2 \). In Eqs. (4) and (5), \( \rho_0 = \rho(T = 0) \) and \( \rho_1 = (\hbar/e^2)(n_i/n_\pi x^2) \) respectively are the impurity-scattering induced semiclassical resistivities (hence \( \rho_i \)) characterizing the low and the high temperature limits, and \( T_F = E_F/k_B \) is the Fermi temperature (with \( n_i, n_\pi \) being the respective 2D carrier density and impurity density in the system, and \( k_F \) and \( q_{TF} \) are the 2D Fermi wave vector and Thomas-Fermi screening wave vector, respectively).

While Eqs. (4) and (5) provide the metallic contributions to \( \rho(T) \) arising from the temperature dependence of the screened Coulomb disorder, the acoustic phonon scattering by itself contributes also to the temperature dependence\(^{4,6} \) given in the high \( (T \gg T_{BG}) \) and low temperature limits by

\[
\rho_a(T \gg T_{BG}) \approx \rho_0 + A_{ph} \left( \frac{T}{T_{BG}} \right),
\]

\[
\rho_a(T \ll T_{BG}) \approx \rho_0 + B_{ph} \left( \frac{T}{T_{BG}} \right)^5,
\]

where \( T_{BG} = 2\hbar k_F v_{ph} \) is the Bloch–Glüeisen temperature with \( v_{ph} \) as the relevant phonon velocity – the constants \( A_{ph}, B_{ph} \) depend on the elastic properties of the semiconductor\(^{46} \).

We immediately note that strong metallicity necessitates \( T_{BG} > T_F \), which means that we must have \( 2\hbar k_F v_{ph} > \hbar^2 k_F^2/2m \), i.e., \( k_F < 4m v_{ph}/\hbar \). Since \( k_F \propto \sqrt{n} \), the observation of metallicity is a low-density phenomenon restricted to \( n < 8g_c m^2 v_{ph}^2/\pi \hbar^2 \equiv n_p \) wherephonon effects are suppressed.

Second, the screening induced metallic temperature dependence [Eqs. (4) and (5)] can only apply for \( n > n_c \) since, for \( n < n_c \), strong localization induced insulating behavior will dominate (and the metallic theory does not apply). Thus, the metallic behavior is only allowed in an intermediate density window \( n_c < n < n_p \). It follows right away that if the sample is so dirty that \( n_c \gtrsim n_p \), the metallic behavior simply cannot be observed in an experimental sample under any circumstance. Thus, a minimal necessary condition for the manifestation of 2D metallic temperature dependence is that

\[
n_c < n_p = 8g_c m^2 v_{ph}^2/\pi \hbar^2,
\]

where \( n_c \) is the crossover carrier density for the metal-to-insulator transition. Since \( n_c \) obviously increases (decreases) with increasing (decreasing) disorder in the system, a minimal condition for the observation of metallicity is that the system must have low disorder or, equivalently, high mobility, at least satisfying Eq. (8) above. We also note that Eq. (8) implies [using the Si(100)-MOSFET materials parameters] an \( n_i \lesssim 1.2 \times 10^{11} \) cm\(^{-2}\) for Si(100)-MOSFETs consistent with experimental observations in the sense that the modern 2D MIT era started with the Kravchenko-Pudalov seminal 1994-95 2D transport measurements where the critical density is
indeed less than $10^{11}$ cm$^{-2}$ whereas the older MOSFETs manifested an insulating phase (with activated conductivity) at a much higher density of $n_c \gtrsim 10^{12}$ cm$^{-2}$.

To see the role of high mobility in the modern 2D MIT phenomenon of current interest more clearly we consider the specific criterion of the impurity scattering induced collisional broadening energy scale defined by the level broadening parameter $\Gamma = k_BT_D$ (where $T_D$ is the so-called Dingle temperature).

Using the Ioffe-Regel criterion, our condition discussed above, i.e., $n_c \gtrsim n_p$, becomes equivalent to the condition $T_D < T_F < T_{BG}$ for the unambiguous manifestation of the metallic temperature dependence. Using the fact that $\Gamma = \hbar/2\pi$ and $\mu = e\tau/m$, we then get the following necessary condition on mobility for the manifestation of the metallic phase [which is equivalent to the condition defined by Eq. (9)]

$$\mu > \frac{\hbar e}{2k_BT_F}. \quad (9)$$

Using a carrier density $n_c \approx 10^{11}$ cm$^{-2}$, we get for the Si-MOSFETs, $\mu > 21,000$ cm$^2$/Vs. For proportionally lower values of $n_c$, the required minimum mobility would be proportionally higher, again reinforcing the fact that high mobility is a necessary prerequisite for the 2D metallic phase (i.e., $n > n_c$) to manifest a strong temperature dependence in the resistivity. It is reassuring to note that indeed all modern Si-MOS samples showing the canonical 2D MIT behavior after the Kravchenko-Pudalov discovery typically have $\mu > 20,000$ cm$^2$/Vs.

We note that the above constraints on the critical density [Eq. (9)] and the sample mobility ($\mu$) are only the necessary conditions, which may not be sufficient for the actual manifestation of a strong temperature dependent 2D resistivity on the metallic (i.e., $n > n_c$) side. To discuss this issue of sufficient conditions we go back to Eq. (11) and note that for $\rho_0(T)$ to manifest strong temperature dependence, we must have $x > 1$ (at least $x > 1$) so that $\frac{\partial \rho_0}{\partial T} \approx 2\pi/(1 + x)$ is not too small. This requires $x = q_{TF}/2kF > 1$, i.e., $q_{TF} > 2kF$ which translates to

$$n < n_M = 2g_0^2m^2e^4/\kappa^2\hbar^4\pi^2, \quad (10)$$

where $\kappa$ is the background lattice dielectric constant (assuming $g_0 = 2$). Obviously $n > n_c$ has to be satisfied for the 2D system to be in the metallic phase, and so metallicity requires the additional sufficient condition of

$$n_c < n_M \lesssim n, \quad (11)$$

with $n_M = 2g_0^2m^2e^4/\kappa^2\hbar^4\pi^2$. For Si-MOSFETs we get $n_M \approx 1.2 \times 10^{12}$ cm$^{-2}$, which is much larger than $n_c \approx 10^{11}$ cm$^{-2}$ for the post-1995 era 2D MOSFET samples manifesting metallicity in the $T < T_{BG}$ regime of temperatures.

It is gratifying that simple considerations involving $T_F$, $T_{BG}$, $T_D$, and $q_{TF}/2kF$ immediately lead to the prediction that in Si-MOSFETs there would be an $n_c$ low enough ($n_c \lesssim 10^{11}$ cm$^{-2}$) for high-mobility ($\mu \gtrsim 20,000$ cm$^2$/Vs) samples to show strong metallic $\rho(T)$ behavior for $n > n_c$ exactly as observed experimentally in the post-Kravchenko (> 1995) samples whereas in older low-mobility samples with $n_c \approx 10^{12}$ cm$^{-2}$, there would be no metallic $\rho(T)$ behavior (except for phonon effects for $T > T_{BG}$) exactly as seen in older MOSFET systems.

What about other 2D systems such as high-mobility 2D n-GaAs and p-GaAs systems? Below we briefly discuss quantitative the implications of Eqs. (9) – (11) for 2D GaAs systems with respect to the 2D MIT phenomena.

First, 2D n-GaAs has $m = 0.07m_e$, $g_0 = 1$, $\kappa = 13$, and $v_{ph} = 4 \times 10^5$ cm/s in contrast to Si(100)-MOSFETs (considered above in depth) which have $m = 0.19m_e$, $g_0 = 2$, $\kappa = 12$, and $v_{ph} = 9 \times 10^5$ cm/s. Applying Eqs. (9) – (11) to 2D n-GaAs system, we get

$$n_p \approx 1.5 \times 10^{10}$ cm$^{-2}$; $n_c < 10^{10}$ cm$^{-2}$; $\mu \approx 500,000$ cm$^2$/Vs; $n_m \approx 4 \times 10^{10}$ cm$^{-2}$. \quad (12)$$

This indicates that one would have to go to very low carrier density, way below $10^{10}$ cm$^{-2}$, to see any metallicity in 2D n-GaAs system. Since $T_F(K) \approx 4h/n_{\text{GaAs}}$ where $\tilde{n}$ is the carrier density measured in $10^{10}$ cm$^{-2}$, the temperature range ($T < T_F$) for any possible metallic behavior would be well below 1K. In addition, $q_{TF}/2kF \approx 0.4/\sqrt{n}$, which means that $q_{TF}/2kF = 1$ is reached only for $n \approx 1.6 \times 10^5$ cm$^{-2}$, implying that observing strong metallicity (i.e., relatively large $dp/dT$) in 2D n-GaAs would necessitate going to carrier density in the range of $1-2 \times 10^8$ cm$^{-2}$ and $T < 100$ mK, requiring electron mobility of $10^5$ cm$^2$/Vs. Indeed there is only one experimental report\textsuperscript{12} of observing strong metallic behavior in 2D n-GaAs, and it required an ultrahigh mobility of $10^7$ cm$^2$/Vs and a sample of very low carrier density ($\sim 10^9$ cm$^{-2}$).

It is easy to convince oneself using Eqs. (3) – (11) and 2D p-GaAs parameters that for GaAs 2D holes, the metallic behavior should be routinely observable in samples with mobilities of $10^5 - 10^6$ cm$^2$/Vs at carrier densities around $10^{10}$ cm$^{-2}$. This is indeed the experimental situation.

Thus, we have established in this section why older Si-MOSFETs did not see 2D MIT phenomenology: It is simply because the sample quality was too low and consequently the critical density was too high. In concluding this section, it may be worthwhile to obtain some rough comparative quantitative estimates for the metallic temperature dependence (for $n > n_c$) in samples with high and low disorder in order to contrast older and newer Si-MOSFET samples. We provide such a quantitative comparison below for two hypothetical Si-MOSFET samples: A (high disorder) and B (low disorder) with high-density mobilities of $5,000$ cm$^2$/Vs (high $n_c$ for sample A) and $50,000$ cm$^2$/Vs (low $n_c$ for sample B).

Sample A (high disorder) has $n_c = 10^{12}$ cm$^{-2}$, which, using Eq. (3) leads to

$$(\Delta\rho/\rho_0)_A \lesssim 0.2, \quad (13)$$
where $\Delta \rho = \rho(T_{BG}) - \rho(T = T_D)$ is the temperature induced increase in the metallic resistivity for $n \gtrsim n_c$ arising from the screening effect.

Sample B (low disorder) has $n_c = 10^{11} \text{ cm}^{-2}$, which, using Eq. 4, leads to

$$\frac{(\Delta \rho/\rho_0)_{B}}{1.2}. \quad (14)$$

Thus, sample A (B) would manifest a less than 20 % (more than 120 %) increase in the metallic resistivity for $n \gtrsim n_c$ between $T_D < T < T_{BG}$ arising from screening effects, clearly establishing that having low (high) values of the crossover density $n_c$ is the crucial element of physics determining strong (weak) metallic temperature dependence in the system. Since the temperature range for metallicity ($T \lesssim T_{BG} \sim \sqrt{n}$) is much smaller for the lower-disorder sample B, as it has a much lower $T_{BG} \sim 14K$ compared with $T_{BG} \sim 35K$ in sample A, the actual manifested temperature dependence would look much stronger in sample B, where $\rho(T)$ will increase by a factor of 2 in the $T = 0 - 10K$ regime compared with only a $< 10\%$ increase in $\rho(T)$ for sample A in the same temperature ($1-10K$) range. This simple estimate shows why older MOSFET samples, extensively studied in the 1970s and 1980s with mobilities around 5,000 cm$^2$/Vs (or less) never manifested any strong metallic behavior because of their relatively high values of $n_c$ whereas the more recently studied MOSFET samples with mobilities above 20,000 cm$^2$/Vs and $n_c \sim 10^{11}$ cm$^{-2}$ or less always manifest strong metallic temperature dependence in its resistivity. The mystery of the so-called strong 2D metallic behavior is thus connected directly to the relative magnitude of $n_c$ as determined by the 2D sample quality.

We conclude this section by summarizing our finding for the materials dependence of 2D systems manifesting strong 2D MIT behavior (i.e., a strong metallic temperature dependence with large $d\rho/dT > 0$ for $n \gtrsim n_c$). We find that $T_c$ defined by $T_c = T_F(n = n_c)$ must be small enough so that $T_D < T_c < T_{BG}$ for phonon effects to be negligible at low temperatures. We also need disorder to be small enough so that $T_D < T_c$, and therefore $T_D < T_c < T_{BG}$ must be satisfied as the necessary condition for the manifestation of 2D MIT. The sufficient condition is given by $q_{TF}^2/2k_F > 1$ (or at least, not too small) for $n \gtrsim n_c$ so that $d\rho/dT$ is not too small. Using the known expressions for the relevant variables $T_{BG}$, $T_F$, $T_D$, $q_{TF}$, and $k_F$ we conclude that $g_v$, $m$, and $v_{ph}$ should be as large as possible [see Eq. (8)], disorder should be as small as possible so that $\mu$ is large [see Eq. (9)], and $g_v^2m^2/\kappa^2$ should be as large as possible [see Eq. (10)], implying not only large $g_v$ and $m$, but also small $\kappa$. This immediately leads to the conclusion that high-mobility Si(111) 2D systems will manifest the strongest 2D MIT behavior (since $g_v = 6$ here, and $m$ is large) whereas 2D n-GaAs will have the weakest 2D MIT behavior (since $m = 0.07m_e$ is the smallest here with $g_v = 1$), and high-mobility Si(100), 2D p-GaAs, and 2D SiGe systems will have strong 2D MIT behaviors since $m = 0.19$ and $g_v = 2$ [for Si(100)], $m = 0.4$ and $g_v = 1$ (for 2D p-GaAs) are consistent with strong 2D MIT behavior. It is gratifying to know that this material-dependence is exactly what is manifested experimentally with high-mobility Si(111) 2D systems showing the strongest 2D metallic behavior and 2D n-GaAs showing the weakest 2D metallic behavior. Of course, if the sample mobility is low so that the condition $T_D < T_c < T_{BG}$ is violated, then there would be no 2D MIT behavior at all, as happened in all low-mobility 2D systems prior to 1995. We also mention that if the spin degeneracy is lifted (so that $g_v = 1$ instead of 2), for example, by the application of a strong parallel magnetic field, then 2D MIT behavior is suppressed according to the above considerations, and as observed experimentally.

