Dipole trap model for the metal-insulator transition in gated silicon-inversion layers

T. Hörmann
Institute for Semiconductor Physics, Johannes Kepler University, 4040 Linz, Austria and Christian Doppler Labor for Surface Optics, Johannes Kepler University, 4040 Linz, Austria

G. Brunthaler*
Institute for Semiconductor Physics, Johannes Kepler University, 4040 Linz, Austria
(Received 30 December 2009; revised manuscript received 21 August 2010; published 9 November 2010)

In order to investigate the metal-insulator transition in high-mobility Si-metal-oxide-semiconductor structures, we have precised and further developed the dipole trap model as originally proposed by Altshuler and Maslov [Phys. Rev. Lett. **82**, 145 (1999)]. Our additional numerical treatment enables us to drop several approximations and to introduce a limited spatial depth of the trap states inside the oxide as well as to include a distribution of trap energies. Depending on the type and width of distribution, the metallic state appears more or less pronounced as observed in experiments on samples with different quality.

DOI: 10.1103/PhysRevB.82.205310 PACS number(s): 71.30.+h, 73.40.Qv, 72.10.Fk

I. INTRODUCTION

The discovery of the metal-insulator transition (MIT) in two-dimensional (2D) electron systems in 1994 (Refs. 1 and 2) has attracted large attention, as it was in apparent contradiction to the scaling theory of localization.[3,4] This theory points out that in the limit of zero temperature, a metallic state should exist only in three-dimensional (3D) systems whereas in two-dimensional disorder should always be strong enough to induce an insulating state. The MIT in high-mobility n-type silicon inversion layers shows a strong decrease in resistivity ρ toward low temperature T for high electron densities, manifesting the metallic region, whereas an exponential increase in ρ demonstrated the insulting regime at low densities. A similar but weaker behavior was observed in many other semiconductor systems at low densities and low temperatures [e.g., p-GaAs,[5] n-GaAs,[6] SiGe,[7] and AlAs (Ref. 8)].

Several models were suggested in order to explain the unexpected finding of metallic behavior in 2D. The most important ones are (i) temperature-dependent screening,[9–14] (ii) quantum corrections in the diffusive regime,[15–18] and (iii) quantum corrections in the ballistic regime.[19,20] Numerous argumentations for the different models are given in literature[21–25] but a clear decision for one of them could not been drawn yet.

In order to take into account the high density of defect states in the oxide of Si-metal-oxide-semiconductor (MOS) structures, Altshuler and Maslov (AM) introduced the dipole scattering scenario for such structures in which charged trap states form dipoles together with the image charge of the screening 2D electrons.[26] The interplay between the gate-voltage-dependent energetic position of the trap states and the height of the chemical potential may lead as well to a metal-insulator transition in that system. It should not be assumed that the dipole scattering is the only effect in action, as the temperature dependence of screening and quantum corrections will contribute at low temperatures. But the charging of defect states and the resulting dipole scattering might be the generator of the particularly large effect in Si-MOS structures. The charged defect states in the thermally grown oxide layer are induced by the misfit at the silicon/silicon-oxide interface.[27–29] Arguments on the importance of trap states in Si-MOS structures were also given by Klapwijk and Das Sarma.[30]

AM could show within their analytical calculations that a trap level at energy E_T which is either filled or empty, depending on its position relative to the chemical potential μ, can lead to a critical behavior in electron scattering if E_T and μ are degenerate. This dipole trap model is able to explain the main properties of the metal-insulator transition in gated Si-MOS structures.

For the analytical calculations, AM made a number of assumptions. These are: (a1) the trap states possess a δ-like distribution in energy (i.e., have all the same energy), (a2) the spatial trap density distribution in the oxide is homogeneous, (a3) the states occupied with electrons behave neutral and cause no scattering of 2D electrons whereas the unoccupied states are positively charged and lead to scattering (AM work in the hole-trap picture, we describe occupation in terms of electrons), (a4) a charged trap state is screened by the 2D electrons so that the resulting electrostatic potential can be described by the trap charge and an apparent mirror charge with opposite sign on the other side of the interface, (a5) the scattering efficiency of the 2D electrons is described by a dipole field of the trap charge and its mirror charge, (a6) a parabolic saddle-point approximation for the effective potential between the Si/SiO2 interface and the metallic gate was used in order to perform analytical calculations, (a7) the energy of the trap state E_T is fixed relative to the quantization energy E_0 of the 2D ground state inside the inversion potential, and (a8) the chemical potential μ in the 2D layer has (A) either the same temperature dependence as in the bulk substrate or (B) as in a 2D electron system with constant electron density.

In this work, we precise and develop the dipole trap model further in order to better understand the influence of charged traps on the metallic state and on the metal-insulator transition in Si-MOS structures. We present some improvements in the analytical description but perform mainly detailed numerical calculations. In the analytical treatment, some expressions are rearranged so that the scattering effi-

ciency is expressed in the same way as the usual Drude formulation. The two main integrals can then be rewritten as Fermi-Dirac integrals which lead to simple analytical approximations of the resistivity ρ at low temperatures.

Due to the numerical treatment, we are able to drop approximations (a6), (a7), and (a8) of the analytic AM model. In general we find good agreement with the calculations of AM in a wide parameter range. There are deviations in the overall behavior of the resistivity at low electron densities and low temperatures as well as for all densities at high temperatures. A detailed calculation of the common chemical potential μ for the 2D electron layer, the depletion layer, and the Si-bulk gives a behavior similar to case (B) of assumption (a8) and does not show a distinct insulating behavior in that the resistivity increases exponentially for low densities toward low temperatures. Nevertheless, there is a clear transition with electron density between a region where the resistivity strongly drops (metallic region) and one in which ρ saturates. A true insulating region may occur either due to further effects which influence the temperature dependence of μ or by the inclusion of, e.g., quantum corrections in the weak and strong localization regime.

