Long-range Coulomb interaction in arrays of self-assembled quantum dots

A. I. Yakimov, A. V. Dvurechenskii, V. V. Kirienko, Yu. I. Yakovlev, and A. I. Nikiforov
Institute of Semiconductor Physics, Lavrent'eva 13, 630090 Novosibirsk, Russia

C. J. Adkins
Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, United Kingdom
(Received 16 July 1999; revised manuscript received 22 September 1999)

An array of 3×10^7 Ge self-assembled quantum dots is embedded into the active channel of a Si metal-oxide-field-effect transistor. Conductance oscillations with a gate voltage resulting from a successive loading of holes into the dots are observed. Based on measurements of the temperature dependence of the conductance maxima, the charge-transfer mechanism in the channel is identified as being due to variable-range hopping between the dots, with the typical hopping energy determined by interdot Coulomb interaction. The characteristic spatial dimension of the hole wave functions as well as the charging energies of the dots are determined from the conductance data. The effect of the proximity of a bulk conductor on hopping transport is studied. We find that putting a metal plane close to the dot layer causes a crossover from Efros-Shklovskii variable-range hopping conductance to two-dimensional Mott behavior as the temperature is reduced. At the crossover temperature the hopping activation energy is observed to fall off. The metal plane is shown not to affect the conductance of samples which show Mott hopping. In the Efros-Shklovskii hopping regime, the conductance prefactor was found to be $\simeq e^2/h$, and the conductance to scale with the temperature. In the fully screened limit, the universal behavior of the prefactor is destroyed, and it begins to depend on the localization length. The experimental results are explained by a screening of long-range Coulomb potentials, and provide evidence for strong electron-electron interaction between dots in the absence of screening.

I. INTRODUCTION

In a single quantum dot (QD) weakly coupled by tunneling barriers to two leads, the interplay of single-electron charging effects and resonant tunneling through quantized states leads to conductance oscillations as the electrochemical potential of the dot is tuned.[1] This phenomenon underlies the working of nanoscale single-electron transistors which have a number of practical uses, ranging from metrology to computing. Recently, research focused on double-dot systems[2] whose behavior is found to be mainly affected by electrostatic coupling between the two dots inside the artificial molecule. The next step is to create and study large arrays of QD's in close proximity, allowing Coulomb interaction and tunneling between them.[3] Such systems can be considered as potential electronic networks for quantum computers,[4] and therefore are particularly valuable in future high-power digital processors. The behavior of a multidot structure is expected to be more complicated for several reasons: (i) the QD's are inevitably not sufficiently identical in size, which can cause smearing of their atomiclike properties; (ii) in contrast to a single dot, the interaction of the dots in an ensemble can be significant; and (iii) transport through the system may be dominated by thermally assisted hopping between the dots rather than by resonant tunneling between source and drain electrodes.

Variable-range hopping (VRH) is a general conduction mechanism in systems with strongly localized carriers at sufficiently low temperatures. In a regime of VRH, the hopping distance increases as temperature is lowered, and the temperature dependence of conductivity is given by

$$G(T)=G_0\exp[-(T_0/T)^x], \tag{1}$$

where, in the two-dimensional case in the absence of long-range Coulomb interaction, the exponent $x=1/3$ and $T_0 \equiv T_M \propto [g(E_F)\xi^3]^{-1}$ (Mott VRH), $g(E_F)$ being the density of states in the vicinity of the Fermi level E_F and ξ the localization radius. If the interaction energy of a displaced electron and the hole it leaves behind are large compared with disorder energies, the conductivity is decribed by the Efros-Shklovskii (ES) law with $x=1/2$ and $T_0 \equiv T_{\mathrm{ES}} \propto e^2/(\epsilon_r \xi)$, where ϵ_r is the relative permittivity, e is the electron charge.

In a previous paper,[5] we reported measurements of impurity hopping transport in a modulation-doped Si field-effect structure with a layer of Ge QD's embedded in proximity with the *p*-type conductive channel. At low temperatures ($T\simeq 6$ K), a small ($\sim 10-20\,\%$) contribution to conductance which intriguingly oscillated with gate voltage was detected, and was ascribed to direct hopping of holes between dots.

In this work we describe a set of experiments in which we studied hopping transport in Si metal-oxide-semiconductor field-effect structures containing a two-dimensional array of Ge self-assembled quantum dots as a conductive channel. The dots are separated from each other by weakly doped silicon, and the only conduction mechanism at low temperatures is the tunneling of holes between them. Pseudomorphic Ge islands grown epitaxially on a Si(001) surface exhibit a large band discontinuity in the valence band, and can be viewed as doping ''artificial atoms.'' They provide a system in which the number of confined holes, the structure of the energy levels, the shape of the wave functions, and the strength of the interaction can be controlled. First, we discuss the field effect and the temperature dependence of conductance of the samples with dots whose occupation with holes

was changed by varying the potential of a gate electrode. Second, we discuss the effect of screening in samples in which the holes on the dots are supplied by a boron δ-doping layer near to the QD layer.

II. EXPERIMENTAL DETAILS

Samples were fabricated on a silicon-on-insulator wafer (p-type Si substrate, 400-nm buried SiO_2 and 170-nm top Si) or on a semi-insulating p-Si substrate with a resistivity of 1000 Ω cm by molecular-beam epitaxy in the Stranskii-Krastanov growth mode. This produces spontaneous formation (self-assembling) of Ge nanoclusters (quantum dots) randomly distributed in the plane.[5–7] The growth temperatures were 500 and 700 °C for the cap and buffer (10–50 nm thick) Si layers, respectively, but, during the growth of Ge, the temperature was lowered to 300 °C. The growth rates were 2 ML s^{-1} for Si and 0.2 ML s^{-1} for Ge. The nominal thickness of Ge deposited was 10 ML (1 ML = 1.4 Å). From scanning tunneling and transmission electron micrographs of similarly grown samples,[6,7] we observe the Ge dots to be approximately 15 nm in diameter and 1.5 nm in height. Their dimensions vary within a 20% range. The areal density of the dots is 3×10^{11} cm^{-2}.