### III. Calculation of $n_c$

In this section, we apply the Ioffe-Regel criterion

$$k_F l = 1, \quad (15)$$

to calculate the crossover 2D MIT ‘critical’ density as a function of system parameters. We first note an immediate problem (or ambiguity) arising from the uncritical direct application of Eq. (15) in obtaining the critical density $n_c$. Interpreting $\nu$ as the transport (or conductivity) mean free path we can write

$$l = v_F \tau_1, \quad (16)$$

where $v_F$ and $\tau_1$ are respectively the Fermi velocity and the transport relaxation time. Using the well-known Drude-Boltzmann formula for the electrical conductivity $\sigma$ given by

$$\sigma = n e^2 \nu_1/m, \quad (17)$$

and $k_F = (4\pi n/g_v g_e)^{1/2}$ for the 2D electron gas and the definition $v_F = p_F/m = \hbar k_F/m$, we get from Eq. (15) the following condition for the critical conductivity $\sigma_c = \sigma(n_c)$

$$\sigma_c = g_v g_e \frac{e^2}{2 \hbar}. \quad (18)$$

Equation (18), which is precisely equivalent to the Ioffe-Regel criterion of Eq. (15) for a 2D electron system, can also be written as

$$\rho_c \equiv \sigma_c^{-1} = \frac{2 \hbar}{g_v g_e e^2}. \quad (19)$$

Thus, in 2D systems the Ioffe-Regel criterion is precisely equivalent to $\rho_c = \hbar/e^2$ if we take the usual situation of $g_v = 2$ and $g_e = 1$ whereas, for Si(100)-MOSFETs with $g_v = 2$, we get $\rho_c = \hbar/2e^2$. It is interesting to note that the straightforward application of Ioffe-Regel criterion leads to a critical metallic resistivity of only
$h/6e^2 \sim 4, 400 \text{ }\Omega$ for the 6-fold valley degenerate Si(111)-MOSFETs which have recently been fabricated using Si-vacuum interfaces.$^{37}$

Such a universal critical resistivity characterizing 2DMIT, with $\rho_c = h/e^2 \approx 25, 600 \text{ }\Omega$ for 2D n-GaAs and p-GaAs and $\rho_c \approx 12, 800 \text{ }\Omega$ for Si(100)-MOSFETs [or 4, 400 $\Omega$ for Si(111)-MOSFETs], is, however, in quantitative disagreement with experimental observations where the critical resistivity for the insulating behavior to manifest itself is certainly not universal in a single material system (i.e., for a given $g_s$ and $g_v$) and typically varies between 5, 000 $\Omega$ and 50, 000 $\Omega$ in various experimental studies although it is often typically within a factor of two of the resistance quantum $h/e^2$, thus indicating that the naive consideration given by the direct application of the Ioffe-Regel criterion defined by Eqs. $^{18}$ and $^{19}$ is certainly reasonably, but not perfectly, accurate. In fact, the pioneering experimental studies of Kravchenko et al.$^{33}$ found $\rho_c \approx 1.5 \text{ }h/e^2$ in low-disorder MOSFETs and the older highly disordered MOSFETs typically manifested $\rho_c \approx h/4e^2$ although both classes of systems presumably involved $g_v = 2$ and $g_s = 2$ with the only difference between the two being the level of disorder and the concomitant value of $n_c$ (being around $\sim 10^{11}$ cm$^{-2}$ and $\sim 10^{12}$ cm$^{-2}$ in two classes of systems, respectively). The most recent experimental investigations of 2DMIT in high quality Si(100) MOSFET based and Si-Ge based 2D systems (both should have $g_s = g_v = 2$) manifest $\rho_c \approx 2h/e^2$ and $2h/3e^2$, respectively, in contrast to the canonical value $h/2e^2$ [Eq. $^{19}$] expected on the basis of $g_s = 2$ and $g_v = 2$. In n-2D GaAs systems, $\rho_c \approx h/2e^2$ has been found$^{22}$ whereas in 2D p-GaAs, the observed $\rho_c$ seems to vary widely with $\rho_c \approx 2h/e^2$ $^{22}$ and $\rho_c \approx h/2e^2$ $^{22}$ both being reported in contrast to the theoretically expected $\rho_c = h/e^2$ for $g_s = 2$, $g_v = 1$ 2D systems.

Thus, we have a conundrum in using the Ioffe-Regel criterion with ‘l’ interpreted as the transport mean free path since this would lead to an inconsistency with the experimentally observed variations in $\rho_c \approx \rho(n_c)$ at the 2D MIT crossover point. We note that this problem of a non-universal experimental $\rho_c$ for 2D MIT within the same material system (i.e., constant $g_s$ and $g_v$) in contrast with the universal theoretical $\rho_c$ prediction from the Ioffe-Regel criterion (with ‘l’ interpreted as the transport mean free path) cannot be resolved by altering the criterion to a different form such as the Mott-Ioffe-Regel criterion$^{39}$ where the transition is defined by $k_F l = \pi$ [rather than $k_F l = 1$ as in the original Ioffe-Regel condition defined by Eq. $^{18}$]. Such a modification will only alter Eq. $^{19}$ for $\rho_c$ to $\rho_c = (2/\pi g_s g_v)(h/e^2)$, a different (and smaller) universal critical resistivity for constant values of $g_s$ and $g_v$ which is, of course, still in disagreement with the experimental observations.

A simple modification of the Ioffe-Regel criterion, where ‘l’ is interpreted as the quantum mean free path given by $l = v_F \tau_q$ where $\tau_q$ is the quantum single-particle scattering time (rather than the transport relaxation time $\tau_l$), actually provides a variable critical resistivity since there is no simple relationship connecting the conductivity $\sigma$ with the quantum scattering time $\tau_q$. Using the identity that impurity scattering induced quantum level broadening $\Gamma$ is related to $\tau_q$ by

$$\Gamma = h/2\tau_q,$$  \hspace{1cm} (20)

it is easy to see that the Ioffe-Regel criterion Eq. $^{15}$, based on using $l = v_F \tau_q$, becomes

$$\Gamma = E_F.$$  \hspace{1cm} (21)

We will use this modified Ioffe-Regel criterion in some of our theoretical analyses since this condition implies a non-universal critical resistivity $\rho_c$ at the 2D MIT crossover even for the same values of $g_s$ and $g_v$.

We note, however, that the calculation of the crossover critical density $n_c$ itself, either using the transport mean free path or the quantum mean free path, would give similar qualitative (but different quantitative) results. Since the precise sample disorder is never quantitatively known (and since we use approximations in treating disorder scattering effects), our goal in this work is a qualitative (and not quantitative) evaluation of the dependence of $n_c$ on various physical variables such as disorder (i.e., $n_i$ and $d$), mobility (at high density), temperature, magnetic field, and materials parameters (e.g., $g_v$, $g_s$). For such qualitative considerations, either form of the Ioffe-Regel criterion may be equally useful although, as discussed above, the corresponding critical resistivity would have the incorrect behavior if ‘l’ is taken as the transport mean free path since it would manifest a universal theoretical value for a given 2D material system (at $T = 0$ at least).

We note that in most systems where the Ioffe-Regel criterion has been applied and discussed so far in the literature (see Graham et al.$^{29}$ and references therein) there is virtually no difference between the transport relaxation time $\tau_l$ and the quantum scattering time $\tau_q$ since the effective disorder potential is essentially short-ranged. In high-mobility modulation-doped 2D systems, however, the charged dopants are placed far from the plane of the 2D layer where the carriers (either electrons or holes in n- or p-modulation doped GaAs or Si-Ge quantum wells and heterostructures) are located, leading to an essentially unscreened very long-range disorder potential in the 2D system. In such modulation-doped high-mobility 2D systems, it is possible for $\tau_l \gg \tau_q$ since most of the disorder scattering would be forward scattering, suppressing $\tau_q$ without affecting $\tau_l$. In such a situation, where forward scattering by remote dopants ($k_F d \gg 1$) dominates transport, it is possible for $\rho_c$ given by Eq. $^{21}$ to be smaller than the $\rho_c$ defined by Eq. $^{19}$. Consequently, the crossover critical density $n_c$ will then be higher as given by Eq. $^{21}$ with $\Gamma = h/2\tau_q$ compared with that given by Eq. $^{19}$ with $l = v_F \tau_q$ defined by the transport mean free path. For Si-MOSFETs, most of the disorder is of short-ranged nature (either charged impurities near the interface or surface roughness scattering), and
therefore, $\tau \approx \tau_q$, so that Eqs. (15) and (21) should give similar (but not identical) estimates for $n_c$ and $\rho_c$ with

\[ n_c \geq n_c \quad (\text{obtained by Eq. (21)}) \]
\[ \rho_c \leq \rho_c \quad (\text{from Eq. (4)}) \] (22)

The inequalities given in Eq. (22) above are general applying to all 2D and 3D systems, and follow simply from the fact that $\tau_q \lesssim \tau_\ell$ always.

**A. Theory**

1. Disorder dependence

Starting with Eq. (15) and writing $l = v_F \tau$ where $\tau$ is an impurity-induced scattering time (either $\tau_l$ or $\tau_q$), we can derive the following scaling relation

\[ n_c \sim n_i^\gamma \] (23)

where

\[ \gamma = (1 + \delta)^{-1} \] (24)

and the exponent ‘$\delta$’ defines the density dependence of $\tau$

\[ \tau \sim n^{\delta}. \] (25)

In deriving Eq. (23), we assume that all parameters, other than $n_i$, are fixed and disorder is entirely parameterized by the 2D impurity density $n_i$. There is an implicit dependence of $n_c$ on the background dielectric constant $\kappa$ and on the impurity location parameter $d$ not explicitly shown in Eqs. (23) – (25). We assume (at this stage) that the $d$-parameter (and the dielectric constant $\kappa$) characterizing the samples is approximately a constant so that the disorder strength can be characterized by the single parameter 2D impurity density $n_i$ in Eq. (23). This should certainly apply approximately to the same class of systems such as Si-MOSFETs fabricated in the same laboratory.

The sample mobility itself is, by definition, inversely proportion to $n_i$

\[ \mu \sim n_i^{-1}, \] (26)

enabling us to eliminate $n_i$ in Eq. (23) in favor of some “maximum mobility” $\mu_m$ defined at a high fiduciary carrier density $n_m \gg n_c$. Eliminating $n_i$ in favor of $\mu_m^{-1}$ we get

\[ n_c \sim \mu_m^{-\gamma}. \] (27)

This gives us a scaling relationship connecting the crossover density $n_c$ to the sample quality as characterized by the typical “maximum mobility” $\mu_m$ defined at some high carrier density $n_m \gg n_c$ deep in the metallic phase.

A detailed theory has recently been developed by us\(^{44}\) for the density scaling of 2D metallic conductivity (and mobility), where we find that the exponent $\delta$ (with $\mu \sim n^\delta$) given by

\[ \delta \approx 0.7 \quad (n - \text{GaAs}); \quad 0.5 \quad (p - \text{GaAs}); \quad 0.3 \quad (n - \text{Si}), \] (28)

restricting to the $n \gg n_c$ situation. This then implies

\[ \gamma = 1/1.7 - 1/1.3 \approx 0.59 - 0.77, \] (29)

with

\[ \gamma \approx 0.59 \quad (n - \text{GaAs}), \quad 0.67 (p - \text{GaAs}); \quad 0.77 (n - \text{Si}). \] (30)

We emphasize that Eqs. (28) – (30) are very approximate and are derived assuming that the 2D impurity density $n_i$ is the only variable determining the disorder, and therefore the quantitative applicability of the numerical values of the exponent $\gamma$ (defined $n_c \sim \mu_m^{-\gamma}$) is very approximate.

It is, therefore, important to emphasize that such a scaling relationship, $n_c \sim \mu_m^{-\gamma}$, with $\gamma \approx 0.67$ approximately (but with some fluctuations in the distribution of $\gamma$ values around $\gamma \sim 0.67$) was noted empirically by Sarachik\(^{42}\) more than ten years ago based on a careful numerical analysis of the existing 2D MIT experimental data. Thus, our Ioffe-Regel criterion based theoretical analysis of the dependence of the crossover critical density $n_c$ on the typical sample mobility $\mu_m$ is completely consistent with the 2D MIT experimental data\(^{42}\).

Similar theoretical considerations can also be applied to the case where $\tau$ is interpreted to be the quantum scattering time $\tau_q$ [rather than the transport scattering time $\tau_l$ as in the analysis of Eqs. (23) – (30) above]. For this situation the Ioffe-Regel criterion is better written as $\Gamma = E_F$ [see Eq. (21)], and it is straightforward to show that we get the following results for $n_c$

\[ n_c = n_i q_{TF}^2 \int_0^1 dx \frac{e^{-4k_F dx}}{\sqrt{1-x^2} (x + q_{TF})^2}, \] (31)

which leads immediately to

\[ n_c = \frac{1}{4\sqrt{2\pi}} \left( \frac{n_i}{d} \right)^{2/3} \text{ for } k_F d \gg 1, \] (32)

and

\[ n_c = \frac{\pi n_i}{2} \text{ for } k_F d \ll 1. \] (33)

Assuming the impurity separation parameter ‘$d$’ to be fixed and the impurity density $n_i$ to be the sole determinant of the system mobility, we can adapt Eqs. (31) – (33) to provide a dependence of the critical density $n_c$ on some fiduciary maximum mobility $\mu_m$ (defined as the sample mobility at some high characteristic density)

\[ n_c \sim \mu_m^{-0.67} \text{ for } k_F d \gg 1, \] (34)

and

\[ n_c \sim \mu_m^{-1} \text{ for } k_F d \ll 1. \] (35)
This immediately leads to the same conclusion we already discussed above [see the discussion above following Eq. (30)] that the dependence of \( n_c \) on the sample quality follows an approximate power law \( n_c \propto \mu_m^\gamma \) where \( \gamma \approx 0.5 - 1.0 \), as has been already pointed by Sarachik based on an empirical analyses of the experimental data.