As the main feature of this work, we have further extended our calculations for the realistic case with energetic broadening and spatial distribution profile of the defect states, i.e., dropping approximations (a1) and (a2), respectively. Depending on the strength of the energetic broadening, the transition between metallic and insulating state is more or less smeared out, the limitation of the spatial extent leads to a lower and thus more realistic resistivity at high temperatures.

The paper is organized as follows. In Sec. II, we give a short introduction into the trap model and discuss the improvements, which we have performed. Section III shows a direct comparison between analytical approximation and numerical results under equivalent conditions. Section IV treats the changes which occur, if the trap state energy if fixed relative to the conduction-band (CB) edge instead of the electronic ground state inside the inversion layer. The main results of this work are presented in Sec. V where a spatial trap profile is introduced and in Sec. VI for an energetic broadening of the trap states. These two additional extensions of the trap model bring the calculated temperature and density dependence of the resistivity into very good agreement with experimental observations. Conclusions are drawn in Sec. VII. In Appendix A details of the trap model, Appendix B the behavior of the chemical potential, Appendix C a low-temperature approximation, Appendix D the ground-state energy of the inversion layer, and Appendix E calculation parameters for the Si-MOS structures are given. Please note that we use SI units throughout this work, in contrast to cgi units in the work of AM.

II. DIPOLE TRAP MODEL

In the AM model, it is assumed that a large number of hole-trap states exists in the oxide at a certain trap energy E_T. If the trap energy lies above the chemical potential μ, the trap is empty (in the electron picture, or has captured a hole in an equivalent description) and is positively charged whereas if E_T lies below μ it is filled with an electron and thus is neutral, see Fig. 1. Please note that we use the terminology $\mu(T)$ for the chemical potential and the Fermi energy E_F denotes $\mu(T=0)$ as in AM.[26]

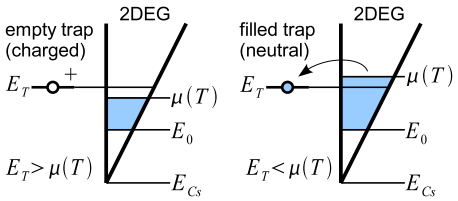

FIG. 1. (Color online) Schematic representation of the trap states together with the 2D electron system in the Si inversion layer. For $E_T > \mu(T)$ the trap state is positively charged and scatters electrons in the 2D layer whereas for $E_T < \mu(T)$ the trap is neutral, with E_{Cs} being the CB edge and E_0 the electronic ground-state energy at the oxide-semiconductor interface.

An applied gate voltage V_g at the Si-MOS structure causes a potential gradient in the oxide and a corresponding decrease in the trap energy,

$$E_T = E_{Ts} - eV_{\text{ins}}Z/D \tag{1}$$

inside, with the unscreened trap energy E_{Ts} at the oxide-semiconductor (OS) interface, the voltage drop across the oxide (insulator) V_{ins}, the distance from the OS interface Z, and the thickness of the oxide D (AM use the symbol V_g instead of V_{ins} in the corresponding equation, but with a somewhat different meaning, see also Appendix A).

As the charges within the depletion zone (2D charge density $-en_{\text{depl}}$) of the Si inversion layer also contribute to the gradient of the potential, we use the equation

$$V_{\text{ins}} = e(n_s + n_{\text{depl}})D/\epsilon_{\text{ins}}\epsilon_0 \tag{2}$$

together with

$$V_g = V_{\text{depl}} + V_{\text{ins}} \tag{3}$$

with V_{ins} the potential difference across the oxide layer, ϵ_{ins} the dielectric constant of the oxide, and ϵ_0 the electric field constant.

As one of the basic concepts behind the dipole trap model, AM describe the interaction between the charged trap state at energy E_T and the 2D electron gas (2DEG) by an image force, which corresponds to an efficient screening of the trap potential inside the semiconductor by the 2DEG. For the interaction term, we use $-e^2/16\pi\epsilon_{\text{ins}}\epsilon_0 Z$ (in SI units) whereas AM have a factor 8 in the denominator (see also, for instance, Ref. 27 or Ref. 28). But the size of this factor does not change the results of the trap model qualitatively, it mainly shifts the maximum in the effective trap potential somewhat and so also the corresponding values for n_s. With the interaction term the effective trap energy can be formulated as

$$E_T(Z) = E_{Ts} + \varepsilon_e(Z), \tag{4}$$

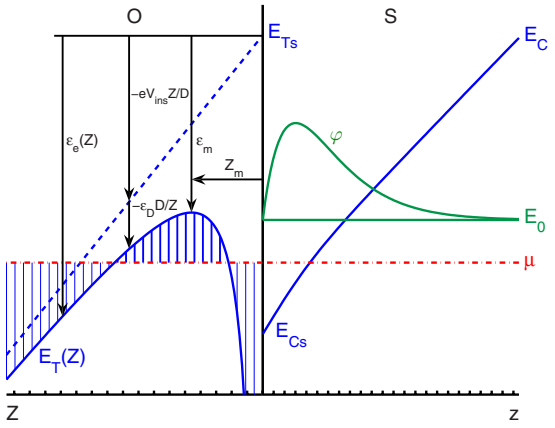

FIG. 2. (Color online) Energy diagram of the oxide-semiconductor interface with relevant notation. The effective trap energy E_T reaches its maximum at $Z=Z_m$, the unscreened trap energy E_{Ts} at the OS interface. The diagram further shows the electrostatic energy ε_e, its components $-eV_{\rm ins}Z/D$ and $-\varepsilon_D D/Z$ and the value ε_m at the maximum, the chemical potential μ, the ground-state energy of the inversion layer E_0 and the corresponding wave function φ, CB edge E_C and its value at the interface E_{Cs}. For simplicity we use different coordinate systems for the oxide (Z) and the semiconductor side (z) so that both, Z and z, are positive on their sides. At those positions where E_T lies above μ, the traps are positively charged (with the usual smearing over k_BT) which is indicated by the bold vertical hatch lines.

$$\varepsilon_e(Z) = -eV_{\rm ins}\frac{Z}{D} - \varepsilon_D\frac{D}{Z}, \tag{5}$$

where the subscript e in ε_e stands for "electrostatic" and

$$\varepsilon_D = \frac{e^2}{16\pi\epsilon_{\rm ins}\epsilon_0 D} \tag{6}$$

is used to abbreviate the image force interaction term. Please note that we use throughout this work the notation of capital E for absolute energies and greek ε for energy differences. The shape of effective trap energy $E_T(Z)$ is shown in Fig. 2 together with the relevant energies.