For the field-effect measurements (Sec. III A), the channel was patterned by photolitography to form a Si island of 100-μm width and 108-μm length, etched down to the underlying SiO_2 [Fig. 1(a)]. The thickness of the Si cap layer (d_{Si}) in this case was 40 nm. Source and drain electrodes were made using Al evaporation and annealing at 450 °C in a N_2 atmosphere. A plasma-enhanced chemical-vapor deposition oxide of 60-nm thickness was deposited as the gate insulator and, finally, a square-shaped (100×100 μm^2) Al gate was deposited. (The distance between the gate and the dot layer in this set of samples, 100 nm, was large enough to avoid screening effects in the temperature range investigated.) The active channel of this type of samples contains about 3×10^7 Ge dots. The drain current (I_d) as a function of the gate voltage (V_g) was measured in the temperature range from 300 to 4.2 K, with the drain voltage fixed at $V_d = 10$ mV, which ensured ohmic conduction at all experimental temperatures.

In the samples used for the experiments with screening (Sec. III B), the holes on the dots are supplied by a boron δ-doping layer near to the QD layer [Fig. 1(b)]. The number of holes per dot was varied from $N = 1/2$ to 13/2 by varying the doping. The silicon cap layer has a thickness of $d_{Si} = 10$ nm. Au source and drain electrodes were deposited on top of the structure and heated at 400 °C to form reproducible Ohmic contacts. A thin ($d_{SiO_2} = 25$ nm) layer of anodic SiO_2 was grown to separate the conductive channel (dot layer) from a Au screening electrode (100 nm thick) which was deposited over the oxide parallel to the dot layer. The screening layer was only deposited between contact 1 and 2 [Fig. 1(b)], and this area (7×5 mm^2) provides a screened sample with a screening length $d = d_{Si} + d_{SiO_2} = 35$ nm. A corresponding unscreened sample is provided by the area between contacts 2 and 3 which has no screening layer. Special care was taken to check the leakage of the insulator. We made sure that throughout the range $T = 4.2$–300 K the resistance between the screening plane and the source or drain is much larger than the channel resistance. Samples prepared in a similar way, but containing no dots, were not conductive at low temperatures.

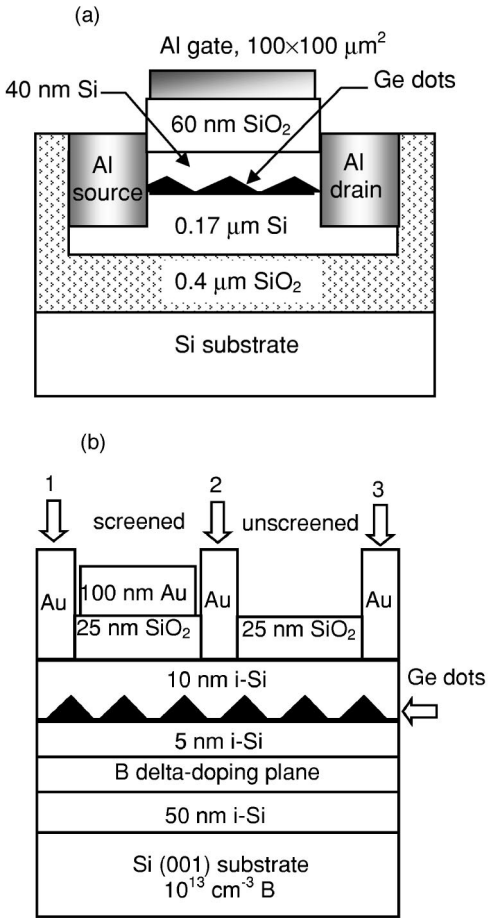

FIG. 1. Schematic diagram of the structure of the samples used for the field-effect experiments (a) and for the experiments with screening (b). Screened samples are defined by contacts 1 and 2, whereas uscreened samples are measured between contacts 2 and 3. Each sample is a strip 7 mm long and 5 mm wide.

III. EXPERIMENTAL RESULTS

A. Field effect in array of charge-tunable dots

The channel conductance ($G \equiv I_d / V_d$) of a sample shown in Fig. 1(a) versus the gate voltage in linear and semilogarithmic plots is depicted in Fig. 2. At room temperature, the G-V_g characteristic shows a shoulder which evolves into a broad conductance peak in the voltage range from 2 to 6 V as the temperature decreases. In order to analyze the peak's structure, we subtract the smoothly varying background (see below). After the background is subtracted, the conductance modulation can be very well described by a sum of four Gaussian peaks. Figure 3 demonstrates the result of decomposition of the several experimental curves into four Gaussians, labeled as QD6, QD5, QD4, and QD3.[8] The position of those peaks as a function of temperature is displayed in Fig. 4. Clearly, the fourfold structure with a gate voltage separation $\Delta V_g \approx 0.7$ V between the peaks is well-defined and completely reproducible over the whole temperature range.[9] A

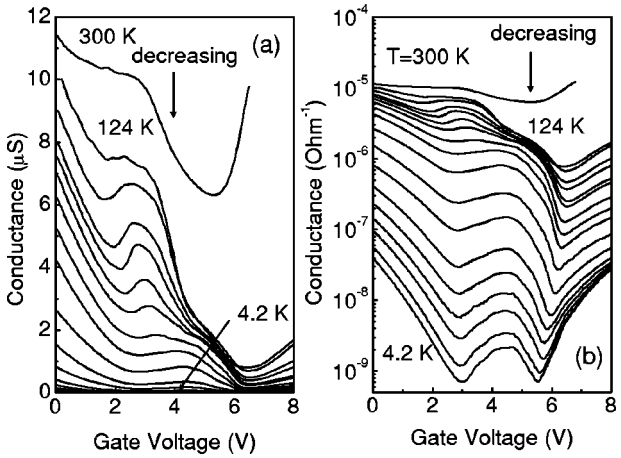

FIG. 2. Channel conductance vs gate voltage characteristics in linear (a) and semilogarithmic (b) plots. The source-drain voltage is 10 mV. The temperature decreases from top to bottom.

very similar fine-structure consistent with four maxima has been observed by us in the admittance spectra of Si-based Schottky diodes with Ge quantum dots,[10] and ascribed to the individual charging of the fourfold-degenerate first excited state in the Ge nanoclusters. The average energy splitting due to hole correlation or charging energy was found to be $\Delta E = 29 \pm 16$ meV.[10]