2. Temperature dependence

The Ioffe-Regel criterion strictly applies at \( T = 0 \), but can be generalized to finite temperatures by considering the temperature dependence of the mean free path \( l(T) \) on the metallic side. This is, of course, particularly relevant for the 2D MIT problem since its hallmark (and the raison d’être for its huge impact in contrast to the corresponding MIT phenomenon in 2D semiconductor systems in the 1970 – 90 era) is the strong temperature dependence of the 2D metallic conductivity for \( n > n_c \). The strong temperature dependence of the 2D metallic conductivity for \( n > n_c \) immediately leads to a strong temperature dependence of the mean free path \( l(T) \), which should strongly affect the critical density \( n_c(T) \) derived on the basis of the Ioffe-Regel criterion \( k_F l = 1 \). Since the 2D metallic conductivity decreases with increasing temperature for \( n > n_c \), the corresponding \( l(T) \) also decreases with increasing temperature whereas the Fermi wave vector \( k_F \propto \sqrt{n} \) is, by definition, temperature-independent. This implies that \( n_c(T) \), defined by the Ioffe-Regel criterion, will increases with increasing temperature at the lowest temperatures, where strong metalliclicity is observed in high-quality 2D systems. The situation is, however, complicated by the fact [see Eqs. (31) and (32)] that the 2D conductivity is non-monotonic as a function of temperature, and eventually decreases with increasing temperature for \( T > T_F \) which is reached at pretty low temperatures if \( T_c = T_F(n_c) \) is low. Thus, \( n_c(T) \) could manifest non-monotonic behavior as a function of temperature for a given sample with \( n_c(T) \) increasing with \( T \) at the lowest temperatures, and then decreasing with \( T \) at higher temperatures. Using Eqs. (31) and (32), and the Ioffe-Regel criterion, we get for \( T \ll T_c \)

\[
n_c(T) \approx n_c \left[ 1 + \left( \frac{x_0}{1 + x_0} \right) \frac{T}{T_c} \right],
\]

and for \( T \gtrsim T_c \)

\[
n_c(T) \approx \frac{T_c}{T},
\]

where \( n_c = n_c(T = 0) \); \( T_c = T_F(n = n_c) \), and \( x_0 = q_{TF}/2k_F c \), where \( k_F c = k_F(n = n_c) \). (We note that the screening wave vector \( q_{TF} \) is a constant independent of carrier density in 2D because of the constant 2D density of states.) From Eqs. (30) and (37), we conclude that \( n_c(T) \) would in principle manifest non-universal behavior, but at sufficiently high temperatures, \( n_c(T) \) will decrease with increasing temperature approximately linearly, i.e., \( n_c(T > T_c) \sim 1/T \). This 1/T decrease in \( n_c(T) \) has been experimentally observed,\(^{21,22,43} \) however, the predicted increase of \( n_c(T) \) with \( T \) at lower temperatures has not yet been reported, perhaps because the lowest temperatures are not reached yet in the experiments to see this decrease or because of complications arising from weak localization corrections not included in our theory.

3. Magnetic field dependence

One of the most important experimental discoveries in the 2D MIT phenomena is the observation of a strong enhancement of \( n_c \) by the application of an applied parallel (to the 2D layer) magnetic field, which is equivalent to the suppression of the 2D effective metallic phase by the applied parallel field. In addition to the enhancement of \( n_c \) compared with its zero-field value, an applied parallel magnetic field also leads to a suppression of the metallic temperature dependence which becomes weaker as the applied field is made stronger. While this latter effect of the magnetic field induced weakening of the metallicity dependence has been extensively studied theoretically,\(^{45} \) there has been no theoretical analysis in the literature of the magnetic field induced enhancement of \( n_c \) itself.

The Ioffe-Regel criterion provides a nature explanation for the field induced enhancement of \( n_c \) as arising from the suppression of the metallic mean free path \( l(T) \) (or the enhancement of the quantum level broadening \( \Gamma \)) due to the enhancement of the effective Coulomb disorder in the metallic phase as the effective carrier screening is reduced by the application of the applied parallel magnetic field \( B \). Screening is suppressed at finite \( B \) since the system gets spin polarized by the \( B \)-induced Zeeman splitting so that the effective Thomas-Fermi screening wave vector \( q_{TF} \) defined by

\[
q_{TF} = \frac{g_s g_n n_c e^2}{\hbar^2},
\]

decreases from being \( g_s = 2 \) for \( B = 0 \) to \( g_s = 1 \) for \( B = B_s \) where \( B_s \) is the density dependent field strength that completely spin-polarizes the 2D system, i.e., \( B_s \) is defined by

\[
2g_\mu_B B_s = E_F,
\]

where \( g \) is the Landé \( g \)-factor for the specific semiconductor, \( \mu_B \) is the Bohr magneton, and \( E_F \) is the Fermi energy at \( B = 0 \). For \( B = B_s \), the 2D system is completely spin-polarized at the Fermi level with \( g_s = 1 \). We note, however, that \( k_F \equiv (4\pi n/g_s g_n)^{1/2} \) itself is now \( B \)-dependent in the presence of spin-polarization, and becomes \( \sqrt{2} \) times larger at \( B = B_s \) compared with its value at \( B = 0 \) since \( g_s \) decreases from 2 to 1. Thus, if the mean free path \( l \) remains unaffected by the applied
field $B$, then the effect of spin-polarization-induced enhancement of $k_F$ itself would lead to a decrease of $n_c$ in finite $B$ in apparent disagreement with experimental observations. Thus, spin-polarization-induced (or equivalently $B$-induced) suppression of screening (and the consequent enhancement of Coulomb disorder) is essential in understanding the enhancement of $n_c$ in the presence of finite applied parallel field. If the effective disorder underlying the 2D MIT phenomenon is purely short-ranged $\delta$-function white-noise disorder, where carrier screening should play no key role, then the application of the parallel magnetic field would decrease $n_c$, effectively enhancing the metallic phase rather than suppressing it as observed experimentally. This latter effect (but not the former), i.e., the effect of the enhancement of $k_F$ by spin-polarization (but not the effect of suppressed screening), is already implicit in Eq. (23) which implies a decreasing $n_c$ with decreasing $g_s$ (i.e., with increasing applied field). We mention that for very large $n_c$, so that $q_{TF}/2k_F$ is very small and screening is unimportant, we do predict that an applied parallel field will either have almost no effect on $n_c$ because spin-polarization effects are negligible due to the very small spin-polarization induced by an applied field at very large density (or will decrease it because of the increasing $k_F$ with increasing spin-polarization). One reason that the applied field effect on 2D MIT was not discovered until 2000 is indeed the fact that $n_c$ was simply too large in the older (and dirtier) 2D samples for the field-induced screening suppression to play any role.

To include the effect of suppressed screening in the presence of finite spin-polarization (i.e., $g_s < 2$), we must take into account the variation of $q_{TF}$ with $g_s$ as shown in Eq. (23). We can obtain an analytical formula by noting that the screened Coulomb disorder $u(q)$ behaves in the following manner

$$u(q) = \frac{\nu(q)}{\epsilon(q)} = \frac{2\pi e^2}{\kappa q + q_{TF}},$$

where $\nu(q) = 2\pi e^2/kq$ is the unscreened 2D Coulomb interaction and $\epsilon(q) = 1 + q_{TF}/q$ is the 2D carrier dielectric screening function. In the strong screening limit ($q_{TF} \gg k_F$), we can write $u(q) \sim q_{TF}^{-1} \sim g_s^{-1}$, and this limit enables an analytical calculation by noting that $l^{-1} \propto n_i u^2 \propto n_i/g_s^2$, which allows us to replace $n_i$ in Eq. (23) by $n_i/g_s^2$, producing the following equation for $n_c(B)$ taking into account dual effects of the enhancement (suppression) of both $k_F$ ($q_{TF}$) by the applied field compared with their $B = 0$ values

$$n_c(B) \sim \left( \frac{n_i}{g_s} \right)^\gamma,$$  \hspace{1cm} (41)

which is valid in the strong screening ($q_{TF} \gg k_F$) limit. In the weak screening limit, the spin-polarization dependence in the screening may be ignored and we get

$$n_c(B) \sim (g_s n_i)^\gamma.$$  \hspace{1cm} (42)

To obtain the explicit $B$-dependence of $n_c(B)$ we need to express the spin-degeneracy factor $g_s$ as a function of the magnetic field $B$, which then leads to

$$n_c(B) \sim \left( 1 + \frac{B}{2B_s} \right)^\gamma,$$  \hspace{1cm} (43)

and

$$n_c(B) \sim \left( 1 - \frac{B}{2B_s} \right)^\gamma,$$  \hspace{1cm} (44)

respectively for Eqs. 41 and 42. Since the 2D MIT phenomenon mostly occurs in the strong-screening ($q_{TF} \gg k_F$) regime, Eq. (13) applies to most situations indicating an increase in the critical density with increasing applied field. At high carrier densities where $2k_F \gg q_{TF}$ or in a situation where short-range scattering dominates transport so that screening is not relevant, Eq. (13) should apply.

Before concluding this section we mention that all our considerations above for the effect of spin polarization $g_s$ (as modulated by the parallel field $B$) applies equally well to the valley degeneracy $g_v$, since the combination $g_sg_v$ appears in all physical quantities. In particular, if $g_v$ could be modified somehow by an external valley-symmetry-breaking field $A$ (for example, an applied strain), then Eqs. (43) and (44) will apply to describe the valley degeneracy dependence of the critical density with the $A$-field replacing the $B$-field. In both cases, the lifting of the spin (or valley) degeneracy by an external field would typically lead to an increasing critical density with increasing field since most 2D MIT phenomena happen in the $q_{TF} \gtrsim k_F$ strong screening regime [as characterized by Eq. (13)]. Indeed, such a symmetrical situation of increasing $n_c$ with increasing external symmetry breaking field for either spin or valley degeneracy has been experimentally observed by Shayegan and his collaborators in the multivalley AlAs 2D systems. Our qualitative findings in this section are in excellent agreement with these experimental results showing an equivalence of increasing $n_c$ with the decrease of $g_s$ or $g_v$. In particular, a given sample with a fixed carrier density ($n$) is most likely to be in the insulating phase [i.e., $n < n_c(g_s, g_v)$] when the 2D system is maximally polarized to have the minimum possible values of $g_s$ and $g_v$, as precisely observed by Shayegan and his collaborators.

4. Materials dependence

To consider how $n_c$ depends on the materials parameters (e.g., $m$, $\kappa$, $g_s$, $g_v$) of the 2D system we imagine a situation with fixed bare disorder while varying only the materials parameters to see how the Ioffe-Regel criterion $k_Fl = 1$ is affected.

Expressing the mean free path $l = v_F\tau$ in terms of the relaxation time $\tau$, and then using the Boltzmann equation to obtain $\tau$ assuming scattering from random
screened charged impurities we get the following integral equation for $n_c$ from the Ioffe-Regel condition (at $T = 0$)

$$\frac{1}{\tau} = \frac{\hbar k_F^2}{m} + \frac{n_c m}{\pi \hbar k_F^2} \left( \frac{2\pi e^2}{\kappa} \right)^2 \int_0^1 dy \frac{y^2 e^{-2qy}}{\sqrt{1-y^2}} \left( y + x \right)^2,$$

where $x = q_{TF}^2/2k_F$ and $q_0 = 2k_Fd$. All other quantities in Eq. 15 are defined as $k_F = (4\pi n/g v_c g_s)^{1/2}$ and $q_{TF} = g_s g_m e^2/\hbar k_F$. We can rewrite the integral equation defined by Eq. 15 as

$$\pi \hbar k_F^4 = n_c m^2 \left( \frac{2\pi e^2}{\kappa} \right)^2 \int_0^1 dy \frac{y^2 e^{-2qy}}{\sqrt{1-y^2}} \left( y + x \right)^2.$$

with the only difference between Eqs. 16 and 17 being the additional factor of $y^2$ inside the integral on the right hand side of Eq. 16. This $y^2$ factor arises from the well-known ‘1 - cosθ’ vertex correction term in the Kubo formula for the current-current correlation function in the conductivity.

In Eqs. 16 and 17, materials parameters $m$, $g_s$, $g_v$, $\kappa$ enter through $k_F = (4\pi n/g v_c g_s)^{1/2}$, $d_0 = 2k_Fd$, and $q_{TF}/k_F \propto n^{-1/2}$. Before discussing the materials dependence of $n_c$ implied by Eq. 16, we note that the above relationship [i.e., Eq. 16] has been derived by assuming $'l'$ to be the transport mean free path, i.e., $\tau = \tau_l$. If, instead of $\tau_l$, we use the quantum relaxation time $\tau = \tau_q$, the only difference is that the vertex correction term disappears from the integral on the right hand side, leading to the following integral equation for $n = n_c$ using the $\Gamma = \Gamma_E$ Ioffe-Regel criterion (i.e. $\tau = \tau_q$ in $k_F = k_F \sqrt{\pi T} \tau = 1$)

$$\pi \hbar k_F^4 = n_c m^2 \left( \frac{2\pi e^2}{\kappa} \right)^2 \int_0^1 dy \frac{y^2 e^{-2qy}}{\sqrt{1-y^2}} \left( y + x \right)^2,$$

2. Comparison with percolation transition

Ioffe-Regel criterion provides one of two possibilities for conduction-to-insulator crossover in 2D semiconductor systems. The other possibility, which has been studied rather extensively in the 2D MIT literature\cite{21,22,44,47-50} is a semiclassical percolation transition at $n = n_c$ arising from the 2D Fermi level moving through the potential fluctuations (“mountains and lakes” landscape) associated with the long-range Coulomb disorder in the 2D system. Simple theoretical considerations\cite{20} and direct numerical simulations\cite{20} indicate that a semiclassical percolation-conductor-to-insulator transition may occur at a critical carrier density given by

$$n_c \approx \frac{1}{4\pi} \sqrt{\frac{n_i}{d}} \approx 0.1 \sqrt{\frac{n_i}{d}},$$

where the long-range-Coulomb disorder is created in the 2D layer by random charged impurities of 2D density $n_i$ located a distance ‘$d$’ from the 2D layer (i.e., exactly the same model for disorder we have used in this work for applying the Ioffe-Regel criterion $k_F l = 1$). Since the percolation transition in the context of 2D MIT has already been extensively studied in the literature, we do not provide any details in the current work on the potential fluctuations driven percolation transition and accept
Eq. (19) for the crossover density as a given. Our goal in the current work is to compare the percolation transition with the Ioffe-Regel transition in the context of the 2D MIT phenomena. It is, however, important here to point out that the percolation transition is manifestly a classical phenomenon with the MIT being driven by the chemical potential or the Fermi level (which is proportional to the carrier density in 2D) crossing through the percolation point in the potential fluctuation driven inhomogeneous 2D “mountains and lakes” landscape with the high-density ($n \gg n_c$) metallic phase being essentially the homogeneous (and well-screened) “all-lakes” situation whereas the low-density ($n \ll n_c$) insulating phase being the highly inhomogeneous (and unscreened) “all-mountains” situation. By contrast, the Ioffe-Regel criterion defines a completely quantum condition for the localization crossover. In some loose sense, the two criterion (percolation and Ioffe-Regel) are complementary and describe the 2D MIT as a high-temperature classical and a low-temperature quantum crossover phenomenon, respectively. We provide some results in the current work comparing and contrasting Ioffe-Regel and percolation results for the critical density $n_c$.