The number of positively charged trap states $p_+(Z)$ is described by a modified Fermi-Dirac statistics into which the energy difference $E_T(Z)-\mu$ enters, see Appendix A. The scattering of the 2D electrons is described in the AM trap model by the dipoles which are formed between the positive traps and the negative image force. The scattering is described by a classical transport cross section σ_t for dipoles (Appendix A). With that assumption, one gets the resistivity ρ from the Drude theory with the Boltzmann equation in the relaxation-time approximation. AM expressed the result in a very compact way, with the disadvantage that the connection to the standard Drude theory is somewhat hidden [(see AM (Ref. 26)].

We have repeated the calculations (Appendix A) and express the resistivity ρ in a form equivalent to the Drude formula

$$\rho = \frac{m_c}{n_s e^2\langle\tau\rangle}, \tag{7}$$

and find a term for the scattering rate $1/\langle\tau\rangle$ in the usual form that scattering rate is equal to scattering cross section times density of scattering centers times velocity of scattered particles, which leads to one of the basic equations used throughout this work,

$$\rho = \frac{m_c\sigma_t(\bar{\varepsilon},\bar{Z})n_T^+ v(\bar{\varepsilon})}{n_s e^2}. \tag{8}$$

Here $\sigma_t(\bar{\varepsilon},\bar{Z})$ is the transport scattering cross section according to the dipole scattering for an appropriately averaged effective electron energy $\bar{\varepsilon}$ and an effective distance $\bar{Z}$ between the traps and the OS interface, $n_T^+=\int_0^D N_T^+(Z)dZ$ is the 2D density of charged traps, $v(\bar{\varepsilon})$ the electron velocity which corresponds to $\bar{\varepsilon}=m_c v^2/2$, and $N_T^+(Z)$ denotes the 3D density of charged traps. For detailed calculations see Appendix A.

An important point to note is that AM have assumed that the energy E_{Ts} is constant relative to the 2D electron ground-state energy E_0, i.e., $\varepsilon_{TsE0}=E_{Ts}-E_0=$const. But E_0 depends on the width of the inversion potential (and thus on the gate voltage V_g) and there is no reason that the deep trap energy E_{Ts} should follow E_0. This assumption was necessary in order to be able to solve the model within an analytical treatment. But due to the numerical treatment in our work, we can drop this assumption [mentioned as (a7) in Sec. I] and assume that the trap energy is rather fixed relative to the CB edge at the interface E_{Cs} (the index s denotes energies at the oxide-semiconductor interface). We thus use

$$\varepsilon_{TsCs} = E_{Ts} - E_{Cs} = \text{const.} \tag{9}$$

This issue will be further treated in Sec. IV.

For the temperature dependence of the chemical potential AM described two different scenarios in their work: (A) the chemical potential of the 2DEG and of the Si substrate coincide, (B) the 2DEG is disconnected form the substrate. Case (A) implies that the temperature behavior in the 2DEG is the same as in the bulk. However, this does not take into account that the chemical potential in the 2DEG is measured against the ground-state energy E_0 and in the bulk against the conduction- or valence-band edge, i.e., it assumes E_0 and the band bending to be fixed. To our opinion, case (A) of AM is not very precise and has to be improved by the missing potential and temperature dependencies. Case (B) seems to be even less realistic, as it does not take into account the temperature and density dependence of E_0 relative to E_{Cs} and the changes in the inversion layer. Thus we modify also assumption (a8) on temperature dependencies like (A) or (B).

As we perform numerical calculations, it is possible to treat a realistic scenario with a common chemical potential μ for 2DEG, inversion layer, and Si-bulk material in thermal equilibrium. For that purpose, the potential run in the 2DEG and the inversion layer, the energy of the electronic ground state E_0 and the charge neutrality have to be solved self-consistently (see Appendices B and D). The parameters for

TABLE I. Values used for calculations.

$g_s=2$
$g_{v2D}=2$
$g_{v3D}=6$
$m_e=9.1094\times10^{-31}$ kg
$m_{d2D}=0.1905m_e$
$m_z=0.9163m_e$
$m_{de3D}=0.322m_e$
$m_{dh3D}=0.59m_e$
$\varepsilon_{sc}=11.9$
$\varepsilon_{ins}=3.9$
$\rightarrow\varepsilon^*=7.9$
$D=200$ nm
$C=16$
$\rightarrow\varepsilon_D=0.4615$ meV
$N_A=2\times10^{15}$ cm^{-3}
$N_D=0$
$\varepsilon_{g0Si}=1.17$ eV
$\alpha_{Si}=4.73\times10^{-4}$ eV K^{-1}
$\beta_{Si}=636$ K

the calculations performed for a {001} silicon surface plane are collected in Appendix E. It will be shown that this realistic scenario is close to case (A) of AM.

Although we perform mainly numerical calculations in this work, in Appendix C an analytical approximation for the resistivity ρ in the low-temperature limit is given for the cases where the maximum of the effective trap energy E_T is either above or below the chemical potential μ or coincides with it.

III. COMPARISON OF ANALYTIC AND NUMERICAL RESULTS

In order to get rid of the restrictions from the saddle-point approximation, we have performed numerical calculations of $\sigma_t(\bar{\varepsilon},\bar{Z})$ and n_T^+ as described in Appendix A. At first we compare the numerical calculation without the saddle-point approximation with the saddle-point approximation of AM for case (A).