To demonstrate that the observed conductance peaks come from a charging of the quantum dots, we first estimate the charge density induced by the change in the gate voltage ΔV_g and compare this quantity with the density of QD's. A change of gate potential ΔV_g induces a change Δn in the two-dimensional carrier density given by $\Delta n = C_g \Delta V_g / e$, where C_g is the capacitance per unit area between the gate and the dot layer. Taking the geometrical parameters of the gate, relative permittivity $\epsilon = 3.9$ for SiO_2, and $\Delta V_g = 0.7$ V

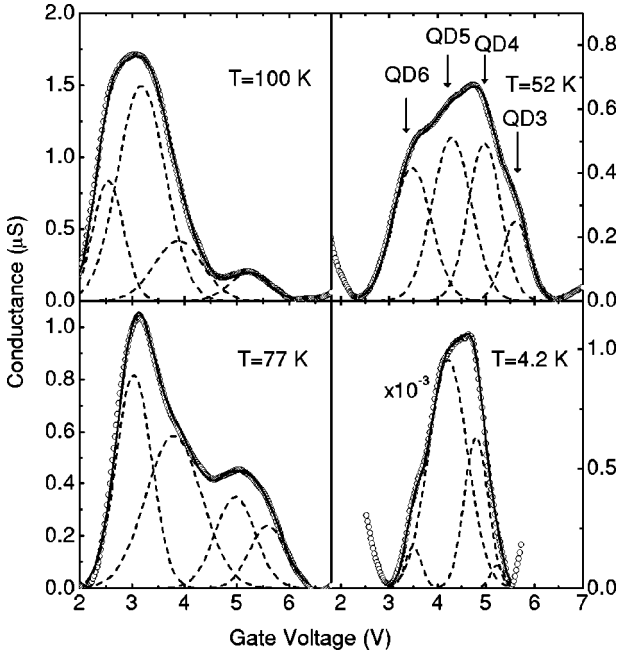

FIG. 3. Dot conductance vs gate potential for four temperatures. Circles show the experimental data with the background subtracted, and solid lines give the result of decomposition into Gaussians.

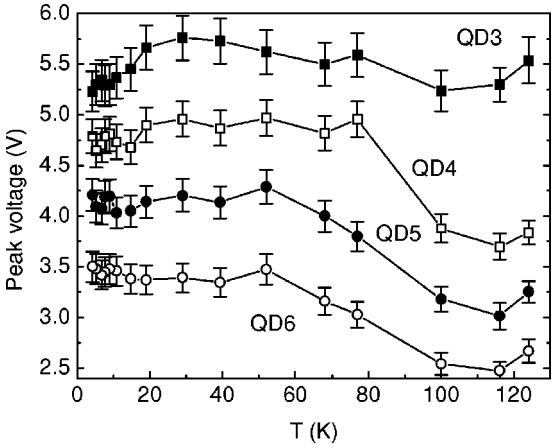

FIG. 4. The position of the peaks in Fig. 3 as a function of temperature.

for the separation of successive peaks, and supposing the electrostatic fields are one dimensional, we find $C_g = 4.7 \times 10^{-4}$ F m^{-2} and $\Delta n \approx 2.1 \times 10^{15}$ m^{-2}. The latter value is consistent with the density of quantum dots, $n_{QD} \approx 3 \times 10^{15}$ m^{-2}, strongly supporting the interpretation that each constituent conductance peak originates from loading of *one hole* into each dot. The maximum conductivity occurs when the given level is half-filled, as this maximizes the product of possible initial and final states for a tunneling process that avoids increasing on-site correlation energy.

The above analysis can be checked by using C-V_g measurements to verify the electrostatics necessary for charging of the dots. We measured the capacitance between the source and drain connected together and the gate. Remember that the source and drain contact the underlying Si. The 100-kHz C-V_g characteristics at $T = 300$ and 4.2 K are shown in Fig. 5. The capacitance falls with increasing V_g corresponding to the increasing depletion layer thickness in the Si below the gate oxide. At 4.2 K, we see, superimposed on the falling capacitance, a structure in the 3–6-V range associated with a filling of the excited state, and also a structure in the 8–9.5-V range which we attribute to a filling of the twofold-degenerate ground state. We may use values of C_g together

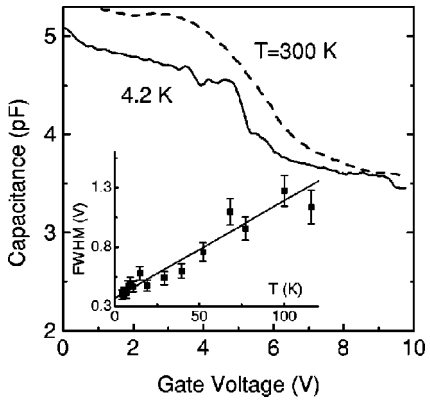

FIG. 5. Capacitance-voltage characteristics. The capacitance was measured between the source and drain joined together and the gate. Values have been corrected for an estimated stray capacitance of about 0.7 pF. Inset: Temperature dependence of the FWHM of the QD3 peak in Fig. 3 with a linear fit to the data.

with the background acceptor concentration in the Si, $N_B = 7\times10^{22}$ m^{-3} (determined by Hall measurements) to estimate the electrostatic configuration at various values of V_g. At $V_g = 9.5$ V, just before filling of the ground-state levels, $C_g \approx 3.4$ pF, which corresponds to the depletion layer extending to some 88 nm below the dot layer. The corresponding band bending between the deep Si and the dot layer is 414 mV. Just before the extended states start to fill, $C_g \approx 3.7$ pF, and the depletion thickness below the dot layer is 62 nm with a corresponding band bending of 220 mV. These simplistic results are compatible with the known Si-Ge valence-band offset and the expected energies of the hole bound states.[11] After the excited states are filled, $C_g \approx 4.7$ pF implying that the effective boundary of the depletion region is now at the dot layer. The depletion layer thickness d_d then continues to fall as V_g is decreased. At $V_g = 0$ V, $d_d = 23$ nm, and the dot layer is about 17 nm into the unperturbed Si. The capacitance measurements are thus consistent with our interpretation of the conductivity results.

At large positive V_g, capacitance only decreases weakly with increasing gate voltage in contrast to the conductance which shows a strong rise (see Fig. 2). Separate measurements confirm that in this region leakage current through the insulator becomes comparable with the source-drain current. Therefore, we conclude that the apparent increase of background conductance at $V_g > 6$ V is a result of leakage through the gate SiO_2.