We note that for fixed ‘d’, the percolation criterion implies $n_c \sim \sqrt{\mu_m}$, i.e., the exponent [see, Eqs. (20) and (21)] $\gamma = 1/2$ in the percolation picture whereas $\gamma = 0.6$–1.0 in the Ioffe-Regel theory as discussed already in great details above. If we assume $n_i$ to be fixed and ‘d’ to be the relevant variable characterizing impurity disorder, then the percolation theory gives the simple dependence $n_c \sim d^{-1}$, which can be converted to the following dependence on the high-density mobility $\mu_m$ assuming that $k_Fd \gg 1$ condition applies

$$n_c \sim d^{-1} \sim \mu_m^{-1/3},$$

where we have used the fact that 2D mobility $\mu \sim d^3$ for $k_Fd \gg 1$ (and fixed $n_i$). Thus percolation theory gives the following exponent $\gamma$ (where $n_c \sim \mu_m^{\gamma}$) for the critical density, assuming $d$ to be fixed,

$$\gamma = 1/2, \quad (51)$$

and, assuming $n_i$ to be fixed,

$$\gamma = 1/3. \quad (52)$$

In addition, the percolation critical density, being dependent only on $n_i$ and $d$ (i.e., just the bare disorder), is independent of materials parameters $g_v, g_s, m_v$ and $\kappa$ in contrast to the critical density $n_c$ based on the Ioffe-Regel criterion. We mention the corresponding Ioffe-Regel exponents for Eqs. (41) and (42) are $\gamma = 2/3$ (fixed $d$) and $2/9$ (fixed $n_i$) for $k_Fd \gg 1$.

One particular aspect of percolation induced 2D MIT not discussed above is worth mentioning here (and we will present numerical results on this aspect later in this paper). The critical density defined by percolation theory is completely independent of any transport considerations and thus the constraint on the critical resistivity $\rho_c \lesssim (2/g_v g_s)(h/e^2)$ defined by Eqs. (19) and (22) does not apply to the percolation critical resistivity. In principle, therefore, $\rho_c$ for the 2D MIT percolation crossover could be any value much larger or smaller than the quantum resistance value of $h/e^2$ whereas, by contrast, the Ioffe-Regel condition implies a critical resistance of $O(h/e^2)$. In practice, however, we find numerically (as shown in the next section) that the calculated $\rho_c = \rho(n_c)$ at the 2D MIT percolation transition turns out to be $\sim h/e^2$ for most, if not all, 2D MIT experimental parameters in realistic 2D systems. The possibility, however, remains that $\rho_c$ could be very different from $h/e^2$ in a percolation 2D MIT crossover since the percolation transition is simply a classical transition between immobile and mobile states in an inhomogeneous potential fluctuation landscape where localization or quantum interference plays no special role.

B. Numerical Results

We now present detailed numerical results for our calculated critical crossover density $n_c$ as a function of various physical parameters using the Ioffe-Regel criterion. These results are obtained by directly numerically solving the integral equations defined by Eq. (10) or (14), which correspond respectively to using $l_i = v_F \tau_l$ or $l_q = v_F \tau_q$ in the Ioffe-Regel criterion $k_Fl = 1$. Both equations give similar qualitative results, and our goal in this work is an investigation of the qualitative dependence of $n_c$ on disorder, temperature, applied in-plane magnetic field, and system parameters, and we do not therefore distinguish between these two closely related versions (i.e., $l_i$ or $l_q$) of the Ioffe-Regel criterion. We also provide a comparison between $n_{cIR}$ and $n_{cper}$ as obtained respectively by the Ioffe-Regel criterion and percolation transition in same situations. We believe that a direct comparison between $n_{cIR}$ and $n_{cper}$ as a function of disorder could shed considerable light on the nature of the 2D MIT.

1. Pure 2D case

In Fig. 2 we show our numerically calculated critical density $n_c$ for both the Ioffe-Regel and the percolation theory as a function of $n_i$ (with $d$ fixed) and $d$ (with $n_i$ fixed). All numerical results presented in this subsection assume the 2D carriers to be confined in an ideal strict 2D layer of zero thickness. For the percolation theory, of course, $n_c = 0.141/\sqrt{\tau_d}$ is trivial to plot, and we provide these results only for the sake of comparison with the nontrivial Ioffe-Regel results for $n_c$, which we obtain by numerically solving the integral equation defined by Eq. (14), which uses $\tau = \tau_q$ (i.e., $\Gamma = E_F$ Ioffe-Regel condition). We show results for the three most commonly studied 2D systems: n-GaAs, p-GaAs, and n-Si(100)-MOSFET.

Several general comments can be made about the results shown in Fig. 2. (i) The analytically derived scal-
FIG. 2. The calculated critical density of n-GaAs (a) as a function of impurity density \( n_i \) for fixed impurity locations \( d = 10, 50, 100, 200 \text{ nm} \) (from top to bottom) and (b) as a function of impurity location \( d \) for fixed impurity densities. Solid (dashed) lines represent the percolation critical density, \( n_c^{\text{per}} \) (Ioffe-Regel critical density, \( n_c^{IR} \)). (c) and (d) show the results for p-GaAs with the same impurity parameters of (a) and (b), respectively. (e) and (d) show \( n_c \) for Si-MOSFET. In (e) the impurity locations \( d = 1, 10, 100 \text{ nm} \) (from top to bottom) are used.

FIG. 3. The calculated resistivity \( \rho \) as a function of carrier density. The solid (dashed) lines show results for n-GaAs (Si-MOSFET). The sets of the remote impurity density \( n_i \) and the distance \( d \) are \( (10^{10} \text{ cm}^{-2}, d = 100 \text{ nm}) \) (i.e., \( n_i d^2 = 1 \)); \( (n_i = 10^{10} \text{ cm}^{-2}, d = 10 \text{ nm}) \) (i.e., \( n_i d^2 = 0.01 \)); \( (n_i = 10^{12} \text{ cm}^{-2}, d = 30 \text{ nm}) \) (i.e., \( n_i d^2 = 9 \) [from left to right]). The vertical dot lines indicate the percolation critical density for given impurity conditions (i.e., \( n_c = 0.1/\sqrt{n_i/d} \)).

The percolating behavior derived earlier in this paper apply in their respective regimes of validity, but as already stated before, the dependence of \( n_c \) on both \( n_i, d \) characterizing disorder precludes any definitive dependence of \( n_c \) on the system mobility since \( n_c(n_i, d) \) and \( \mu_m(n_i, d) \) at some high density \( n_m \gg n_c \) are two independent functions of \( n_i \) and \( d \). (ii) In general, \( n_c^{IR} > n_c^{\text{per}} \) for larger values of \( n_i \) and/or \( d \). We see the clear trend in Fig. 2 that as \( n_i \) (d) increases for fixed \( d \) (\( n_i \)) respectively, \( n_c^{IR} \) lines cross above the \( n_c^{\text{per}} \) lines for all three 2D systems we study. For lower disorder (i.e., smaller \( n_i \)), which is of particular interest to 2D MIT phenomena, \( n_c^{IR} \) always is smaller than \( n_c^{\text{per}} \). We expect the percolation theory to be of validity only for rather large values of \( d \) (since only then the Coulomb disorder is effectively unscreened and leads to long-range potential fluctuations in the 2D landscape), and again for \( d' \) not too large, we always find \( n_c^{IR} < n_c^{\text{per}} \). (iii) For similar disorder parameters (i.e., same values of \( n_i \) and \( d \)), our results

in Fig. 2 indicate very similar (but not identical) values of \( n_c^{IR} \) for all three systems we study – of course \( n_c^{IR} = 0.1/\sqrt{n_i/d} \) is, by definition, independent of the materials parameters. This finding of similar \( n_c^{IR} \) in all three systems, while being surprising at first sight, turns out to be consistent with experimental observations where the discrepancy in the reported \( n_c \) values among different systems (with \( n_c^{\text{Si}} \sim 10^{11} \text{ cm}^{-2} > n_c^{p-GaAs} \sim 10^{10} \text{ cm}^{-2} > n_c^{p-GaAs} \sim 10^9 \text{ cm}^{-2} \) appears to arise almost entirely from the very different disorder parameters in these systems (with \( \mu_m^{\text{Si}} \sim 5 \times 10^4 \text{ cm}^2/\text{Vs} < \mu_m^{p-GaAs} \sim 5 \times 10^3 \text{ cm}^2/\text{Vs} < \mu_m^{p-GaAs} \sim 5 \times 10^2 \text{ cm}^2/\text{Vs} \)), where \( \mu_m \) is the typical high-density mobility value, more or less explain the difference in their observed \( n_c \) values based on the approximate scaling law \( n_c \sim \mu_m^{\gamma} \) with \( \gamma \approx 0.67-1.0 \) derived earlier in this paper. (iv) To the extent the numerical results in Fig. 2 allow us to discern any materials trend in the \( n_c^{IR} \) values, we find that for the same disorder strength (i.e., same values of \( n_i \) and \( d \)) Si-MOSFETs tend to have the lowest \( n_c^{IR} \) with n-GaAs and p-GaAs having almost the same calculated \( n_c \) thus verifying the effective mass independence of \( n_c^{IR} \) we derived before. A clear prediction of this finding is that 2D n-GaAs and 2D p-GaAs will have very similar values of \( n_c \) provided they have similar disorder configurations.

To reinforce the point that the Ioffe-Regel criterion typically leads to \( n_c \) values which depend strongly on the disorder, but only weakly on the material, we show in Fig. 3 our calculated resistivity \( \rho(n) \) as a function of 2D carrier density \( n \) for the n-Si-MOSFET and the n-GaAs system for exactly the same set of values of \( (n_i, d) \)
with three different sets of disorder configurations (i.e., $n_i$ and $d$ values) shown in the plots. The $k_Fl = 1$ Ioffe-Regel criterion translates into $\rho = h/e^2$, which gives similar $n_c$ values for the three sets of disorder shown in Fig. 3. For the purpose of comparison we also shown $n_{c_{\text{typ}}}=0.1\sqrt{d}/d$, which again is reasonably close to the calculated $n_{c_{\text{IR}}}^R$ value for each disorder configuration. At first sight, it appears that for the intermediate disorder strength (red curves with $n_id^2 = 0.1$), the $n_{c_{\text{IR}}}^R$ values for Si and $n$-GaAs are very different from each other, but this discrepancy is resolved once the valley degeneracy effect (i.e. $\nu_v = 2$ for Si) is taken into account so that the critical resistivity $= \rho_c h/2e^2$ for the system. It becomes clear that if we use $\rho_{Si}^c = 0.5h/e^2$ and $\rho_{GaAs}^c = h/e^2$, then indeed the resultant $n_c$ values for the two systems are very close to each other, indicating the approximate materials universality of $n_c$ among different 2D systems, with disorder being the primary determinant of $n_c$.

One unexpected aspect of the results shown in Fig. 3 is that the critical resistivity $\rho_c = \rho(n_c)$ for the percolation transition seems to be not very different from that (i.e., $\rho_c \sim h/e^2$) implied by the Ioffe-Regel criterion, which is, of course, a direct manifestation of $n_{c_{\text{typ}}}^R$ and $n_{c_{\text{IR}}}^R$ being not that different (typically $n_{c_{\text{IR}}}^R \gg n_{c_{\text{IR}}}^R$ in Figs. 2 and 3) from each other. We show in Figs. 4(a) - (d) our calculated $\rho_{c_{\text{IR}}}^R$ and $\rho_{c_{\text{typ}}}^R$ defined by

$$\rho_{c_{\text{IR}}}^R = \rho(n = n_{c_{\text{IR}}}^R),$$

$$\rho_{c_{\text{typ}}}^R = \rho(n = n_{c_{\text{typ}}}^R).$$

(53)

Although we expect $\rho_{c_{\text{IR}}}^R \ll h/e^2$, by definition, there is no reason for $\rho_{c_{\text{typ}}}^R$ to have anything to do with $h/e^2$ since it is a nonuniversal quantity not determined by quantum interference or quantum localization. Our results, however, indicate that in general $\rho_{c_{\text{typ}}}^R \ll h/e^2$ as well! We emphasize that this is just a mere coincidence with no deep significance.

In Fig. 4 we show our numerically calculated $\rho_{c_{\text{IR}}}^R$ as a function of $n_i$ and $d$ for 2D $n$-GaAs, $p$-GaAs, and $n$-Si-MOS systems. We emphasize that $\rho_{c_{\text{IR}}}^R = 2h/(\nu_v g_v e^2)$ universally by definition if the quantity $'t'$ in the $k_Fl = 1$ Ioffe-Regel criterion is interpreted as the transport mean free path [see Eq. (11)] of the 2D system. This means that $\rho_{c_{\text{IR}}}^R = h/e^2$ (GaAs); $h/e^2$ (Si) for all $n_i$ and $d$ if we take $'t'$ to be the transport mean free path $l = l_t$ as in Eq. (10). All our numerical $\rho_{c_{\text{IR}}}^R$ results therefore interpret $l = l_t$ as the quantum mean free path [using Eq. (17) without the conductivity vertex correction term] where $k_Fl = 1$ becomes equivalent to $\Gamma = E_F$ strong localization condition. The most important qualitative conclusion based on the numerical results of Fig. 4 is that $\rho_{c_{\text{IR}}}^R$ is large (small) for small (large) $n_i$ and small (large) $d$. In Fig. 4 $\rho_c(n_i, d)$ falls off monotonically either as a function of increasing $n_i$ or increasing $d$, which of course makes sense since small $n_i$ and $d$ implies very small $n_c$, and hence rather large $\rho_{c_{\text{IR}}}^R$ (which is still bounded from above by $h/e^2$ since $\rho_{c_{\text{IR}}}^R \leq h/e^2$ by definition since $l_t \leq l_i$). The decrease of $\rho_{c_{\text{IR}}}^R$ to incredibly small values as a function of increasing $n_i$ or $d$ may appear completely unphysical (perhaps even ridiculous) at first, but this is a direct manifestation of our model of disorder which is entirely characterized by a 2D impurity plane containing $n_i$ random charged impurities per unit area separated by a distance $'d'$. For large $'d'$, this model fails completely since there would always be some unknown and unintentional background charged impurities which will cause the strong localization crossover at some higher value of $\rho_c$ (i.e., lower value of $n_c$). In principle, however, the qualitative result emerging from Fig. 4 is that more disordered the system (i.e., larger the value of $n_c$), lower is the critical resistance $\rho_c$ at the transition (and higher is the $n_c$). This is certainly qualitatively correct since older MOSFETs (before 1994 – 95 when the current era of 2D MIT physics commenced with the Kravchenko et al. work) typically had very high values of $n_c$ ($> 10^{12} \text{ cm}^{-2}$) with consequently rather low values of $\rho_c$ ($\sim h/10e^2 \approx 2 \text{ k}\Omega$). We also mention in this context the empirical finding of Sarachik[6] that $n_c \sim \mu_m^{-0.67}$ which implies very large $n_c$ (and hence rather low $\rho_c$) for samples with very large values of $n_i$. Results with very large $d$-values in Fig. 4 are shown only for the sake of completeness since other unknown disorder with small ‘$d$’ (not included in the model) will intervene making our large $d$ results inapplicable to experimental systems. Our results, however, do indicate that extremely pure modulation doped 2D samples with large values of ‘$d$’ should have relatively small values of $\rho_c$, if all other disorder effects are absent.