In Fig. 3, we show how the resistivity ρ depends on temperature T and 2D electron density n_s in the inversion layer according to the analytic saddle-point approximation of AM (Ref. 26) (dashed lines), the analytic approximations for low temperatures where the Fermi-Dirac integrals are replaced by corresponding approximations as described in Appendix C (dotted lines), and the numerical integration (full lines). We chose $\varepsilon_{TsE0}\simeq42.02$ meV in order to get $n_{sc}=10^{11}$ cm^{-11} and for the 3D trap density we assumed $N_T=10^{18}$ cm^{-3}. The value for N_T seems to be quite high, but within small distance from the oxide-semiconductor interface the number of defects in the oxide is indeed high and only there the traps will be charged (where the states are above the chemical potential) and contribute to scattering. A further limitation of the available trap states to a narrow region besides the interface will follow later on in this work.

The critical behavior of $\rho(T)$ versus n_s for the homogeneous trap density N_T can be seen most clearly on the double logarithmic plots in Figs. 3(a) and 3(c) for n_s in the vicinity of n_c. In Fig. 3(a), the spacing of n_s is linear whereas in Fig. 3(c), the deviations n_s-n_c increase logarithmically. We added the linear plot Fig. 3(b) as it is in this form directly comparable to Fig. 1(b) in the work of AM.[26]

We explain the shape of $\rho(T,n_s)$ starting from a value $n_s>n_{sc}$ (lower curves in Fig. 3). Due to the corresponding large gate voltage, the maximum of the effective trap energy $\varepsilon_T(Z)$ is shifted downward (see Fig. 2) whereas μ is relatively high for the large n_s. The effective trap energy is completely below μ and only a small part of the traps is empty due to thermal excitations [as described by the $p_+(Z)$-peak], leading to small scattering rates and small resistivity (metallic regime). On the contrary for $n_s<n_{sc}$, the effective trap energy $\varepsilon_T(Z)$ is increased, μ has decreased (see Fig. 9 in Appendix B) and so the resistivity ρ increases strongly due to the large number of charged traps.

For increasing temperature T, the chemical potential μ decreases relative to E_0, leading to an increased number of charged trap states n_T^+ which cause an increase in resistivity ρ due to increased dipole trap scattering rate at all densities. The broadening of the $p_+(Z)$-peak due to the increased smearing of the Fermi distribution function by k_BT also contributes to an increase in n_T^+ and thus of ρ.

The saddle-point approximation works best for small temperatures T and densities $n_s>n_{sc}$ where the $p_+(Z)$-peak is very narrow and therefore the Taylor approximation of the trap energy is quite accurate within the peak. For large T and/or $n_s<n_{sc}$, the deviations of the saddle-point approximation from the numerical calculations become visible in Fig. 3 (dashed vs full lines).

As can be seen from Fig. 3, a critical behavior around n_{sc} is visible in that the behavior of the resistivity is qualitatively different for $n_s<n_{sc}$ and $n_s>n_{sc}$ and the curves spread apart around n_c toward lower T. But this does not mean necessarily that ρ increases toward infinity for low n_s and low T. Whether the separate behavior for $n_s<n_{sc}$ leads to an insulating behavior or not depends in the trap model on the T dependence of the chemical potential μ. For case (A) in AM, the slope $d\mu/dt$ is negative and does not lead to the insulating behavior whereas for case (B) $d\mu/dt$ is positive at low T and leads to the insulating behavior. In case (A), the resistivity approaches in the dipole trap model a constant value for $T\rightarrow0$. But the inclusion of quantum interaction in the weak and strong localization regime should nevertheless drive the system into an insulating state (see, e.g., Refs. 3 and 4). Furthermore, there might be other effects, which influence the temperature run of the chemical potential μ like in case (B)—which we do not hold for very realistic—and drive ρ into a strongly insulating state. To our opinion, the important characteristic feature of ρ within the trap model is the qualitative different behavior for $n_s<n_{sc}$ and $n_s>n_{sc}$.

Furthermore, it should be noted that according to the calculations, the resistivity ρ for $T\rightarrow0$ drops toward zero in the metallic regime. This is caused by the fact that electron scattering is taken into account only from the trap states at a