The energy-level separation (ΔE) of the different charge states in the dots can be estimated by using $\Delta E = \eta e \Delta V_g$, where the gate modulation coefficient η relates the gate voltage to the hole energy inside the dot. This coefficient can be determined in two ways. One way is to calculate η from the temperature dependence of the full width at half maximum (FWHM) of the conductance peaks, which, for a single dot showing Coulomb blockade oscillations, should be broadened with T as $3.5k_BT/(\eta e)$ (Ref. 12) (k_B is Boltzmann's constant). By measuring the FWHM of the peak QD3 as a function of temperature (Fig. 5), we obtain $\eta = 3.4\times10^{-2}$, with a residual FWHM of about 0.37 V which is a result of statistical fluctuation in the dot ensemble. Another way of calculating η was proposed in Ref. 5. When most of the charge induced by a change of gate voltage, ΔV_g, is captured by the QD's as discussed above, then the change in potential of the dot is given by $\Delta\phi = e\Delta n/C_{\rm QD} = C_g\Delta V_g/(n_{\rm QD}C_{\rm QD})$, where $C_{\rm QD}$ is the dot self-capacitance. Thus $\eta = \Delta\phi/\Delta V_g = C_g/(n_{\rm QD}C_{\rm QD})$. The value of $C_{\rm QD}$ for a disk-shaped dot with diameter D in classical electrostatics is given by $C_{\rm QD} = 4\epsilon_r\epsilon_0 D$. For $D = 15$ nm and $\epsilon_r = \epsilon_{\rm Si} = 11.7$, this yields $C_{\rm QD} = 5.5$ aF and $\eta = 3.5\times10^{-2}$. Obviously, the agreement between the two above estimates of the gate-voltage–dot-energy modulation coefficient is very satisfactory. Based on these calculations, the estimated charging energy is about 23 meV. Again, this value is consistent with our previous admittance experiments.[10]

We may ask why we do not observe conductance maxima associated with a filling of the hole ground states (QD1 and QD2 maxima) as we did with our earlier structures.[5] The probable explanation is that the gate leakage current at $V_g \approx 9$ V, where the capacitance measurements show the ground states to be filling, prevents sufficient accuracy in a measurement of hopping conductivity through the ground-state levels which is expected to be smaller as a result of the reduced localization length of the wave functions.

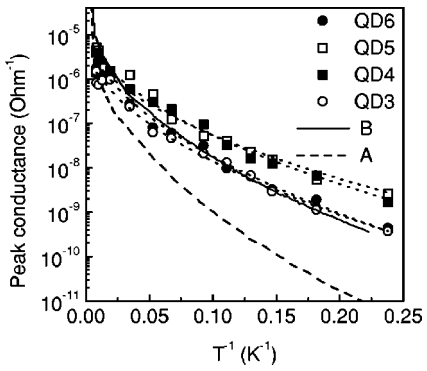

FIG. 6. Temperature dependence of the conductance maxima (symbols). Dotted lines are the best fit of the experimental data to Eq. (1). Broken and solid lines represent the tempeature dependence of conductance in the test samples A and B, respectively (see the text).

A useful way to identify the mechanism of carrier transport is to study the temperature dependence of the conductance. In the regime of resonant tunneling through discrete energy levels, conductivity depends weakly on temperature. The current peak height should increase as the temperature is reduced due to a diminished thermal broadening of the resonance. In contrast, hopping conduction is thermally activated (see Sec. I). Figure 6 shows the temperature dependence of the four conductance peaks QD3–QD6. For all maxima (symbols) we see a temperature-dependent activation energy reminiscent of variable-range hopping. A best fit to these curves (dotted lines) indicates that below 100 K the temperature dependence can be described by Eq. (1), with $x \simeq 0.5$ and $T_0 = 395$–565 K (see Table I).

We can check the hopping model for our structure and extract a value of localization radius ξ by making a quantitative comparison with the theoretical prediction. With $k_BT_0 = 6.2e^2/(4\pi\epsilon_r\epsilon_0\xi)$,[13] the spatial dimension of the wave functions is found to be $\xi = 15$–21 nm (Table I). For variable-range tunneling to occur, the temperature-dependent optimum hop distance $R_{\rm opt} = 0.25\xi(T_0/T)^{1/2}$ must be larger than both the localization length and interdot distance (3–4 nm). At $T = 10$ K and with $T_0 = 395$–565 K, we have $R_{\rm opt} = 29$–34 nm. Therefore these conditions are satisfied.

To obtain further evidence to support a hopping mechanism in the metal-oxide-semiconductor field-effect transister (MOSFET), we have fabricated two test samples without any

TABLE I. Fitting parameters for variable-range-hopping conduction through the charge-tunable quantum dots in the field-effect transistor.

Conductance maximum	x	T_0 (K)	ξ (nm)
QD6	0.49 ± 0.11	565 ± 58	15.4 ± 1.6
QD5	0.49 ± 0.14	395 ± 48	21.4 ± 3.1
QD4	0.51 ± 0.14	405 ± 49	21.1 ± 2.2
QD3	0.51 ± 0.10	536 ± 52	16.2 ± 1.5
Sample A	0.50 ± 0.01	1176 ± 36	7.6 ± 0.2
Sample B	0.51 ± 0.01	581 ± 37	15.0 ± 0.9

oxides or gate. Both samples contain a remotely doped layer of Ge QD's grown on a semi-insulating Si(001) substrate. In sample A the doping level is such that only the ground state contains holes, there being 3/2 holes per dot taken from impurities, while in sample B, the ground state is full and the first excited state is partially occupied with a total of 5/2 holes per dot. The results of $G(T)$ measurements for the two samples are shown in Fig. 6 by solid and broken lines, respectively. For both samples, best fits again give $x \simeq 0.5$, with $T_0 = 1176$ K for sample A,[14] and $T_0 = 581$ K for sample B. The latter is close to the value found for the corresponding state (QD3) in the MOSFET structure (see Table I). Moreover, the actual values of $G(T)$ for sample B match the temperature dependence of conductance maximum QD3. These results provide strong support for the assertion that the observed conductance oscillations originate from the hopping of holes through the discrete energy levels of the first excited state.