In Fig. 4 we show the same results as in Fig. 4 ex-
conjecture now for the percolation theory (i.e. $\rho_{c}^{per}$ is shown as a function of $n_i$ and $d$ in Fig. 5 in contrast to Fig. 4 where $\rho_{c}^{IR}$ is shown). It is clear that $\rho_{c}^{per}$ (Fig. 5) behaves qualitatively very differently than $\rho_{c}^{IR}$ (Fig. 4) with $\rho_{c}^{per} \sim h/e^2$ within a factor of 2 for most values of $n_i$ and $d$. (We emphasize again that $\rho_{c}^{IR} = h/e^2$ within a factor of 2 also if the Ioffe-Regel criterion is taken to be $l = l_T$ in the $k_F l = 1$ condition.) The fact that the percolation transition which defines $n_p^{per} = 0.1 \sqrt{n_i/d}$ with $\rho_{c}^{per} = \rho_{c}(n_i)$ as obtained from our standard Drude-Boltzmann semiclassical transport theory provides a very reasonable value of $\rho_{c} \sim h/e^2$ for a wide range of realistic disorder parameters is certainly somewhat of a surprise. We should mention that for unrealistically large $n_i$ and/or unrealistically small $d$, $\rho_{c}^{per}$ takes on unrealistic values, but for realistic physical combinations of $(n_i, d)$ values operational in real 2D systems our theoretical $\rho_{c}^{per}$ seems to agree well with the experimental results. We do not know at this stage whether this is simply a coincidence or indicates some deep truth about the importance of percolation transport in the 2D MIT phenomena.

Results shown in Figs. 4 and 5 hint at the dimensionless parameter $n_i d^2$ being the important disorder parameter determining $\rho_{c}^{IR}$ and $\rho_{c}^{per}$. This, in fact, follows from the definitions of these two critical resistivities. Using the Boltzmann transport theory for charged impurity scattering limited transport at $T = 0$ we find

$$\rho_{c} = \frac{8h}{e^2 n_c \pi^2} \int_0^1 dy \frac{dy^2 e^{-2y d_c}}{\sqrt{1 - y^2 (y + x_c)^2}} ,$$  \hspace{1cm} (54)$$

where

$$k_c = k_F(n_c) = \left( \frac{4\pi n_c}{g_s g_v} \right)^{1/2} ; x_c = q_{TF}/2k_c ; d_c = 2k_c d .$$  \hspace{1cm} (55)$$

Putting $n_p^{per}$ or $n_p^{IR}$ for $n_c$ in Eq. (54) we obtain $\rho_{c}^{per}$ and $\rho_{c}^{IR}$ respectively. We note that the explicit dependence of $\rho_{c} \sim n_i$ in Eq. (54) is misleading since $n_i$ itself has an $n_i$ dependence also.

The integral on the right hand side of Eq. (54) can be analytically evaluated in various asymptotic limits for both $\rho_{c}^{IR}$ and $\rho_{c}^{per}$, giving the following results (with $k_c \sim$...


\[ n_{\text{crit}} = \frac{1}{4\sqrt{\pi}} d \]

For \( k_c d \gg 1 \)

\[ \rho_c^{IR} \propto (n_i d^2)^{-2/3}; \quad n_c \sim (n_i / d)^{2/3}; \]
\[ \rho_c^{PER} \propto (n_i d^2)^{-1/4}; \quad n_c \propto \sqrt{n_i / d}. \]  

(56)

For \( k_c d \ll 1 \)

\[ \rho_c^{IR} \propto (n_i d^2)^{-3/2}; \quad n_c \sim n_i d^2; \]
\[ \rho_c^{PER} \propto (n_i d^2)^{1/2}; \quad n_c \propto \sqrt{n_i / d}. \]  

(57)

This shows that \( n_i d^2 \) is an important dimensionless parameter determining the disorder scaling of the crossover resistivity \( \rho_c \). We have explicitly checked numerically that these equations [Eqs. (56) and (57)] are in quantitative agreement of our numerical results.

We note that \( \rho_c \) and \( n_c \) have very different qualitative dependence on the disorder parameters \( n_i \) and \( d \), and this might enable an experimental distinction between them possible if quantitative information about the underlying disorder becomes available. In Figs. 7–8 we show our numerically calculated \( \rho_c \) as a function of \( n_i d^2 \) to explicitly depict the dimensionless dependence of \( \rho_c / (h/e^2) \) on the dimensionless disorder parameter \( n_i d^2 \).

In Figs. 6 and 7 we show our calculated \( \rho_c^{IR} \) and \( \rho_c^{PER} \) as a function of \( n_i d^2 \) for various fixed values of \( n_i \) and \( d \) (as shown in the figures) for 2D n-GaAs, p-GaAs, and n-Si-MOSFET systems. It is clear (which is also obvious from our analytical results) that the Ioffe-Regel and the percolation criteria provide very different qualitative dependence of \( \rho_c \) on disorder parameters. Finally, in Fig. 8 we show the calculated resistivity \( \rho \) at different values of \( n \geq n_c^{PER} \) in order to emphasize the scaling behavior.

2. Realistic 2D structures

All results shown in Figs. 2–8 are for strict 2D systems with the appropriate effective mass, lattice dielectric constant, and valley degeneracy (\( q_v = 1, 2 \) for GaAs, Si, respectively) defining each semiconductor material. Results given in Figs. 2–5 serve to provide the qualitative dependence of the critical density and resistivity on disorder parameters, but are not quantitatively realistic even if the disorder parameters (i.e., \( n_i \) and \( d \)) were precisely known. In particular, the finite quantum thickness of the realistic quasi-2D system softens the Coulomb disorder arising from the charged impurities since the 2D Coulomb interaction changes from \( 2\pi e^2/kq \) to \( (2\pi e^2/kq) f(q) \) where \( f(q) \leq 1 \) is the quasi-2D form factor due to the finite quantum thickness effect [and \( f(q) = 1 \) in the ideal 2D limit]. Since the modification to the transport theory for \( f(q) < 1 \) is well-known we do not provide any details, concentrating instead on the numerical results for \( n_c \) in the realistic quasi-2D situation.

In Fig. 6A we show our \( n_i^{IR} \) and \( n_i^{PER} \) results [Fig. 3(a)] for n- and p-GaAs quantum wells (using \( \tau = \tau_r \) so that \( \rho_c^{IR} = h/e^2 \)). For the purpose of comparison, we also provide our results for the strict 2D limit (i.e., zero quantum well thickness \( a = 0 \)) in Fig. 3(b). For \( n_c^{PER} = 0.1\sqrt{n_i / d} \),
the effective value of 'd' changes by a, changing \( n_{c,IR}^{\text{per}} \) to \( n_{c,IR}^{\text{per}} = 0.1 \sqrt{n_i/(d + a/2)} \). For \( n_{c,IR}^{\text{per}} \), the finite thickness increases the effective mean free path \( l \), and thus suppress the resultant \( n_c \). Thus, both \( n_{c,IR}^{\text{per}} \) and \( n_{c,IR}^{\text{per}} \) are suppressed by the finite thickness with this suppression effect being very strong for \( n_{c,IR}^{\text{per}} \) when \( d < a \). A comparison of Figs. 9(a) and (b) bear this out, and thus the finite thickness effect is only quantitative with the qualitative power law dependence of \( n_c \) on \( n_i \) being approximately the same.

In Fig. 10, we show our realistic quasi-2D results for n-Si-MOSFETs where the quasi-2D quantum thickness is determined self-consistently by the carrier density \( n \) itself, which we incorporate through the variational Stern-Howard wavefunction\(^1\). We note that for small values of \( d \) (which is the expected situation in Si-MOSFETs since the charged impurities are typically in the SiO\(_2\) layer close to the Si-SiO\(_2\) interface), \( n_{c,IR}^{\text{per}} \sim 10^{10} - 10^{11} \text{ cm}^{-2} \) for \( n_i \sim 10^{10} - 10^{11} \text{ cm}^{-2} \) whereas \( n_{c,IR}^{\text{per}} > 10^{11} \text{ cm}^{-2} \) for \( n_i > 10^{11} \text{ cm}^{-2} \). These finding are consistent with the higher- and lower-mobility Si-MOSFET devices, respectively.

One important qualitative point to note in Figs. 9 and 10 is that while there is a large difference between percolation and Ioffe-Regel predictions for \( n_c \) for large values of 'd', for small values of \( d \), they are virtually indistinguishable. We also note that the materials difference (e.g., n- versus p-GaAs 2D systems in Fig. 9) is rather small with respect to the calculated \( n_c \) for the same disorder. It may be worthwhile to point out that writing \( n_c \sim n_i \delta \) in Fig. 10 we get \( \delta = \delta(d, n_i) \), and our best numerical estimate for the exponent \( \delta \) is: \( \delta \approx 0.8 - 1 \) for \( d = 1 \text{ nm} \), \( \delta \approx 0.6 - 0.9 \) for \( d = 5 \text{ nm} \), and \( \delta \approx 0.5 - 0.8 \) for \( d = 15 \text{ nm} \). Since \( \mu \sim n_i^{-1} \), we can approximate \( \gamma = \delta \) (where \( n_c \sim \mu_m^{-\gamma} \)), and thus our earlier estimate of \( \gamma \approx 0.67 \) for Si-MOSFET is consistent with \( d = 1 - 2 \text{ nm} \). This is a stringent consistency check on our theory since, indeed, the random charged impurities in Si MOSFETs are known to be located 1 – 2 nm inside the oxide layer.

In Fig. 11, we show our calculated Ioffe-Regel value of \( n_c \) as a function of a fiduciary “maximum mobility” defined as the mobility calculated for exactly the same value of disorder parameters (i.e., the same sample), but at a

FIG. 9. (a) Calculated \( n_{c,IR}^{\text{per}} \) (dashed lines) and \( n_{c,IR}^{\text{per}} \) (solid lines) as a function of impurity density for fixed several \( d = 10, 50, 100, 200 \text{ nm} \) for GaAs quantum wells with a well width \( a = 200 \text{ nm} \). Thick (thin) dashed lines represent results for n-GaAs (p-GaAs). (b) The same results as (a) for zero quantum well thickness (i.e., \( a = 0 \)).

FIG. 10. The calculated critical density of n-MOSFET as a function of impurity density \( n_i \) for fixed impurity locations \( d = 3, 10, 30, 100 \text{ nm} \) (from top to bottom). Here the finite thickness of quasi-2D system is considered. Solid (dashed) lines represent the percolation critical density, \( n_{c,IR}^{\text{per}} \) (Ioffe-Regel critical density, \( n_{c,IR}^{\text{IR}} \)).

FIG. 11. (a) and (b) show the calculated Ioffe-Regel critical density \( n_{c,IR}^{\text{IR}} \) of Si-MOSFET as a function of mobility for different temperatures, \( T = 0, 1, \) and 2K (from bottom to top). The mobility is calculated (a) at a given high density \( n = 5 \times 10^{11} \text{ cm}^{-2} \) and (b) at \( n = 5n_c \). (c) and (d) show the \( n_c \) of n-GaAs as a function of mobility for different temperatures, \( T = 0, 0.2, \) and 0.5K (from bottom to top). The mobility is calculated (c) at a given high density \( n = 5 \times 10^{10} \text{ cm}^{-2} \) and (b) at \( n = 5n_c \).
much higher density $n_m \gg n_c$. The precise dependence of $n_c$ on the high-density mobility $\mu$, of course, depends somewhat on the fiduciary density chosen for the high-density mobility, but the basic finding is that the power law ($\gamma$) dependence, $n_c \sim \mu_m^{-\gamma}$, is a function of temperature, and typically $\gamma \sim 0.7 - 0.8$ as already pointed out empirically by Sarachik a long time ago. The fact that $\gamma \approx 0.6 - 1$ is consistent with experimental findings in different systems is an indication that the experimental 2D MIT may very well be a strong localization crossover phenomenon.

One salient feature of the results presented in Fig. 12 is that the effective exponent $\gamma$, $n_c \sim \mu_m^{-\gamma}$, is the mobility at same high density $n_m \gg n_c$, depends strongly on the temperature (as one would expect because of the strong temperature dependence of the 2D metallic resistivity for $n \gtrsim n_c$ provided $n_c$ is not too large). In Fig. 13 we show our numerically calculated $n_c(T)$, based on the finite-temperature Ioffe-Regel criterion $k_F \ell = 1$, as a function of temperature. As discussed earlier, $n_c(T)$ first increases with $T$ and then decreases when $T \lesssim T_F$. However, the overall variation in $n_c(T)$ is less than a factor of 2 in our results.

In Fig. 13 we show that the calculated maximum mobility dependence of $n_c$ is to some extent dependent on how the maximum mobility is chosen, and thus one cannot really discuss a unique dependence of $n_c$ on the maximum mobility, which is obvious from the fact that both $n_c = n_c(n_i, d)$ and the mobility $\mu = \mu(n_i, d)$ are independent functions of $n_i$ and $d$. What is interesting, however, is the finding that the exponent $\gamma$ (with $n_c \sim \mu_m^{-\gamma}$) remains within our analytical finding of $\gamma \approx 0.6 - 1$ for a wide range of definitions of the maximum mobility $\mu$.

In Fig. 14 we show how our realistic numerical results change if the quantum mean free path with $l = l_q = v_F \tau_q$ is used in the $k_F \ell = 1$ criterion for 2D MIT. There is no qualitative change in the results with $\tau_q$ replacing $\tau_i$ in the Ioffe-Regel criterion as we already emphasized earlier in this paper.

Finally, in Figs 15 and 16 we show the effect of an applied parallel magnetic field $B$ on the critical density $n_c(B)$ due to the spin-polarization-induced lifting of spin degeneracy $g_s$ from $g_s = 2$ at $B = 0$ to $g_s = 1$ at $B = B_p$ where $B_p$ is the applied field strength to fully spin-polarize the 2D electrons. We show numerical results only for Si-MOSFETs here since the qualitative effect of the parallel field on 2D MIT is the same for all 2D systems since the relevant physics is the suppression of screening (and hence suppression of the transport mean

FIG. 12. Calculated Ioffe-Regel $n_c(T)$ of (a) Si-MOSFET as a function of temperature for different impurity densities $n_i = 0.5, 1.0, 1.5 \times 10^{11}$ cm$^{-2}$ (from bottom to top) and (b) n-GaAs for $n_i = 0.2, 0.5, 1.0 \times 10^{10}$ cm$^{-2}$ (from bottom to top).