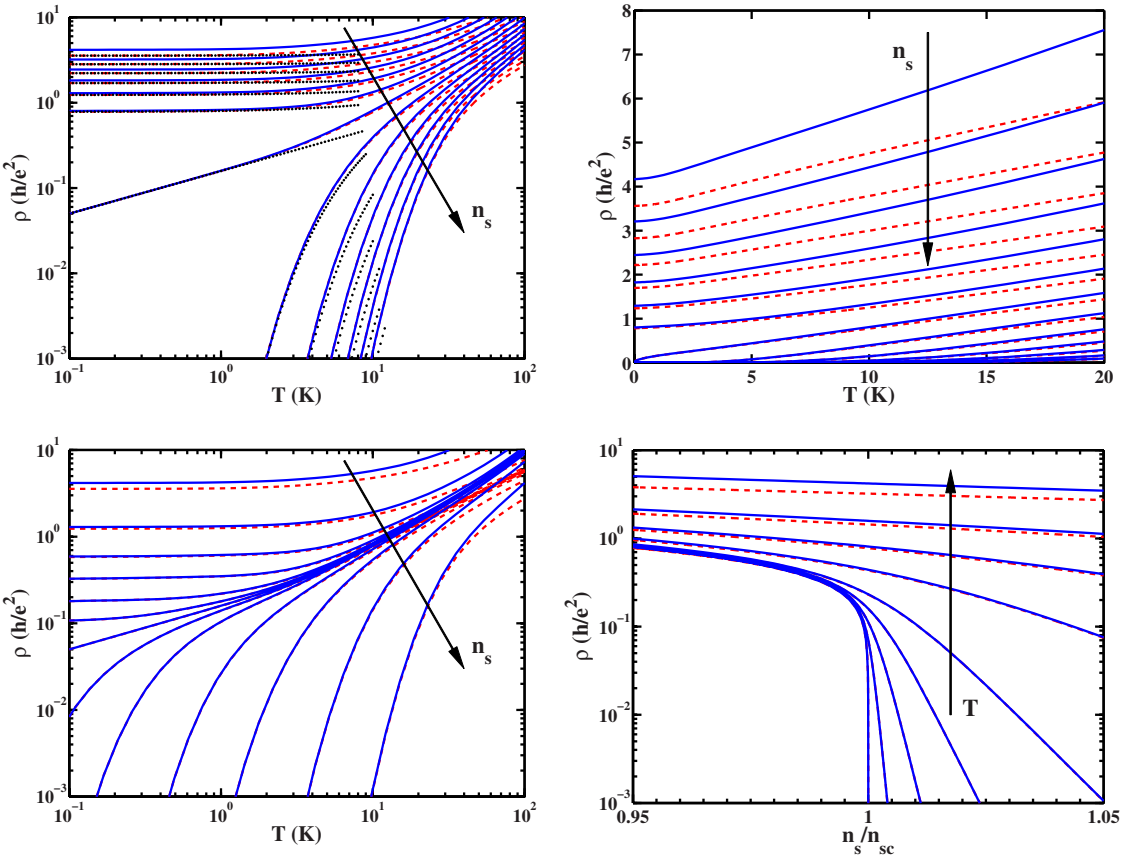

FIG. 3. (Color online) Behavior of the resistivity ρ depending on temperature T and 2D electron density n_s in the inversion layer. Full black (blueonline) numerical integration, dashed gray (red online) saddle-point approximation, doted black lines: analytic low-temperature approximations. In (a)–(c) n_s and in (d) T is constant for the individual curves, all for a critical density of $n_{sc}=10^{11}$ cm^{-2}. (a) logarithmic and (b) linear display of $\rho(T)$ with $n_s=0.70,\ldots,1.30\times n_{sc}$ in steps of $0.05\times n_{sc}$ in top-down order (indicated by arrow) for the individual curves, (c) logarithmic view of $\rho(T)$ with logarithmic spacing of n_s away from n_{sc} with $n_s=[0.70,0.90,0.97,0.99,0.997,0.999,1.00,1.001,1.003,1.01,1.03,1.1,1.3]\times n_{sc}$ in top-down order, (d) $\rho(n_s)$ with parameter $T=0,0.2,0.5,1,2,5,10,20,50$ K in bottom-up order.

single trap energy. If this trap energy is below the Fermi energy, with decreasing T the number of charged scattering centers goes to zero. Only if other scattering effects, such as residual impurities, surface roughness, acceptor states in the depletion layer, etc., are included, the low-T resistivity would be limited. It will be seen later in this work that an energetic broadening of the trap states will have a similar effect in limiting the decrease in ρ, as there remain some charged trap states.

IV. CONDUCTION BAND AS REFERENCE ENERGY

If the bands in the semiconductor and in the oxide are bent due to an applied gate voltage all local states shift with the bands. This holds as well for the trap states E_{Ts}, as they are related to the host material. Their energetic position is fixed relative to the CB edge and not to the electronic ground-state energy E_0 of the inversion layer as assumed by AM. The latter assumption (a7) was necessary in order to get an analytical description of the model.

From Eq. (A1) we see that $E_T-\mu=E_{Ts}+\varepsilon_e(Z)-\mu$ (see Fig. 8) determines the probability p_+ of a trap to be charged. So if we measure E_{Ts} against the CB edge at the interface E_{Cs} we also need to know $\mu_{Cs}=\mu-E_{Cs}$. We find

$$\mu_{Cs}=\mu-E_0+E_0-E_{Cs}=\mu_{E_0}+\varepsilon_{0Cs}, \tag{10}$$

where $\varepsilon_{0Cs}=E_0-E_{Cs}$ is the electronic ground-state energy in the potential of the inversion layer relative to the CB edge at the OS interface.

The chemical potential μ_{E_0} follows from Eq. (B1) whereas an accurate calculation of ε_{0Cs} is rather complex. For simplicity we follow the calculation method of Ando, Fowler, and Stern (AFS),[28] neglect the exchange interaction and correlation effects, and use the Ritz variational principle. (In the mentioned article also more sophisticated methods for the calculation of ε_{0Cs} are given.)

With the Fang-Howard envelope wave function approximation according to AFS,[31] we find an energy eigenvalue ε_{0Cs} between 20 and 60 eV for n_s increasing from 0 to 10^{12} cm^{-2}, respectively, in a nearly linear manner. The increase in ε_{0Cs} with n_s comes mainly from the steeper potential of the inversion layer due to a correspondingly large applied gate voltage. See Appendix D with Fig. 10 for details.