B. Screening of long-range Coulomb interaction in arrays of quantum dots

1. Crossover from Efros-Shklovskii to Mott VRH

It should be remarked that the ES form of $G(T)$ does not necessarily result from intersite correlations. It only requires a density of states having an appropriate dependence on energy in the vicinity of the Fermi level.[15] In this section we demonstrate that the hopping conductance of an array of QD's may be enhanced substantially, and show a crossover from ES VRH to Mott VRH with decreasing temperature by putting the dot layer in proximity with a metal plane. The results provide strong evidence of the dominant role played by long-range Coulomb interaction between the dots in electronic transport in ensembles of QD's.

The large spatial scale of intersite correlations allows one to examine the role of Coulomb interaction experimentally by making use of intentionally introduced screening. If one places a metal plane parallel to the conductive channel at a distance d, the interaction may be described by including images of the real charges in the screening electrode. If the separation of initial and final states in a hop is small compared with the distance between a charge and its image $(=2d)$, the screening electrode makes little difference, and the interaction remains monopolar. At large distances, however, a charge and its image behave as a dipole and interactions fall off more rapidly with distance. The general expression for the interaction potential is[16]

$$U(r) = \frac{e^2}{4\pi\epsilon_r\epsilon_0}\left(\frac{1}{r} - \frac{1}{\sqrt{r^2+4d^2}}\right). \tag{2}$$

Thus distance d plays the role of a screening length. This means that, at low temperatures, when the hopping distance becomes longer than about $2d$, initial and final states become electrostatically independent and one should observe a breakdown of ES VRH and a crossover to a Mott regime with $x = 1/3$ in two dimensions. One would expect the screened conductance to be larger than that in the unscreened regime. However, Entin-Wohlman and Ovadyahu[17] found a reduction of hopping conductivity and transition to a simply activated law in a screened In_xO_y film. A similar behavior was obtained by Adkins and Astrakharchik in experiments with ultrathin bismuth films.[18] The explantion put forward was that in those systems the screened hopping was to nearest neighbors with the observed activation energy simply a characteristic disorder energy. Only Van Keuls *et al.*[19] reported the observation of a universal crossover from ES to Mott hopping, driven by a variation of temperature, magnetic field, and electron density in $GaAs/Al_xGa_{1-x}As$ MOSFET's.

It should be pointed out that, in the Mott regime of a screened system, the effective (constant) density of states is not that which would be present in the absence of interaction. It is only the low-energy interactions, those corresponding to long distances, that are screened and they are the interactions responsible for the suppression of the density of states at low energies, close to E_F. At higher energies, the density of states still falls off similarly to the unscreened ES situation. Thus the constant density of states at low energies when there is screening is not equal to some background density of states, but is given by the ES density of states at the energy corresponding to charges separated by the effective screening length $(\sim 2d)$. The density of states at low energies with screening thus depends only on d and the local relative permittivity. One obtains[16]

$$g(0) = \alpha(4\pi\epsilon_r\epsilon_0/e^2 d), \tag{3}$$

where α is a numerical constant estimated in unpublished calculations by Mogilyanskii and Raikh as $\alpha \approx 0.1$ (see Ref. 16).

In this section we present results of low-temperature conductance studies in two types of samples. The samples of the first type (to be referred to as screened samples) have a planar metallic gate close to the dot layer. (The distance between the channel of QD's and the gate is $d = 35$ nm.) Samples of the second type (reference or unscreened structures) contain no gate electrode. The top oxide layer is present in both cases, however. In all structures the holes on the dots are supplied by boron impurities. The temperature dependence of the conductance $G(T)$ of screened and unscreened samples is shown in Fig. 7 as Arrhenius plots. In contrast with the experiments,[17,18] the low-temperature conductance of the screened QD systems is found to be larger than that of unscreened samples except for the $N = 1/2$ sample, where $G(T)$ does not change significantly with screening (in the range of temperature studied).

To analyze the characteristic behavior of $G(T)$, we examine the temperature dependence of the reduced activation energy $w(T) = d\ln G/d\ln T$.[20] For an exponential hopping dependence of G, $w(T) = x(T_0/T)^x$, and

$$\log_{10} w(T) = A - x\log_{10} T, \quad A = x\log_{10} T_0 + \log_{10} x. \tag{4}$$

Thus a plot of $\log_{10} w(T)$ against $\log_{10} T$ yields the values of the exponent x from the slope and of the characteristic temperature T_0 from the y intercept A: $T_0 = (10^A/x)^{1/x}$. Figure 8 contains $w(T)$ data obtained by numerical differentiation of the $G(T)$ curves. Linear regression was used to determine the best slopes x and the best intercepts A for both high- and low-temperature intervals (solid lines in Fig. 8).

The fitting results are summarized in Table II. In unscreened samples with $N > 1/2$, the slope of the w plots yields $x \approx 0.5$, consistent with ES VRH over the whole temperature

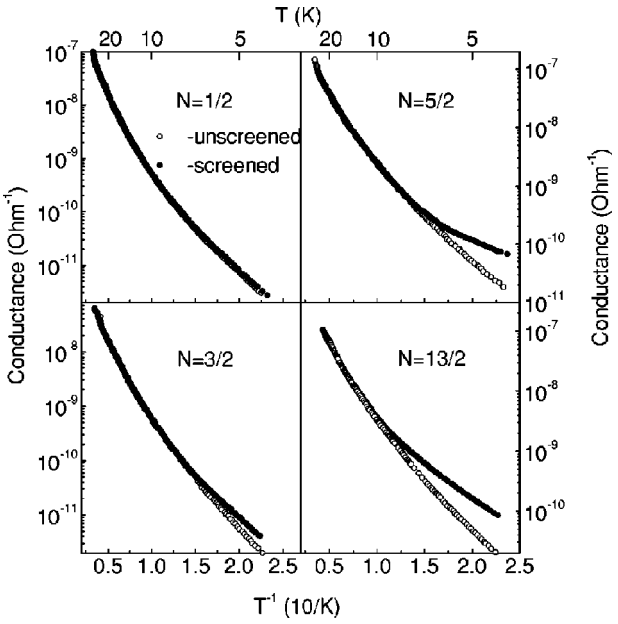

FIG. 7. Temperature dependence of conductance for various dot occupations.

range. Screened samples with $N>1/2$, however, exhibit a crossover from ES VRH at high temperatures to Mott VRH, with $x\approx 1/3$, at low temperatures. As explained above we expect this to occur when the optimum hop distance, $R_{\rm opt}$, becomes greater than the screening length.