FIG. 13. Calculated zero temperature Ioffe-Regel critical densities (solid lines) as a function of a reference mobility for (a) Si-MOSFET and (b) for n-GaAs. The reference mobility is calculated for different densities $n = 5n_c, 10n_c, 20n_c$ (from bottom to top). Zero temperature $n_c$ as a function of mobility calculated at (c) $n = 5, 10, 20 \times 10^{11}$ cm$^{-2}$ (from bottom to top) for Si-MOSFET and (d) $n = 5, 10, 20 \times 10^{10}$ cm$^{-2}$ for n-GaAs (from bottom to top). (e) Zero temperature $n_c$ of Si-MOSFET as a function of a reference mobility calculated at $n = 10 \times 10^{11}$ cm$^{-2}$ for different locations of impurity center, $d = 0, 10, 20$ Å (from bottom to top), and (f) $n_c$ of n-GaAs as a function of a mobility at $n = 10 \times 10^{10}$ cm$^{-2}$ for different locations of impurity center, $d = 0, 50, 100$ Å (from bottom to top).
Thus, the condition for the observation of strong temperature dependence of the metallic resistivity and the condition for the observation of strong magnetic field dependence of 2D MIT are closely related as they both require fairly small \( n_c \) (and therefore very high-quality 2D samples) so that \( T/T_F \) and \( B_0/B_s \) can be relatively large in respective cases. This close connection between the temperature dependence and the magnetic field dependence of 2D MIT phenomena is experimentally well-established, and has already been noted in the literature\(^{45}\).

We note that at very low applied field values in Fig. 15 there is a small upturn in the critical density compared with its zero-field value. This is a real effect arising from the increase in effective \( k_F \) induced by the applied field which always suppresses \( n_c \) at finite field compared with its zero-field value, as noted earlier in this paper. If screening effects are unimportant (e.g. scattering by unscreened short-range disorder or at very high carrier density with \( 2k_F \gg q_{TF} \)), then this Fermi surface effect would dominate the finite field transport properties. The 2D MIT phenomenon occurs at low values of \( n_c \), where \( q_{TF} > 2k_F \), and screening effects dominate, leading to a suppression of the metallic phase and an increase in \( n_c \) at finite applied magnetic field.

Finally, in Fig. 15 we show our numerical results on the valley-degeneracy dependence of \( n_c \) by plotting the numerically calculated \( n_c(g_y) \) as a function of the valley degeneracy \( g_y \) (at fixed \( g_x = 2 \)) which we assume for this purpose to be a fictitious continuous variable — in reality \( g_y \) is 1 or 2 for Si(100)-MOSFETs [whereas for Si(111)-MOSFETs, \( g_y = 6 \) is allowed]. As expected \( n_c(g_y) \) behaves very similarly to the spin-polarization effect on \( n_c \), and with decreasing valley degeneracy, \( n_c \) is

---

**FIG. 14.** (a) and (b) show, respectively, the calculated \( n_c \) with the quantum mean free path using the same parameters of Fig. 13(c) and (d) where the transport mean free path is used. Solid (dashed) lines indicate the results with quantum (transport) mean free path. (c) and (d) show the calculated \( n_c \) with the quantum mean free path using the same parameters of Fig. 13(e) and (f) where the transport mean free path is used, respectively. Solid (dashed) lines indicate the results with quantum (transport) mean free path.

Free path \( l \) due to the applied magnetic field. We neglect all orbital effects of the applied magnetic field which could enhance \( n_c(B) \) even more for systems with large quasi-2D thickness.\(^{23}\) As mentioned already, the maximum possible effect of the magnetic field is an enhancement of \( n_c \) by a factor of \( \sqrt{2} \) due to the reduction of spin degeneracy from 2 to 1. Thus, our results in Fig. 14 show an approximate 40% enhancement of \( n_c \) in the presence of the applied field at \( T = 0 \) whereas at finite temperatures the effect is much smaller. We emphasize that although the spin-polarization-induced enhancement of \( n_c(B) \) compared with its \( B = 0 \) value is a universal qualitative phenomenon as long as screened Coulomb disorder is the dominant underlying transport scattering mechanism, the actual quantitative effect would be miniscule (and experimentally unobservable) if \( n_c(B = 0) = n_c \) is very large (as it is highly disordered 2D systems where the 2D MIT phenomena have no dramatic consequences) since \( B_s = 2E_c/g \mu_B \) with \( E_c = E_F(n = n_c) \) would be very large when \( n_c \) is large, and thus \( B_s \gg 1 \) limit would apply on any physically applicable magnetic field in the laboratory making \( n_c(B) \approx n_c(B = 0) \). It is only when \( n_c \) [and hence \( E_c = E_F(n_c) \)] is sufficiently small that the applied parallel field induced enhancement of \( n_c \) can be experimentally relevant since the available laboratory applied field could reach the \( B/B_s \sim 1 \) regime.

**FIG. 15.** (a) Calculated \( n_c(B) \) as a function of parallel magnetic field \( B \) for an impurity density \( n_i = 10^{12} \text{ cm}^{-2} \) and for different temperatures \( T = 0.0, 0.5, 0.1 \text{ K} \). The critical density is normalized by the critical density calculated at \( B = 0 \) and \( T = 0 \), where \( n_c(0, 0) = 7 \times 10^{10} \text{ cm}^{-2} \). Above \( B_s \) the scattering time \( \tau \) is constant in this model, which gives the saturation of the critical density. (b) Calculated \( n_c(B) \) as a function of parallel magnetic field \( B \) for two different impurity densities \( n_i = 5 \times 10^{10} \) (solid lines) and \( 10^{11} \text{ cm}^{-2} \) (dashed lines) and for two temperatures \( |T = 0 \text{K} \) (thick lines) and \( T = 0.3 \text{K} \) (thin lines)]. The critical density is normalized by the critical density calculated at \( B = 0 \) and at given temperature, \( n_c(0, T) \).

Thus the condition for the application of strong magnetic field enhancement of the magnetic field effect would dominate the finite field transport properties. But the 2D MIT phenomenon occurs at low values of \( n_c \), where \( q_{TF} \gg 2k_F \), and screening effects dominate, leading to a suppression of the metallic phase and an increase in \( n_c \) at finite applied magnetic field.
enhanced since screening is reduced. This dependence of \( n_c \) on spin and valley degeneracy of the 2D system is consistent with detailed experimental results reported in 2D AlAs systems\(^{46} \). We emphasize that our calculated \( n_c(g_v) \) for fixed \( g_v \) is identical to results shown in Fig. 16 since both \( g_s \) and \( g_v \) enter the theory equivalently as the product \( g_s g_v \), through the density of states. We mention that for \( T \neq 0 \) (not shown in Fig. 16), \( n_c(g_v) \) softens somewhat showing a weaker dependence of \( n_c \) on \( g_v \). The numerical \( n_c(g_v) \) in Fig. 16 agrees exactly with the analytical dependence \( n^{IR}_c \propto g_v^{-1} \) for \( k_F d \ll 1 \) which is satisfied essentially for all values of \( n_c \) shown in Fig. 16.

**IV. GRAPHENE**

In this section, we consider the strong localization crossover properties of monolayer graphene, which, while being a 2D electronic material of tunable carrier density (with either electrons or holes as the 2D carriers), is fundamentally different\(^{6} \) from the 2D semiconductor systems generally studied in the context of 2D MIT phenomena. It is generally believed theoretically that gapless chiral 2D Dirac fermions do not localize even for arbitrarily strong disorder unless there is intervalley scattering or some other symmetry is broken. Graphene is a gapless, chiral, linearly dispersing 2D system with a massless Dirac-Weyl electron-hole single-particle energy dispersion given by \( E(k) = \hbar v k \) where \( v \) is the graphene velocity \( (v \approx c/300 \) where \( c \) is the velocity of light). The gapless chiral Dirac-fermion nature of graphene leads to fundamentally distinct localization properties\(^{53,54} \) compared with 2D semiconductor systems with 2D graphene manifesting “antilocalization” (i.e., increasing conductivity due to disorder-induced quantum interference) rather than quantum “localization” manifesting in 2D semiconductor systems. Moreover, percolation actually prevents localization from happening in graphene since disorder-induced potential fluctuations lead to electron-hole puddles around the Dirac point in graphene due to its gaplessness – thus conduction can occur at low carrier density through the electron-hole puddles (and the standard percolation driven semiclassical localization picture of “mountains-and lakes” landscape, applying to disordered 2D semiconductor systems which have energy gaps, does not apply to graphene). If an energy gap can be introduced in graphene, as, for example, happens in bilayer graphene in an electric field and in graphene nanoribbons (where latral size confinement introduces energy gaps in the linear spectrum), then percolation transition induced localization indeed becomes possible if the energy gap is larger than the temperature and the typical disorder-induced potential fluctuations. Such percolation localization has been studied in graphene nanoribbons\(^{55} \) and in bilayer graphene\(^{56} \), but our focus in the current work is on regular 2D gapless chiral monolayer graphene, where localization is not allowed unless some symmetry is broken in the system to induce destructive quantum interference effects.

One scattering process enabling localization to occur in graphene is disorder-induced inter-valley scattering, which is, however, very weak since this involves a very large wave vector transfer requiring strong lattice-defect-induced short-range scattering. However, being a relevant perturbation (no matter how weak), at low enough temperatures the presence of any intervalley scattering would induce localization in graphene\(^{56} \). We assume this to be the case (i.e., the presence of a very weak, but nonzero, intervalley carrier scattering), and use the Ioffe-Regel criterion to calculate the critical density \( n_c \) for the strong localization crossover in graphene at \( T = 0 \).

Since in graphene \( \tau_g \gg \tau_i \) (within a factor of 2)\(^{57} \) we simply use the standard Ioffe-Regel criterion \( k_F l = 1 \) with \( l \) as the transport mean free path to consider graphene localization in the presence of some nonzero inter-valley scattering. We also assume that the dominant underlying disorder arises from random charged impurities in the environment as in our 2D semiconductor disorder model. There is considerable experimental evidence in favor of charged impurity disorder being the main disorder mechanism in graphene\(^{58-60} \) at least at low carrier densities where strong localization physics might play a role. We, therefore, use the same 2-parameter disorder model with \( n_i \) and \( d \) to characterize the graphene disorder as we did before for 2D semiconductor systems.

Putting \( k_F \sim \sqrt{n} \) and \( l = v \tau_i \sim \sqrt{n}/n_i \) (as appropriate for \( k_F d \ll 1 \) in graphene for Coulomb disorder\(^{6} \)) we get from the Ioffe-Regel criterion

\[
n_c \sim n_i \sim \mu^{-1},
\]

implying that \( \gamma = 1 \). We note that in graphene screened
Coulomb disorder leads to $\sigma \sim n$, and a constant mobility $\mu$ for a given value of $n_i$ and $d$. We mention here that an applied parallel magnetic field is unlikely to have any observable experimental effect on $n_c$ in graphene because applied magnetic fields necessary to substantially spin-polarize the system are simply too large unless extremely small carrier densities can be explored in ultrapure samples.

It may then appear that it should be straightforward to observe strong localization in graphene simply by having highly disordered samples with very high values of $n_i$ and thus very low values of the mobility $\mu$ since some finite intervalley scattering should always be present as a matter of principle. This may, in fact, be true, but one must go to extremely low temperatures $T \ll \hbar/(k_B \tau_v)$ to see the strong localization effect at $n \leq n_c^{IR}$, where $\tau_v$ is the effective disorder-induced inter-valley scattering time. There is no independent reliable estimate available for intervalley scattering in graphene, but it is expected to be very small. A very rough estimate (actually an upper bound) can be obtained by assuming that very high density graphene transport, where the conductivity $\sigma(n)$ becomes almost a constant, is dominated by short-range scattering which contributes to intervalley scattering since zero-range real space disorder is effectively independent of wave vector transfer in the momentum space (and leads to a constant 2D graphene conductivity). This provides an upper bound for $T_v = \hbar/2k_B \tau_v \approx 1$ mK! Obviously, such a low-temperature strong localization crossover will be rather difficult to observe experimentally unless special efforts are made to enhance intervalley scattering. One experimental technique which could be explored in this context is enhancing lattice defects (or more generally, short-range scattering) in graphene by introducing short-range neutral impurities or vacancies into the graphene sheet artificially in order to induce (or enhance) intervalley scattering, thus increasing $T_v \propto \tau_v^{-1}$ in the process. If intervalley scattering can be enhanced by a factor of 100, thus increasing $T_v$ to $\sim 100$ mK, it could become experimentally feasible to study the strong localization crossover in graphene by tuning the carrier density through $n_c$, which could be made high by having low-mobility samples.

There is, however, a second problem in observing 2D strong localization crossover in graphene (in addition to to the ultralow crossover temperature scale $k_B T_v \sim \hbar/\tau_v$ because of extremely long intervalley scattering time $\tau_v$) which does not apply to 2D semiconductor systems. Graphene is a true semimetal and has no energy gap between conduction and valence bands in contrast to semiconductors, and thus the system goes continuously from being electron-like to hole-like linearly dispersing 2D bands as the external gate voltage tunes the chemical potential through the Dirac point (or more precisely, the charge neutrality point). In reality, however, this gate voltage tuned transition from the electron to the hole band (and vice versa) is not sharp at one precise value of the gate voltage because of disorder-induced potential fluctuations which lead to the formation of inhomogeneous electron-hole puddles in the system over a finite gate voltage or carrier density range $n^*$ around the Dirac point. Once the chemical potential is low enough so that the average density $n < n^*$, the system becomes highly inhomogeneous and in some sense, density fluctuations dominate, becoming larger than the average density (i.e., $n^* > n$) so that further decreasing the chemical potential (i.e., average density) has very little effect on transport (and thermodynamic properties) with the conductivity $\sigma(n < n^*)$ becoming almost a constant around the Dirac point. This phenomenon has been extensively studied theoretically and experimentally, and is well-known. Obviously, for $n^* > n_c$, the strong localization...
crossover will be experimentally unobservable (even for \( T < T_c \)) since the inhomogeneous electron-hole puddle will intervene and make it impossible for localization effect to manifest itself with \( \sigma(n < n^*) \approx \sigma(n^*) = \sigma_{\text{min}} \). Thus, the observation of the strong localization crossover effect necessitates \( n^* < n_c \) (as well as \( T < T_c \)), which may be very difficult, if not impossible to achieve.

In particular, there is good reason\(^5,5^8\) to believe that \( n^* \sim n_i \), and simple estimates indicate that in general \( n^* > n_c \), which implies that strong localization behavior in graphene may never be experimentally observable because of density inhomogeneity and temperature effects. Indeed most graphene experiments find \( n_c \) may be very difficult, if not impossible to achieve.

V. DISCUSSION

We have studied 2D MIT as a strong localization induced crossover phenomenon determined by the Ioffe-Regel criterion (or possibly as a semiclassical percolation transition in the disorder-induced ‘mountains-and-lakes’ inhomogeneous potential fluctuations landscape). There are several distinct aspects of the 2D MIT phenomenology we have addressed in this work theoretically. For the purpose of completeness, we have also studied graphene, concluding that a 2D MIT is unlikely to occur in graphene.