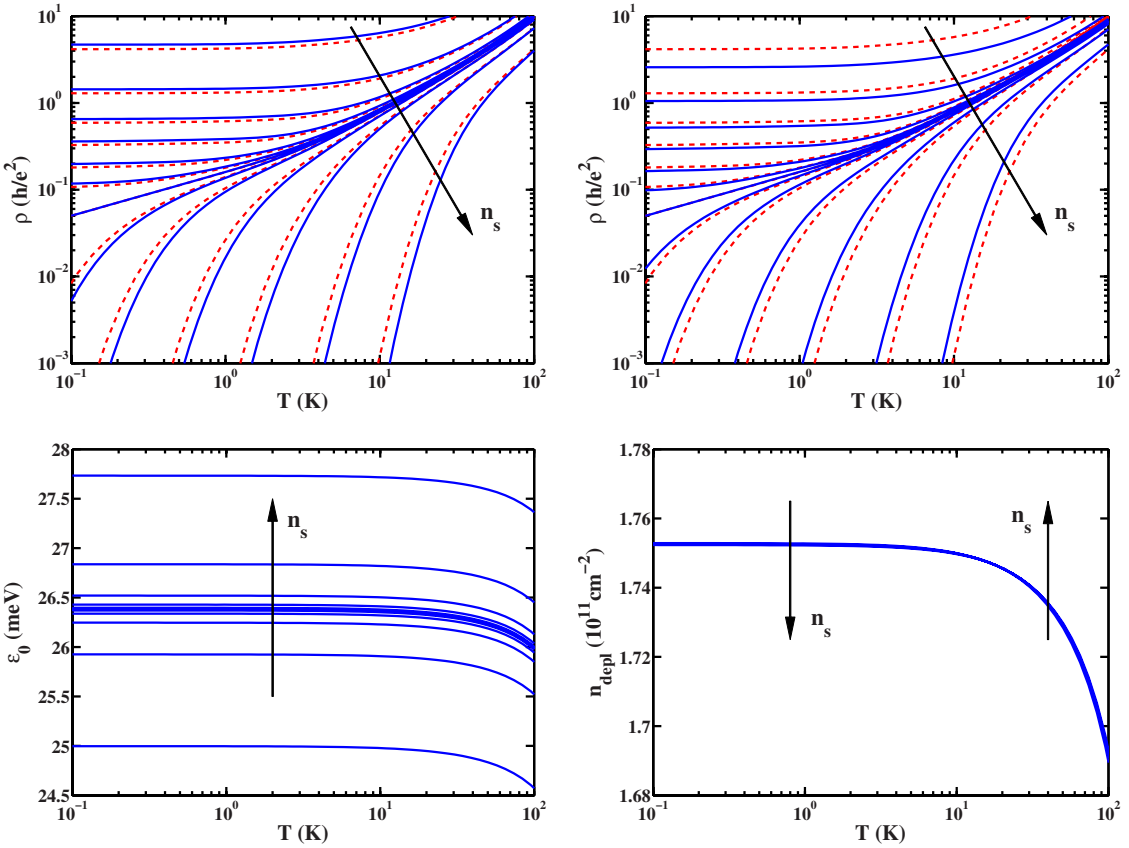

FIG. 4. (Color online) (a) Resistivity $\rho(T,n_s)$ for the potential across the oxide layer $V_{\rm ins}\propto n_s$ with $N_T=10^{18}$ cm^{-3} and $\varepsilon_{TsCs}\simeq 68.4$ meV and (b) for $V_{\rm ins}\propto n_s+n_{\rm depl}$ with $N_T=2.98\times 10^{18}$ cm^{-3} and $\varepsilon_{TsCs}\simeq 95.7$ meV. The parameters are chosen to give a critical density of $n_{sc}=10^{11}$ cm^{-2} for both cases. For (a) and (b), the full black (blue online) represent the realistic case where the trap states are pinned relative to the CB-edge (ε_{TsCs}=const) whereas the dashed gray (red online) show trap energies fixed relative to the electronic ground state E_0 in the inversion potential (ε_{TsE0}=const). Individual curves have different n_s with logarithmic spacing away from n_{sc} with $n_s=[0.70,0.90,0.97,0.99,0.997,0.999,1.00,1.001,1.003,1.01,1.03,1.1,1.3]\times n_{sc}$ in top-down order. In (c) the ground-state energy of the inversion layer $\varepsilon_{0Cs}(T,n_s)$ is shown for same n_s values as before but now bottom-up ordering. In (d) the depletion density $n_{\rm depl}(T,n_s)$ is shown but the n_s dependence vanishes more or less within the linewidth (it is different for low and high T as indicated by arrows and vanishes near 11 K).

Now we hold the difference between trap energy and CB edge at the interface $\varepsilon_{TsCs}=E_{Ts}-E_{Cs}=E_{Ts}-\mu+\mu_{E_0}+\varepsilon_{0Cs}$ constant (instead of ε_{TsE0} as before). When n_s is decreased not only μ_{E_0} decreases but also ε_{0Cs} does so due to the flatter potential, thus μ drops off faster against E_{Ts} and the transition is more abrupt. This can be seen in Fig. 4(a) where results for the case that the trap states shift with the CB edge (ε_{TsCs}=const, full lines) or with the electronic ground state (ε_{TsE0}=const, dashed lines) are compared and where the differences between curves are larger for the former case. In both cases the resistivity was calculated numerically according to the integrals Ω_j as defined in Appendix A. The critical curves $n_s=n_{sc}$ coincide at $T\to 0$ because the values $\varepsilon_{TsE0}\simeq 42.0$ meV and $\varepsilon_{TsCs}\simeq 68.4$ meV were chosen in order to get the same n_{sc} and for higher temperatures because $\varepsilon_{0Cs}(T)$ is nearly constant over a wide temperature range [Fig. 4(c)]. For the 3D trap density, we assumed again $N_T=10^{18}$ cm^{-3}.

It has to be emphasized that we still use $V_{\rm ins}\propto n_s$ like AM. But the 2D density of positive charges in the depletion layer $n_{\rm depl}$ is a by-product of the calculation of ε_{0Cs} and it is no longer a problem to use $V_{\rm ins}\propto n_s+n_{\rm depl}$ instead, i.e., to include the charge in the depletion layer as well. We will do so henceforward. Due to the additional charge $n_{\rm depl}$, the slope of the line $E_{Ts}-eV_{\rm ins}Z/D$ in Fig. 2 is increased, the maximum of $E_T(Z)$ falls off against μ and the $p_+(Z)$-peak gets smaller and narrower. But it is still possible to let coincide the critical curves $\rho(n_s=n_{sc})$ by increasing the trap density N_T in order to compensate the narrower $p_+(Z)$-peak and by increasing ε_{TsCs} in order to get the same critical value n_{sc}. Here it is also important that the 2D charge carrier density of the depletion layer $n_{\rm depl}(T)$ has almost no n_s dependence for low temperatures as can be seen in Fig. 4(d). The resistivity $\rho(T,n_s)$, with $N_T=2.98\times 10^{18}$ cm^{-3} and $\varepsilon_{TsCs}\simeq 95.7$ meV, for which we get the same n_{sc} as before, is shown in Fig. 4(b). The before used depletion density $n_{\rm depl}$ was calculated under the assumption of a background doping density of $N_A=2\times 10^{15}$ cm^{-3}, which is a typical value for high-mobility Si-MOS samples.[32] The values for N_T and ε_{TsCs} were chosen in order to get the requested n_{sc} for the given N_A. In reality the value ε_{TsCs} is determined by the chemical nature of the defect and thus the critical density may change from sample to sample if the background doping density N_A is different.