To check this interpretation of the results, we first calculate the localization length ξ. This may be done in several ways. First, ξ may be obtained directly from $T_{\rm ES}$ and T_M using the relations[16]

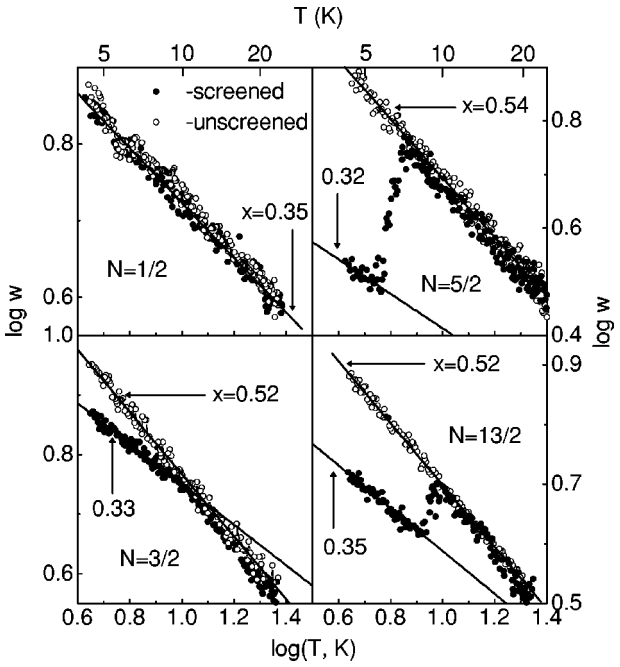

FIG. 8. Plot of $\log_{10}$ of the dimensionless activation energy w against $\log_{10} T$. The solid lines are least-square fits to linear dependencies. The slopes of the solid lines yield the hopping exponents x and the characteristic hopping temperature T_0 is found from the y intercepts. The fitting parameters of all the curves are listed in Table II.

TABLE II. Fitting parameters for variable-range-hopping conduction through the quantum dots in screened and unscreened regimes.

		Data at $T>T_{\rm cross}$		Data at $T<T_{\rm cross}$		
			$T_0=T_{\rm ES}$		$T_0=T_M$	$T_{\rm cross}$
N	Screened	x	(K)	x	(K)	(K)
1/2	no			0.38	4.8×10^4	
1/2	yes			0.35	5.4×10^4	
3/2	no	0.52	1260			
3/2	yes	0.53	1392	0.33	4.3×10^4	6.5
5/2	no	0.54	900			
5/2	yes	0.52	890	0.32	1.0×10^4	6.7
13/2	no	0.52	1044			
13/2	yes	0.54	1072	0.35	2.6×10^4	8.7

$$k_B T_{\rm ES}=6.2e^2/4\pi\epsilon_{\rm r}\epsilon_0\xi, \tag{5}$$

$$k_B T_M=14e^2 d/4\pi\epsilon_{\rm r}\epsilon_0\xi^2\alpha. \tag{6}$$

We estimate the effective relative permittivity in our structures as $\epsilon_{\rm r}\approx 9$.[21] Alternatively, the permittivity can be eliminated between 5 and 6 to give

$$\xi_{ES/M}=(2.26/\alpha)d(T_{\rm ES}/T_M). \tag{7}$$

Fourth, ξ may be estimated from $T_{\rm cross}$, the temperature of crossover from ES to Mott behavior. Aleiner and Shklovskii[16] took this to occur when the ES and Mott VRH tunneling exponenets are equal. This gives

$$\xi_{\rm cross}=\left(\frac{0.822}{\alpha^2}\right)\frac{4\pi\epsilon_{\rm r}\epsilon_0 d^2}{e^2}k_B T_{\rm cross}. \tag{8}$$

These various estimates of ξ, calculated using $\alpha=0.1$, are given in the first part of Table III. We note that values increase from left to right and that $\xi_{\rm cross}$ becomes unreasonably large. We suggest that the value of α (0.1) obtained Mogilyanskii and Raikh is too small. If we take $\alpha=0.4$, we obtain

TABLE III. Localization length ξ calculated as described in the text.

N	Screened	$\xi_{\rm ES}$ (nm)	ξ_M (nm)	$\xi_{ES/M}$ (nm)	$\xi_{\rm cross}$ (nm)
Taking $\alpha=0.1$					
3/2	no	9.1			
3/2	yes	8.3	15	26	350
5/2	no	13			
5/2	yes	13	30	70	360
13/2	no	11			
13/2	yes	11	19	33	470
Taking $\alpha=0.4$					
3/2	no	9.1			
3/2	yes	8.3	7.3	6.4	22
5/2	no	13			
5/2	yes	13	15	18	23
13/2	no	11			
13/2	yes	11	9.3	8.2	30

TABLE IV. Calculated optimum hop distance at $T_{\rm cross}$.

N	$\langle\xi\rangle$ (nm)	$R_{\rm opt,ES}$ (nm)	$R_{{\rm opt},M}$ (nm)
3/2	7.8	29	48
5/2	14	40	53
13/2	10	28	48

the values of ξ given in the second half of Table III. The first three estimates of ξ are now in much better general agreement. The values calculated from $T_{\rm cross}$, however, remain generally large. To explain this we suggest that it is not sufficiently accurate to take $T_{\rm cross}$ to occur when the ES and Mott exponenets are equal (as did by Aleiner and Shklovskii). Fitting VRH equations to the two regions gives different preexponential factors, and these should be included when equating the conductivities. This would introduce a logarithmic corrections when calculating $\xi_{\rm cross}$. While the last point deserves further investigation, we believe our analysis give a consistent interpretation of the data.

It now remains to check that the optimum hop distance $R_{\rm opt}$ is of order d at $T_{\rm cross}$. We use the standard theoretical results for two dimensions:

$$R_{\rm opt,ES}(T_{\rm cross})=(\xi/4)(T_{\rm ES}/T_{\rm cross})^{1/2}, \tag{9}$$

$$R_{{\rm opt},M}(T_{\rm cross})=(\xi/3)(T_M/T_{\rm cross})^{1/3}. \tag{10}$$

These values, calculated using the mean of $\xi_{\rm ES}$ and ξ_M, are given in Table IV, and are satisfactory.