First, we have argued that the crucial factor underlying the strong metallic temperature dependence for \( n > n_c \) in the 2D MIT phenomena is directly connected with the relatively low values of \( n_c \) in high-mobility samples satisfying the condition of \( T_D < T_F < T_{BG} \) where \( T_F \) is the typical Fermi temperature in the metallic phase and \( T_D (T_{BG}) \) is the Dingle (Bloch-Grüneisen) characteristic temperature. Low (high) \( T_D (T_{BG}) \) compared with \( T_F (\times n) \) respectively requires low (low) disorder (carrier) density, and thus 2D MIT can only occur in low-disorder systems. The other necessary condition for the strong metallic temperature dependence for \( n > n_c \) requires \( q_{TF}/2k_F > 1 \) (or at least \( q_{TF}/2k_F \) not too small) in the metallic phase which requires a low density and a high effective mass (indicating that 2D metallicity will be strong in Si and p-GaAs systems, but relatively weak in n-GaAs systems since the low effective mass in n-GaAs, \( m = 0.07m_e \), makes it difficult to satisfy the condition \( q_{TF}/2k_F \); \( n \) and \( q_{TF}/2k_F \) except at very low \( n < 10^9 \) cm\(^{-2}\) in n-GaAs).

The main results obtained in this paper involve theoretical and numerical calculations of the crossover critical density \( n_c^{IR} \) for 2D MIT using the Ioffe-Regel criterion and its comparison with the corresponding percolation transition density \( n_c^{per} \). In 2D semiconductor systems, particularly in high-mobility modulation-doped GaAs structures, the random charged impurities could in principle be rather far from the 2D electrons (or holes) leading to the 2D disorder being mostly long-ranged so that the transport scattering time \( \tau_l \) and the transport mean free path \( l_q = v_F \tau_q \) being much larger than the quantum scattering time \( \tau_l \) and the quantum mean free path \( l_q = v_F \tau_q \). In such a situation, the Ioffe-Regel criterion could be interpreted either as \( k_F l_q = 1 \) or \( k_F l_q = 1 \) leading to \( n_c^l < n_c^q \) since \( l_q > l_q \). We have used both definitions (as well as the percolation definition) for obtaining the critical density.

In addition to obtaining \( n_c^l \), we also provide results for the critical resistivity \( \rho_c = \rho(n = n_c) \), which at \( T = 0 \), is by definition \( \rho_c = 2h/g_F e^2 \). For a percolation transition at \( n = n_c^{per} = 0.1 \sqrt{n_i}/d \), in principle, \( \rho_c^{per} \) could have any value, but in practice \( \rho_c^{per} \sim h/e^2 \) seems to apply extensively for realistic 2D sample parameters.

All our theoretical results use a simple minimal model of Coulomb disorder characterized by a random 2D charged impurity density \( n_i \) and a separation of \( d \) between the impurities and the 2D carriers. More complex models of disorder are straightforward to include in the theory, but will involve more (than two) free parameters, making it difficult to interpret and understand the tra-
theoretical results. Since precise information about the details of disorder is not typically available for high-quality 2D semiconductor systems manifesting 2D MIT, our 2-parameter impurity model is a reasonable starting point for discussing the 2D MIT phenomena.

Below we summarize and critically discuss our important findings and open questions focusing on the key features of our theory as compared with the experimental 2D MIT phenomenology.

A. Localization versus percolation

A question of great importance, of course, is whether the 2D MIT at \( n \approx n_c \) is a strong localization quantum crossover or semiclassical percolation, i.e., whether \( n_c = n_c^{IR} \) or \( n_c = n_c^{per} \). At first sight, it appears that our theory should be able to answer this question with sharp precision since the physical origins and the mathematical descriptions of the Ioffe-Regel quantum crossover and the percolation transition are completely distinct. This turns out to be a much more difficult issue than anticipated at first because \( n_c^{IR} \) and \( n_c^{per} \) have similar magnitudes (both in qualitative agreement with experimental \( n_c \)) in many situations for realistic values of sample parameters. This is a rather surprising finding of our work that could not have been anticipated earlier. Even more surprisingly we find that the critical resistivity \( \rho_c = \rho(n = n_c) \) is similar for both Ioffe-Regel theory and percolation theory! This is a very unexpected and rather strange result since in principle \( \rho_c^{per} = \rho(n = n_c^{per}) \) is allowed to be arbitrary whereas \( \rho_c^{IR} \) is closely related to the resistance quantum \( h/e^2 \approx 25,600 \Omega \) since it arises from quantum localization. But our explicit calculations for 2D n-Si, n-GaAs, and p-GaAs systems show that for realistic experimental parameters (with \( n_c \approx 10^9 \sim 10^{11} \text{cm}^{-2} \)), \( \rho_c \) values calculated from localization and percolation considerations are not widely different although their dependences on system parameters could be quite different. In view of this similarity between absolute values of \( n_c \) (and \( \rho_c \)) in the two theories, it is not easy to manifestly choose one mechanism over the other as determining the 2D MIT crossover density \( n_c \), at least using empirical information and experimental data.

One practical possibility is that as the carrier density is lowered from the high-density metallic phase (\( n \gg n_c \)) to the low-density insulating phase (\( n < n_c \)), whichever transition occurs first (i.e., at higher carrier density) in a given sample dominates the actual crossover behavior in that system. (We mention here that our theoretical calculation of \( n_c \) and \( \rho_c \) explicitly approaches the transition from above, i.e., from the metallic phase.) Thus, \( n_c = n_c^{per} \) if \( n_c^{per} > n_c^{IR} \) and \( n_c = n_c^{IR} \) if \( n_c^{IR} > n_c^{per} \). At very low temperatures, however, quantum interference must always be present and therefore \( n_c \rightarrow n_c^{IR} \) as \( T \rightarrow 0 \). Careful experiments should be carried out to investigate this question of whether \( n_c \) is better described as localization or as percolation. We emphasize that the experimental finding that \( \rho_c \sim h/e^2 \) (typically within a factor of 2 – 3) does not automatically imply that \( n_c \) is described by \( n_c^{IR} \) since our explicit numerical calculations indicate that, perhaps purely coincidentally, \( \rho_c \sim h/e^2 \) (again within a factor of 2 – 3) is also true for \( n_c = n_c^{per} = \rho(n = n_c^{per}) \). We discuss the issue of localization versus percolation more below in the context of comparing theory and experiment.

B. Theory and experiment

An important issue is how our theoretically calculated \( n_c \) compares with the observed experimental dependence of the critical density on various system parameters such as “maximum” mobility, temperature, external magnetic field, valley degeneracy, effective mass, etc. In this respect (i.e., when compared with experimental findings), the Ioffe-Regel criterion describing 2D MIT as a crossover phenomenon seems to be in much better agreement (both qualitative and quantitative) with observations than the percolation theory. In particular, the fact that the 2D MIT behavior (specifically, the value of \( n_c \) itself) depends on an applied parallel magnetic field is difficult to reconcile with the percolation transition which gives a nominally density-independent explicit value of \( n_c = n_c^{per} \approx 0.1\sqrt{n_{BG}}/d \). The Ioffe-regel theory by contrast correctly predicts an increasing \( n_c \) with the applied parallel field (i.e., a field-induced suppression of the metallic phase much discussed in the literature on 2D MIT) arising from the spin-polarization of the 2D system. Similarly, the valley degeneracy dependence of \( n_c \) (and its equivalence to the spin degeneracy dependence), which has been experimentally demonstrated, is very naturally explained in the Ioffe-Regel theory as arising from the variation in screening due to the modification in the density of states, whereas it has no simple explanation within the semiclassical percolation theory.

The most important experimental parameter determining the crossover density \( n_c \) is, of course, the sample quality (or disorder) as discussed throughout this article. We emphasize that although \( n_c \) obviously increases with increasing disorder, and this is the key reason for the 2D MIT phenomena manifesting itself only in the 1990s when sufficiently high-quality 2D systems could be studied with sufficiently low values of \( n_c \) (in order to make the strong temperature dependence of the 2D effective metallic phase apparent by having \( T_D < T_F < T_{BG} \) and \( q_T/q_F > 2k_F \), there does not, in principle, exist a simple relationship between \( n_c \) and the sample mobility \( \mu \) (at a density \( n \gg n_c \)). The reason for this is that the sample disorder is minimally determined by at least two independent parameters (\( n_i \) and \( d \)), and therefore it is, in principle, allowed for \( n_c \) and \( \mu_m \) [with \( \mu_m = \mu(n = n_m) \)] where \( n_m \gg n_c \) is some specific high density] to be completely independent parameters. Thus, in principle, a sample with very high \( \mu_m \) could have much higher \( n_c \) than another sample with low \( \mu_m \) although it is proba-
bly not very likely.

With the above caveat in mind, we can, however, obtain from the Ioffe-Regel (or percolation) theory how \( n_c \) varies with \( n_i \) and \( d \) separately, and we can also calculate how the mobility \( \mu(n) \) varies with \( n_i \) and \( d \) as well as carrier density \( n_c \). Therefore, the disorder dependence of \( n_c \) is completely specified in our theory through the two disorder parameters \( n_i \) and \( d \). Assuming a fixed \( d \), we can convert the \( n_i \)-dependence of \( n_c \) to an effective dependence on the mobility at some high carrier density, finding, \( n_c \sim \mu_m^{-0.7} \) in the Ioffe-Regel theory and \( n_c \sim \mu_m^{-0.5} \) in the percolation theory. It appears that the Ioffe-Regel theory agrees better with experimental data, but the fact that \( n_c \) cannot really be a unique function of a single \( \mu_m \) value makes this agreement between theory and experiment less compelling than it appears at first sight.

Assuming \( k_F d \ll 1 \), we obtain theoretically \( I_c^{IR} \sim n_i d^0 \) and \( n_c^{IR} \sim n_i d^{-1/2} \) whereas for \( k_F d \gg 1 \), \( n_c^{IR} \sim (n_i/d)^{3/7} \) and \( n_c^{IR} \sim (n_i/d)^{1/7} \). In principle, these asymptotic dependence on \( n_i \) and \( d \) can be explicitly checked experimentally, but we know of no detailed experimental study of the critical density on the microscopic parameters defining the disorder.

At this stage, the most convincing agreement between our theory for \( n_c \) and experiment comes from (1) \( n_c^{IR} \sim \mu_m^{-0.7} \) type behavior noted earlier empirically \( ^{42} \), (2) the parallel field induced enhancement of \( n_c^{IR} \), (3) the dependence of \( n_c^{IR} \) on the valley degeneracy and its equivalence to the spin-degeneracy dependence. We note that all three properties mentioned here favor the 2D MIT being a strong localization induced crossover phenomenon as determined by the Ioffe-Regel criterion (in contrast to the semiclassical percolation transition). In this context, we must mention one small (but significant) remaining discrepancy between the Ioffe-Regel theory and the experimental finding on 2D MIT. The Ioffe-Regel theory predicts that the critical resistivity \( \rho_c = \rho(n_c) \) at the transition must necessarily obey the inequality

\[
\rho_c^{IR} \leq \frac{h}{e^2} \frac{2}{g_s g_v},
\]

which means that even if the spin and valley degeneracy are lifted \( \rho_c < h/e^2 \approx 25,6000\Omega \). Experimentally, this inequality is obeyed almost universally with the most important exception being the original Si-MOSFET data of Kravchenko et al. who consistently found \( \rho_c^{SI} \approx 1.5h/e^2 \). We have no way of explaining \( \rho_c > h/e^2 \) (at least at \( T = 0 \)) within the Ioffe-Regel theory. One possibility is that \( \rho_c > h/e^2 \) is a finite temperature effect, and \( \rho_c(T \to 0) \) approaches \( h/e^2 \), but we simply do not know if this is true or not. We emphasize, however, that the vast majority of 2D MIT data are consistent with \( \rho_c \) values obtained from the Ioffe-Regel theory.

C. Transition versus crossover

We have studied the 2D MIT as a crossover in this work (either described by Ioffe-Regel criterion or by percolation), not as a true localization quantum phase transition since two is established to be the lower critical dimensions for the Anderson localization phenomenon, both for noninteracting electrons in a disordered system and in the presence of disorder and interaction. We consider the metallic phase (for \( n > n_c \)) to be an effective metal which at \( T = 0 \) will be insulating in an infinite system. The metal-to-insulator crossover in our theory arises from the strong modification in the effective screened Coulomb disorder which becomes very strong as the carrier density is lowered, leading to the \( k_F d = 1 \) condition defining the MIT crossover point. The fact that our calculated \( n_c \) is in qualitative agreement with experimental observations is persuasive evidence in support of 2D MIT being a crossover phenomenon, but our theory can shed no light on the question of whether quantum criticality in playing a role in this problem or not. In particular, we emphasize that we have no way of ruling out the 2D MIT as a true quantum phase transition since this issue is simply beyond the scope of our work, which treats the problem manifestly as a crossover phenomenon.

D. Interaction versus disorder

Our theory treats disorder (and not interaction) as the key qualitative parameter determining the 2D MIT with interaction effects entering the theory indirectly through the screening of the bare Coulomb disorder arising from the random quenched charged impurities (and as such, interaction is only included in the static RPA screening of the electron-impurity scattering vertex). The theory is purely physically motivated since unscreened Coulomb disorder causes unphysically strong carrier scattering by the impurities. Since we use the Boltzmann transport for calculating the 2D conductivity explicitly, it is not clear how one can incorporate interaction effects (beyond RPA screening) in the theory. The great advantage of the theory is, of course, that all the quantities of experimental interest (e.g., the explicit quantitative dependence of 2D MIT properties on disorder, temperature, magnetic field, spin degeneracy, valley degeneracy, effective mass, background lattice dielectric constant, details of the quasi-2D confinement potential and the associated quasi-2D thickness effect, phonon scattering, etc.) can be quantitatively calculated, which most certainly would be impossible in even the simplest theory including electron-electron interaction going beyond RPA screening. Since our goal in the current work is an explicit calculation of the crossover critical density for 2D MIT, we do not see any way of incorporating electron-electron interaction effects microscopically beyond the reasonable (and physical) approximation of using an RPA-screened
Coulomb disorder for calculating transport properties. We mention that our RPA theory is completely consistent with a weak-coupling perturbation theory for including interaction effects\textsuperscript{12} in the transport properties of the effective metallic phase, but interaction effects on the 2D MIT itself remain beyond the scope of the current work.