If now the slope of the energy $E_{Ts}-eV_{\mathrm{ins}}Z/D$ is higher due to the inclusion of n_{depl}, the variable $\max[E_T(Z)-\mu]$ which is crucial for the resistivity is less sensitive on V_{ins} and n_s, and therefore all together the transition is less abrupt as compared to the assumption of AM with ε_{TsE0}=const and neglection of the depletion layer charge, as can be seen in Fig. 4(d) in the comparison of solid and dashed lines.

V. SPATIAL TRAP PROFILE

At higher temperatures, the large value of k_BT leads also to charged trap states in regions where the trap energy E_T is even below the chemical potential [i.e., the $p_+(Z)$-peak broadens] and therefore the resistivity increases to quite high values (see curves at higher temperature in Fig. 4), which are not observed in experiments. But the very high density of trap states should exist only within a narrow strained region of the oxide,[27] and the broadening of the peak beyond the width of this region leads to an unrealistic description. We can resolve this problem by introducing a spatially limited trap density profile $N_T(Z)$.

If now the trap density N_T is a function of Z it has to be inside the integral Ω_j [compare Eq. (A8)],

$$\Omega_j \equiv \int_0^D N_T^+(Z)Z^j dZ = \int_0^D N_T(Z)p_+(Z)Z^j dZ. \qquad (11)$$

For simplicity we use here a rectangular spatial trap profile from the OS interface to an arbitrary depth $Z_{\max}$,

$$N_T(Z) = \begin{cases} \dfrac{n_T}{Z_{\max}} & \text{for } 0 \le Z \le Z_{\max} \\ 0 & \text{for } Z > Z_{\max}, \end{cases} \qquad (12)$$

where n_T is introduced as a 2D trap density. Figure 5(a) shows $\rho(T,n_s)$ for $Z_{\max}$=4 nm, the CB edge E_{Cs} was used as reference energy for the trap states with $\varepsilon_{TsCs}\simeq 95.7$ meV in order to get a critical density of $n_{sc}=10^{11}$ cm^{-2}. Where the 3D trap density N_T does not vanish its value is assumed to be 2.98×10^{18} cm^{-3} as before [like for Fig. 4(b)], resulting in $n_T=N_T\cdot Z_{\max}=1.19\times10^{12}$ cm^{-3} of which again only a part is charged.

As can be seen in Fig. 5(a), the behavior for low temperatures has hardly changed but for high temperatures $\rho(T,n_s)$ now saturates as a broadening of the $p_+(Z)$-peak beyond $Z_{\max}$ does not lead to a further increase in the number of charged scattering centers. This is due to the limited availability of trap states.

Figure 5(b) shows the averaged (effective) trap distance $\bar{Z}$ [see Eq. (A6)] which enters into σ_t and strongly influences ρ. Except for the lowest density, $\bar{Z}$ coincides for the 4 nm limited (full lines) and unlimited (dashed) trap profile at low T. Only at high T, the curves deviate strongly, as in the unlimited trap density also trap states deep inside the oxide get thermally ionized with increasing k_BT whereas this is not possible for the limited trap profile. In addition the trap states close to the oxide-semiconductor interface, which lie at low effective energies due to the strong Coulomb interaction, become also ionized at high T, which turns the effective distance $\bar{Z}$ to lower values, especially for the latter case where no additional charged traps exist further than $Z_{\max}$. For high T a value of 2 nm is approached finally, which is just the center of the limited trap profile.

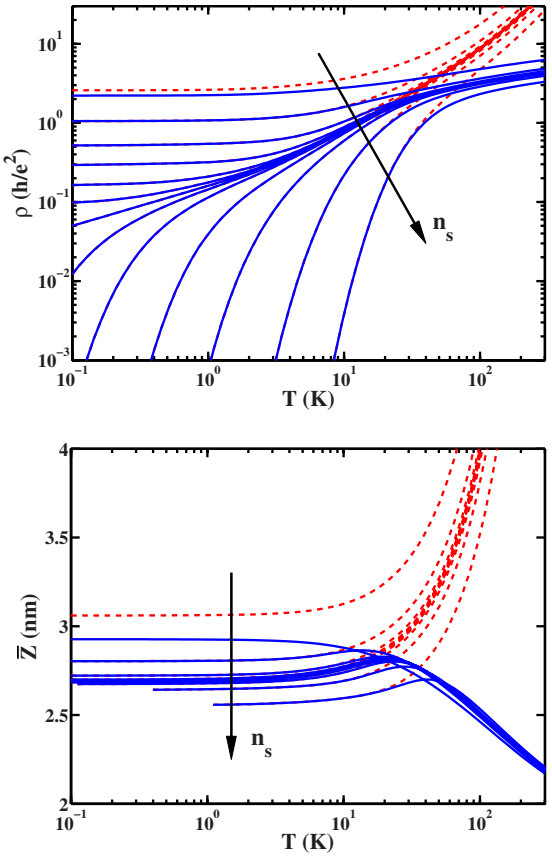

FIG. 5. (Color online) (a) Resistivity $\rho(T,n_s)$ and (b) effective trap distance $\bar{Z}$ for rectangular spatial trap profile with $Z_{\max}$ =4 nm (solid blue/dark lines) and unlimited trap depth (dashed red/gray lines) with curves of constant n_s. The critical density is $n_{sc}=10^{11}$ cm^{-2} which means that ε_{TsCs} has to be 95.7 meV. Densities for individual lines are n_s=[0.70,0.90,0.97,0.99,0.997, 0.999,1.00,1.001,1.003,1.01,1.03,1.1,1.3]$\times n_{sc}$ with logarithmic spacing of n_s away from n_{sc} in top-down order and trap density of $N_T=n_T/Z_{\max}=2.98\times10^{18}$ cm^{-3}. In (b) the $\bar{Z}$ values for high density and low T are not shown, as the numerical integration procedure could not find the very small and narrow occupation peak of $p_+(Z)$.

This saturation of $\rho(T,n_s)$ at high temperatures due to the spatially limited availability of trap states is in fairly good agreement with experiments, where ρ is limited as well.