We return to the $N=1/2$ results. Here screening only produces a very small reduction in x from 0.38 to 0.35, and no crossover is seen. The absence of ES VRH for this sample can be understood as follows. At $N\ll 1$, most of the dots contain neither holes nor nearby impurities and are neutral. A dot is charged only when it contains a carrier (hole) supplied by the rare impurities. Since a displaced hole leaves behind the neutral state, the correlation between initial and final sites in most of the hops becomes unimportant; relevant energies are dominated by disorder, and one observes Mott conduction. In fact, one should remember that not all transitions are from singly charged sites to neutral sites, so correlation energies are not entirely absent. This is probably why the hopping exponent in the $N=1/2$ sample is slightly larger than 1/3 predicted for the ''pure'' Mott hopping.

An interesting feature of our results is the form of the transition as seen in the temperature dependence of w shown in Fig. 8. First, the transition is extraordinarily sharp. If it simply resulted from a situation in which two different processes were present, with the transition occurring as one becomes more dominant than the other, then one would expect it to be much more spread out. Second, with larger values of N, we see an actual discontinuity of w at the transition. There is a certain resemblance to second- (3/2) and first-order ($N=5/2$ and 13/2) thermodynamic phase transitions. One wonders whether there may not be a cooperative aspect to the transition from the Mott to the ES regime as correlation energies become greater.

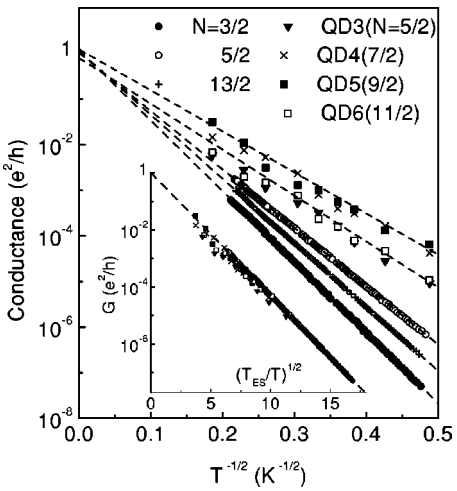

FIG. 9. The conductance $G(T)$ vs $T^{-1/2}$ for different dot occupation N in unscreened samples. Data from Fig. 6 are also included. The inset shows the same data plotted vs $(T_{\rm ES}/T)^{1/2}$.

An important observation is that the crossover, when w is discontinuous, is characterized by a *drop* of w at $T_{\rm cross}$, corresponding to a reduction of the hopping activation energy as a result of the screening. This behavior is direct evidence of the suppression of long-range correlations between initial and final hole sites on the dots.

2. Universal prefactor in unscreened regime of VRH

Notice that the data obtained in Sec. III B 1 imply a temperature-independent prefactor G_0. Figure 9 shows the conductance in units of e^2/h, the quantum of conductance, of unscreened samples with different dot occupation N plottted versus $T^{-1/2}$; the symbols are the experimental points and the broken lines are the least-squares fits to the $T^{-1/2}$ ES dependence. Here we also plot the amplitude of conductance maxima taken from Fig. 6.[22] An impressive feature is that all the curves extrapolate to the same prefactor $G_0\simeq e^2/h$. This is more evident when $G(T)$ is plotted against the dimensionless parameter $(T_{\rm ES}/T)^{1/2}$ (see the inset of Fig. 9). In this plot, the data collapse onto a single universal curve with intercept $G_0=(1.05\pm 0.05)e^2/h$. This observation is similar to that found for the two-dimensional impurity hopping conductance in both ES and Mott unscreened regimes in Si MOSFET's (Ref. 23) and in δ doped GaAs/Al_xGa_{1-x}As heterostructures.[24] The universality of the prefactor signals against the conventional phonon-assisted hopping mechanism, where the prefactor would depend on material parameters such as the localization length. To resolve this discrepancy, it was suggested by Baranovskii and Shlimak[25] that the phononless hopping is assisted by electron-electron interaction. According to this model the current-carrying single electron moves via quantum resonant tunneling between localized states brought into resonance by a time-dependent random Coulomb potential created by fast electron transitions in their environment. The dependence of the fluctuation amplitude of energies of hopping sites on temperature gives rise to the temperature dependence of the conductance.[26]

If the universal prefactor does result from interaction, its universality would be destroyed in the presence of screening.

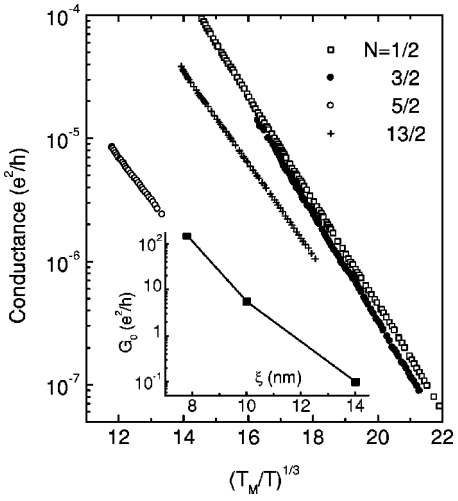

FIG. 10. Plots of the conductance $G(T)$ vs $(T_M/T)^{1/3}$ for screened samples at $T<T_{\mathrm{cross}}$. The inset shows the dependence of the prefactor on the localization length.

The ratio $G(T)/(e^2/h)$ is plotted against $(T_M/T)^{1/3}$ in Fig. 10 for data from screened samples at temperatures below T_{cross}. In contrast to the unscreened regime, the data do not collapse onto a single curve. The best-fit value of $G_0 \equiv G_M$ as a function of localization length $\langle \xi \rangle$ (taken from Table IV) is shown in the inset of Fig. 10. Obviously, in the fully screened limit, there is a strong dependence of the prefactor on the localization radius.

IV. SUMMARY

We have described a set of experiments in which we studied hopping transport in field-effect structures containing from 3×10^7 to 10^9 quantum dots. We demonstrate that below $\sim$100 K this system is able to show conductance oscillations associated with a filling of the dots by successive single holes. From the temperature dependence of the conductance maxima, we identify the conduction mechanism as variable-range hopping in a density of states determined by Coulomb interaction between the dots. In samples with screening, we observe a crossover from Efros-Shklovskii VRH with $\ln\sigma \propto T^{-1/2}$ to Mott VRH with $\ln\sigma \propto T^{-1/3}$ as the temperature is reduced. The data in the ES regime collapses onto a universal curve with a prefactor $G_0 \simeq e^2/h$, while all traces in the screened regime do not show the universal behavior. The results demonstrate the important role of long-range interdot Coulomb interaction in dense ensembles of quantum dots, and they raise interesting issues relating to the mechanisms of the hopping processes.