In our physically motivated picture for 2D MIT, as emphasized throughout this paper, the critical density \( n_c \) can be arbitrary depending on the disorder strength (in fact, \( n_c \) grows monotonically with increasing impurity density \( n_i \)). Thus, the 2D MIT can occur at any value of \( n_c \), including very large values of \( n_c \) where interaction effects are indeed very weak. For such large \( n_c \) values, as discussed in depth in this paper, the 2D effective metallic phase manifests no strong temperature dependence since the dual conditions \( T_D < T_F \lesssim T_{BG} \) and \( q_{TF} \gtrsim 2k_F \) are simply not satisfied, and therefore, the 2D resistivity for \( n > n_c \) cannot manifest any strong temperature dependence. This is indeed the situation for all 2D systems prior to 1994–95 when the 2D MIT simply happened at large values of \( n_c \). Increasing sample quality eventually led to \( n_c \) values low enough so that \( T_D < T_F \lesssim T_{BG} \) and \( q_{TF} \gtrsim 2k_F \) can both be satisfied simultaneously, so that the 2D effective metallic phase for \( n \gtrsim n_c \) manifested strong temperature dependent resistivity. This lowering of \( n_c \) happened purely because bare disorder went down explicitly, and thus the physics is dominated by the physics of disorder. The reason the 2D MIT is more dramatic for larger values of valley degeneracy \( g_v \) in our theory is due to the fact that \( q_{TF} \propto g_v \), and thus effective screening is stronger for larger \( g_v \). We, in fact, predict that 2D n-Si-MOSFETs \((g_v = 2)\) and 2D p-GaAs \((g_v = 1)\) will have very similar 2D MIT properties since \( q_{TF} \propto g_v m \) are similar in the two systems (since \( m_{p-GaAs} \approx 2m_{n-Si} \)). Similarly, we predict that 2D n-GaAs \((g_v = 1)\) has very weak 2D MIT manifestation since \( m_{n-GaAs} \approx m_{n-Si}/3 \approx m_{p-GaAs}/6 \) and thus the effective screening is extremely weak in 2D n-GaAs compared with p-GaAs and n-Si. We believe that interaction effects (beyond screening) are of quantitative, but not qualitative, importance in 2D MIT which is dominated by the behavior of screened effective Coulomb disorder.

In this context, it may be relevant to point out the experimental findings in Lilly et al.\textsuperscript{19,43} on an ultrapure 2D n-GaAs system where the lowest-ever 2D MIT critical density \( n_c \sim 10^9 \text{ cm}^{-2} \) was found for the highest-ever 2D mobility sample \((\mu \sim 10^7 \text{ cm}^2/\text{Vs})\). This small value of critical density corresponds to a dimensionless 2D interaction strength of \( r_s \sim 13 \) which is substantially larger than the corresponding interaction parameter applying to the 2D MIT observed in any 2D Si MOSFETs (where typically \( r_s \sim 10 \)) with \( n_c \sim 10^{11} \text{ cm}^{-2} \). If 2D MIT was indeed primarily an interaction-related phenomenon, then this n-GaAs system should have manifested much stronger metallicity (i.e. stronger temperature dependence in the resistivity of the effective metallic phase for \( n > n_c \)) than the corresponding Si-MOS systems. In reality, however, the n-GaAs sample in Ref. \textsuperscript{19} showed only a \( 10 – 20 \% \) increase in its metallic resistivity with temperature in contrast to Si MOSFETs which typically manifest a more than \( 100 \% \) increase in the temperature-dependent resistivity in the metallic phase. This indicates that 2D MIT is likely to be a disorder-driven phenomenon rather than an interaction-driven phenomenon.

E. Coulomb disorder versus short-range disorder

We have used (screened) long-range Coulomb disorder arising from random quenched charged impurities (parameterized by a 2D impurity density \( n_i \) and a separation distance \( d \) from the 2D carriers) as the underlying scattering mechanism determining the 2D transport properties (i.e., the scattering or relaxation time \( \tau \)). The main reason for this is, of course, the fact that Coulomb disorder is known to be the primary scattering mechanism in 2D semiconductor systems, certainly at the low carrier density of interest in the 2D MIT phenomena.

A relevant theoretical question is how our results would be modified if we consider short-range disorder (e.g., uncorrelated \( \delta \)-function point scattering centers randomly distributed in the 2D layer). Since the short-range real space disorder corresponds to a constant in the momentum space we have for the purely 2D short-range disorder case

\[
\tau_{SR}(n) \sim \text{constant}; \quad \mu_{SR}(n) \sim \text{constant}; \quad \sigma_{SR}(n) \sim n, \quad (60)
\]

where the subscript ‘SR’ refers to the short-range disorder model. We can now use the Ioffe-Regel criterion \( k_F l_{SR} = 1 \) to obtain \( k_F v_{F} t_{SR} = 1 \) implying (since \( k_F \sim \sqrt{n} \))

\[
n_{c,IR} \sim \tau_{SR}^{-1} \sim \mu_{SR}^{-1}. \quad (61)
\]

Thus, the exponent \( \gamma = 1 \) (where \( n_c \sim \mu^{-\gamma} \)) for the short-range disorder in contrast to \( \gamma \approx 0.7 \) for the long-range Coulomb disorder. Since experimental data clearly support \( \gamma < 1 \)\textsuperscript{42} it is clear that the realistic situation corresponds to the underlying disorder being the long-range Coulomb disorder rather than short-range disorder.

We mention that the fact that the applicable disorder in 2D semiconductor systems is Coulomb disorder rather than short-range disorder (at least at low carrier densities of interest to the 2D MIT phenomenology) is also manifestly obvious from the carrier density dependence of the 2D conductivity \( \sigma(n) \) in the metallic regime \((n > n_c)\) where in general the experimental data in all 2D semiconductor systems follow the density scaling \( \sigma(n) \sim n^\beta \) where \( \beta = 2 \) always applies empirically\textsuperscript{21,22,43} in contrast to \( \beta = 1 \) for short-range disorder. The density scaling behavior of 2D metallic conductivity (for \( n > n_c \)) has been extensively discussed in the recent literature\textsuperscript{42,44} conclusively establishing the Coulombic nature of the underlying disorder in high-mobility 2D semiconductor structures.
One last comment in this respect is that all existing theoretical attempts to include interaction effects in the 2D MIT phenomena invariably utilize the unrealistic short-range disorder model \textsuperscript{22,27,28} since long-range Coulomb disorder is essentially impossible to handle in the presence of electron-electron interaction using standard diagrammatic many-body field theories or renormalization group techniques. Our approach, by contrast, has been to use a realistic screened effective Coulomb disorder in the theory (and not the unrealistic and unphysical short-range white noise disorder) so that the calculated 2D conductivity has reasonable density dependence in the metallic phase instead of manifesting a constant density-independent mobility. Whether a complete theory can be developed in the future combining realistic Coulomb disorder and electron-electron interaction effects (beyond RPA screening which acts to regularize the infrared divergence of Coulomb disorder in an appropriate manner) is not known at this stage and remains an important open question.

VI. CONCLUSION

Assuming 2D MIT to be a crossover phenomenon to a strongly localized insulating state, we have developed a theory for the critical density for the transition from the higher-density effective metallic phase to the lower-density strongly localized insulating phase. The calculated critical density is in qualitative agreement with experimental observations on 2D MIT with respect to its dependence on disorder, applied parallel magnetic field, valley degeneracy, and materials parameters. We have critically examined theoretically the conditions necessary for a strongly temperature dependent resistivity in the effective metallic phase just above the crossover critical density, again obtaining qualitative agreement with existing data on Si-MOSFETs and 2D n- and p-GaAs systems. We have also discussed in depth the possibility of observing 2D MIT in monolayer graphene concluding that this is unlikely to happen except perhaps at extremely low temperatures below the very small intervalley scattering energy scale.

Several open questions remain for future investigations. Although our main conclusion is that the 2D MIT (at least at very low temperatures) is strong-localization induced crossover phenomenon as constrained by an Ioffe-Regel type quantum interference condition, the role of semiclassical percolation in the inhomogeneous potential fluctuation landscape in affecting the crossover behavior remains unclear. For Si-MOSFETs, where the random charged impurities are located at the Si-SiO\textsubscript{2} interface close to the 2D electron system, it is hard to see how and why percolation could be relevant, but in modulation-doped GaAs structures, where the dopants are far away from the electrons, percolation could conceivably be relevant. One possibility we have speculated about is that the transition itself crosses over from being percolation like at higher temperatures to being Ioffe-Regel-like at lower temperatures as quantum tunneling and quantum interference become effective. But there are all mere speculations, and we do not have a theory combining percolation and strong localization crossover, which remains an important open issue in the long-range fluctuating potential landscape of Coulomb crossover. In the current work, we have only compared localization and percolation crossovers, concluding that the experimental observations are more consistent with the Ioffe-Regel localization crossover.

Another open question (and an important shortcoming of our theory) is that the theory developed in this paper approaches the transition (i.e., the crossover) from the higher-density effective metallic side (with decreasing density to approach the transition) using Boltzmann transport theory to treat the screened Coulomb disorder induced carrier scattering. An equivalent theory from the lower-density insulating side (with increasing density to approach the transition) is highly desirable, but is out of scope for our work, and in fact, there is no good idea in the literature about how to approach the transition from the insulating side where the whole concept of a quantum mean free path becomes inapplicable (and therefore the Ioffe-Regel criterion is useless). Such a theory from the insulating side, if available, could be a compelling consistency check for the calculated critical density if the same crossover point is reached theoretically from either direction.

Another issue with our theory, in spite of its good qualitative agreement with essentially all aspects of 2D MIT phenomenology, is that our calculated critical resistivity (i.e., the 2D resistance at the crossover critical density) is only in approximate quantitative agreement with experiments. This may not be a serious problem since the crossover nature of the transition makes it problematic to define a unique zero-temperature critical resistivity (particularly since the 2D resistivity is strongly temperature dependent around the critical density), and it is likely that a proper extrapolation of the experimental data to zero temperature would be in reasonable agreement with our theory since the disagreement is mainly in Si-MOSFETs and is by less than a factor of two.

Our theory provides a good qualitative explanation for the dependence of the critical density on disorder, temperature, applied parallel magnetic field, spin and valley degeneracy, and materials parameters indicating that the Ioffe-Regel criterion, in all likelihood, captures the essential features of the transition between the high-density effective (weakly localized) metallic phase and the low-density strongly localized insulating phase in 2D semiconductors.

ACKNOWLEDGEMENTS

This work is supported by LPS-CMTC and Basic Science Research Program through the National Research
30

46. O. Gunawan, T. Gokmen, K. Vakili, M. Padmanabhan, E. P. De Poortere, and M. Shayegan, Nature Phys. 3, 388 (2007).
47. S. Das Sarma, E. H. Hwang, and Qiuzi Li, Phys. Rev. B 88, 155310 (2013).
48. M. M. Fogler, Phys. Rev. B 69, 121409 (2004); J. A. Nixon and J. H. Davies, Phys. Rev. B 41, 7929 (1990); Y. Meir, Phys. Rev. B 61, 16470 (2000); J. Shi and X. C. Xie, Phys. Rev. Lett. 88, 086401 (2002); J. Shi, S. He, and X. C. Xie, Phys. Rev. B 60, R13950 (1999).
49. F. G. Pikus and A. L. Efros, Zh. Eksp. Teor. Fiz. 96, 985 (1989) [Sov. Phys. JETP 69, 558 (1989)]; A. L. Efros, Solid State Commun. 67, 1019 (1988).
50. A. L. Efros, F. G. Pikus, and V. G. Burnett, Phys. Rev. B 47, 2233 (1993).
51. F. Stern and S. Das Sarma, Phys. Rev. B 30, 840 (1984); W. E. Howard and F. F. Fang, Phys. Rev. B 13, 2519 (1976).
52. S. Das Sarma and E. H. Hwang, Phys. Rev. Lett. 84, 5596 (2000).
53. V. I. Fal’ko, K. Kechedzhi, E. McCann, B. L. Altshuler, H. Suzuura, and T. Ando, Solid State Commun. 143, 33 (2007); E. McCann, K. Kechedzhi, Vladimir I. Fal’ko, H. Suzuura, T. Ando, and B. L. Altshuler, Phys. Rev. Lett. 97, 146805 (2006); J. H. Bardarson, J. Tworzydlo, P. W. Brouwer, and C. W. Beenakker, Phys. Rev. Lett. 99, 106801 (2007).
54. I. Aleiner and K. Efetov, Phys. Rev. Lett. 97, 236801 (2006).
55. S. Adam, S. Cho, M. S. Fuhrer, and S. Das Sarma, Phys. Rev. Lett. 101, 046404 (2008).
56. E. Rossi and S. Das Sarma, Phys. Rev. Lett. 107, 155502 (2011).
57. E. Hwang and S. Das Sarma, Phys. Rev. B 77, 195412 (2008).
58. S. Adam, E. H. Hwang, V. M. Galitski, and S. Das Sarma, Proc. Natl. Acad. Sci. U.S.A. 104, 18392 (2007).
59. Y.-W. Tan, Y. Zhang, K. Bolotin, Y. Zhao, S. Adam, E. H. Hwang, S. Das Sarma, H. L. Stormer, and P. Kim, Phys. Rev. Lett. 99, 246803 (2007).
60. J. H. Chen, C. Jang, S. Adam, M. S. Fuhrer, E. D. Williams, and M. Ishigami, Nature Phys. 4, 377 (2008).
61. E. H. Hwang, S. Adam, and S. Das Sarma, Phys. Rev. Lett. 98, 186806 (2007).
62. E. H. Hwang and S. Das Sarma, Phys. Rev. B 80, 075417 (2009).
63. L. A. Ponomarenko, A. K. Geim, A. A. Zhukov, R. Jalil, S. V. Morozov, K. S. Novoselov, I. V. Grigorieva, E. H. Hill, V. V. Cheianov, V. I. Falko, K. Watanabe, T. Taniguchi, and R. V. Gorbachev, Nat. Phys. 7, 958 (2011).
64. S. Das Sarma, E. H. Hwang, and Qiuzi Li, Phys. Rev. B 85, 195451 (2012).
65. K. Kechedzhi, E. H. Hwang, and S. Das Sarma, Phys. Rev. B 86, 165442 (2012).
66. F. Amet, J. R. Williams, K. Watanabe, T. Taniguchi, and D. Goldhaber-Gordon, Phys. Rev. Lett. 110, 216601 (2013).
67. S. Das Sarma and E. H. Hwang, Phys. Rev. B 88, 035439 (2013).
68. D. Belitz and T. R. Kirkpatrick, Rev. Mod. Phys. 66, 261 (1994); D. Belitz, T. R. Kirkpatrick, and T. Vojta, Rev. Mod. Phys. 77, 579 (2005).
69. F. J. Wegner, Z. Phys. B 35, 207 (1979).
70. A. M. Finkel'stein, Zh. Eksp. Teor. Fiz. 86, 367 (1984) [Sov. Phys. JETP 59, 212 (1984)]; Pis’ma Zh. Eksp. Teor. Fiz. 40, 63 (1984) [JETP Lett. 40, 796 (1984)].
71. A. Punnoose and A. M. Finkel'stein, Science 310, 289 (2005).