ACKNOWLEDGMENTS

The authors would like to thank B. Shklovskii, A. Dahm, I. Shlimak, and S.D. Baranovskii for some helpful discussions. This work was supported by grants from the Russian Foundation of Basic Research (Grant No. 99-02-17019), the Interdisciplinary Scientific and Technical Program ''Physics of Solid State Nanostructrures'' (Grant No. 98-1100), and the Intercollegiate Scientific Program ''Universities of Russia-Basic Research'' (Grant No. 4103).

[1] For a review, see *Single-Charge Tunneling-Coulomb Blockade Phenomena in Nanostructures*, in *NATO Advanced Study Institute, Series B: Physics*, edited by H. Grabert and M. H. Devoret (Plenum Press, New York, 1991).

[2] F. Hofmann, T. Heinzel, D. A. Wharam, J. P. Kotthaus, G. Böhm, W. Klein, G. Tränke, and G. Weimann, Phys. Rev. B **51**, 13 872 (1995); D. Dixon, L. P. Kouwenhoven, P. L. McEuen, Y. Nagamune, J. Motohisa, and H. Sakaki, *ibid.* **53**, 12 625 (1996); R. H. Blick, R. J. Haug, J. Weis, D. Pfannkuche, K. von Klitzing, and K. Eberl, *ibid.* **53**, 7899 (1996); R. J. Haug, J. Weis, R. H. Blick, K. von Klitzing, K. Berl, and K. Ploog, Semicond. Sci. Technol. **11**, 381 (1996); R. H. Blick, T. Schmidt, R. Haug, and K. von Klitzing, *ibid.* **11**, 1506 (1996); T. Schmidt, R. J. Haug, K. von Klitzing, A. Förster, and H. Lüth, Phys. Rev. Lett. **78**, 1544 (1997).

[3] C. I. Duruöz, R. M. Clarke, C. M. Marcus, and J. S. Harris, Jr., Phys. Rev. Lett. **74**, 3237 (1995).

[4] M. G. Ancona and R. W. Rendel, J. Appl. Phys. **77**, 393 (1995).

[5] A. I. Yakimov, C. J. Adkins, R. Boucher, A. V. Dvurechenskii, A. I. Nikiforov, O. P. Pchelyakov, and G. Biskupski, Phys. Rev. B **59**, 12 598 (1999).

[6] A. I. Yakimov, A. V. Dvurechenskii, Yu. Prockuryakov, A. I. Nikiforov, O. P. Pchelyakov, S. A. Teys, and A. K. Gutakovskii, Appl. Phys. Lett. **75**, 1413 (1999).

[7] A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, and O. P. Pchelyakov, Thin Solid Films **336**, 332 (1998).

[8] We label the dot states in terms of the number of holes on each dot. For example, the peak observed at $V_g \approx 5.5$ V is labeled QD3 because it corresponds to loading the third hole into the dots, this hole entering the first of the excited states with the two ground-state levels already fully occupied. The peak corresponds to a mean loading of 5/2 holes per dot.

[9] The observed shift of the peak's position to lower gate voltage at $T>80$ K is attributed to the temperature-induced diminishing of the Si and Ge band gaps. This effect is commonly observed in photoluminescence measurements performed on samples with QD's at different temperatures.

[10] A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, and O. P. Pchelyakov, Phys. Low-Dimens. Semicond. Struct. **3/4**, 99 (1999).

[11] S. K. Zhang, H. J. Zhu, F. Lu, Z. M. Jiang, and Xun Wang, Phys. Rev. Lett. **80**, 3340 (1998).

[12] C. W. J. Beenaker, Phys. Rev. B **44**, 1646 (1991).

[13] V. L. Nguyen, Fiz. Teckh. Poluprovodn. **18**, 335 (1984) [Sov. Phys. Semicond. **18**, 207 (1984)].

[14] The characteristic length of the ground state turns out to be less than that of the excited state by a factor of 2–3. This is consistent with our previous magnetoresistance measurements (Ref. 5).

[15] B. I. Shklovskii and A. L. Efros, in *Electronic Properties of Doped Semiconductors* (Springer-Verlag, Heidelberg, 1984).

[16] I. L. Aleiner and B. I. Shklovskii, Phys. Rev. B **49**, 13 721 (1994).

[17] O. Entin-Wohlman and Z. Ovadyahu, Phys. Rev. Lett. **56**, 643 (1986).

[18] C. J. Adkins and E. G. Astrakharchik, J. Phys.: Condens. Matter **10**, 6651 (1998).

[19] F. W. Van Keuls, X. L. Hu, H. W. Jiang, and A. J. Dahm, Phys. Rev. B **56**, 1161 (1997).

[20] A. G. Zabrodskii and K. N. Zinov'eva, Zh. Éksp. Teor. Fiz. **86**, 727 (1984) [Sov. Phys. JETP **59**, 425 (1984)].

[21] We estimate the average relative permittivity as $\epsilon_r = 0.5[\epsilon_{Si} + (d_{Si}\epsilon_{Si} + d_{SiO_2}\epsilon_{SiO_2})/(d_{Si} + d_{SiO_2})] \approx 9$, where $\epsilon_{Si} = 12$ and $\epsilon_{SiO_2} = 3.9$ are the relative permittivities of Si and SiO_2, respectively.

[22] The $N = 5/2$ unscreened sample and the QD3 conductance maximum (which corresponds also to $N = 5/2$) show different slopes due to different effective permittivities.

[23] W. Mason, S. V. Kravchenko, G. E. Bowker, and J. E. Furneaux, Phys. Rev. B **52**, 7857 (1995).

[24] S. I. Knondaker, I. S. Shlimak, J. T. Nicholls, M. Pepper, and D. A. Ritchie, Phys. Rev. B **59**, 4580 (1999).

[25] S. D. Baranovskii and I. Shlimak, cond-mat/9810363 (unpublished).

[26] I. Shlimak, S. D. Baranovskii, and V. I. Kozub, in *Abstracts of the 8th International Conferences on Hopping and Related Phenomena, Murcia, 1999* (Universidad de Murcia, Spain, 1999), p. 30.