VI. BROADENING OF THE TRAP ENERGY

As the trap states in the distorted oxide layer will not all be identical and in the stochastic position distribution they will influence each other, their energetic position will be broadened.

We describe the broadening ΔE_T with the help of a normalized distribution function $g(\tilde{E}_{Ts},E_{Ts},\Delta E_T)$ for the trap energy E_{Ts} which characterizes the trap. Now E_{Ts} has the meaning of a mean value. (Mean value should not be understood

in a strict mathematical sense, e.g., for the Lorentz distribution the mean value does not exist but in this case it is obvious to take the energy E_{Ts} where the symmetric distribution has its maximum.) Furthermore $\tilde{E}_{Ts}$ should be the value for a particular trap. The probability of $\tilde{E}_{Ts}$ to lie within the interval $[E, E+dE]$ is given by $g(E, E_{Ts}, \Delta E_T)dE$. Therefore, we replace the probability p_+ of a trap to be charged (see Appendix A) by

$$P_+(Z) = \int_{-\infty}^{\infty} \frac{g(\tilde{E}_{Ts}, E_{Ts}, \Delta E_T)}{\frac{1}{2}\exp\left(-\frac{\tilde{E}_{Ts} - eV_{\text{ins}}\frac{Z}{D} - \varepsilon_D\frac{D}{Z} - \mu}{k_B T}\right) + 1} d\tilde{E}_{Ts}. \tag{13}$$

The denominator is that of p_+, only E_{Ts} is replaced by $\tilde{E}_{Ts}$. By introducing the dimensionless parameters

$$\alpha = \frac{\Delta E_T}{k_B T}, \tag{14}$$

$$\beta = \frac{E_{Ts} - eV_{\text{ins}}\frac{Z}{D} - \varepsilon_D\frac{D}{Z} - \mu}{k_B T} \tag{15}$$

and a dimensionless distribution function $h(\eta)$ defined by

$$\eta = \frac{\tilde{E}_{Ts} - E_{Ts}}{\Delta E_T}, \tag{16}$$

$$g(\tilde{E}_{Ts}, E_{Ts}, \Delta E_T) = \frac{1}{\Delta E_T} h(\eta), \tag{17}$$

the probability P_+ can be written as

$$P_+(\alpha, \beta) = \int_{-\infty}^{\infty} \frac{h(\eta)}{\frac{1}{2}\exp(-\alpha\eta - \beta) + 1} d\eta. \tag{18}$$

As a rule this integral cannot be calculated analytically. An exception from this rule is the uniform distribution. If we define the width of the "rectangle" as $2\Delta E_T$, we get

$$h(\eta) = \begin{cases} \frac{1}{2} & \text{for } -1 < \eta < 1 \\ 0 & \text{elsewhere} \end{cases} \tag{19}$$

and

$$P_+(\alpha, \beta) = \frac{1}{2\alpha}\ln\frac{1 + 2\exp(\beta + \alpha)}{1 + 2\exp(\beta - \alpha)}. \tag{20}$$

We also use the normal distribution

$$h(\eta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\eta^2}{2}\right) \tag{21}$$

with the standard deviation as ΔE_T and the Lorentz distribution (natural line broadening)

$$h(\eta) = \frac{1}{\pi}\frac{1}{\eta^2 + 1} \tag{22}$$

with the half full width at half maximum as ΔE_T.

The numerical results are presented in Fig. 6 for the three different distribution functions and for different width $\Delta E_T = 0.03$, 0.3 and 3 meV. Again we took the CB edge at the interface E_{Cs} as reference energy for the trap states, assumed that traps exist only in the oxide within 4 nm from the OS interface (with constant trap density in this region as in Sec. V), and used $V_{\text{ins}} \propto n_s + n_{\text{depl}}$ instead of $V_{\text{ins}} \propto n_s$. We chose $\varepsilon_{TsCs} \simeq 95.7$ meV in order to get $n_{sc} = 10^{11}$ cm^{-2} and for the 3D trap density $N_T = n_T / Z_{\max} = 10^{18}$ cm^{-3}.

With increasing ΔE_T (in each column of Fig. 6 from top to bottom) the transition becomes smeared out partly or completely, depending on the broadening width and on the distribution function. In the metallic regime, the mean trap energy is below the chemical potential. As the normal and the Lorentz distribution have tails, there always remain some charged traps from the upper tail when otherwise all traps would be filled and thus be neutral. On a logarithmic resistivity scale the $\rho(T)$ behavior is changed drastically by the small amount of additional charged traps and especially for the Lorentz distribution the transition between the two resistivity regions gradually vanishes. This is in agreement with samples of different quality, possessing different distribution width, where the transition is more or less pronounced.

In the insulating regime, the mean trap energy is above the chemical potential. Here a large part of the traps is always charged and in contrast the few uncharged traps due to the lower tail of the energy distribution do not change the behavior substantially.

By means of further analytical considerations, we got also an estimate for the temperature T_b below which the resistivity becomes almost constant. In general the resistivity ρ varies by some orders of magnitude, so as a criterion for being almost constant we took that region where ρ changes finally by only a factor 2 down to zero temperature. The markers in Fig. 6 represent these temperatures T_b.

According to this definition, for the insulating behavior $n_s < n_{sc}$ for all three distributions we get

$$T_b = \frac{\max E_T(Z) - E_F}{(4 - \ln 2)k_B} \tag{23}$$

and for the metallic behavior $n_s > n_{sc}$

$$T_b = \begin{cases} \dfrac{\Delta E_T^2}{(4 + \ln 2)k_B[E_F - \max E_T(Z)]} & \text{normal distribution} \\ \dfrac{E_F - \max E_T(Z)}{2(4 + \ln 2)k_B} & \text{Lorentz distribution.} \end{cases} \tag{24}$$

The uniform distribution has no tails so in the metallic regime there is no temperature range where ρ is almost constant.

Figure 6 shows that the sharpness of the transition from the strongly metallic to the insulating regime depends strongly on the size and type of trap energy broadening. But

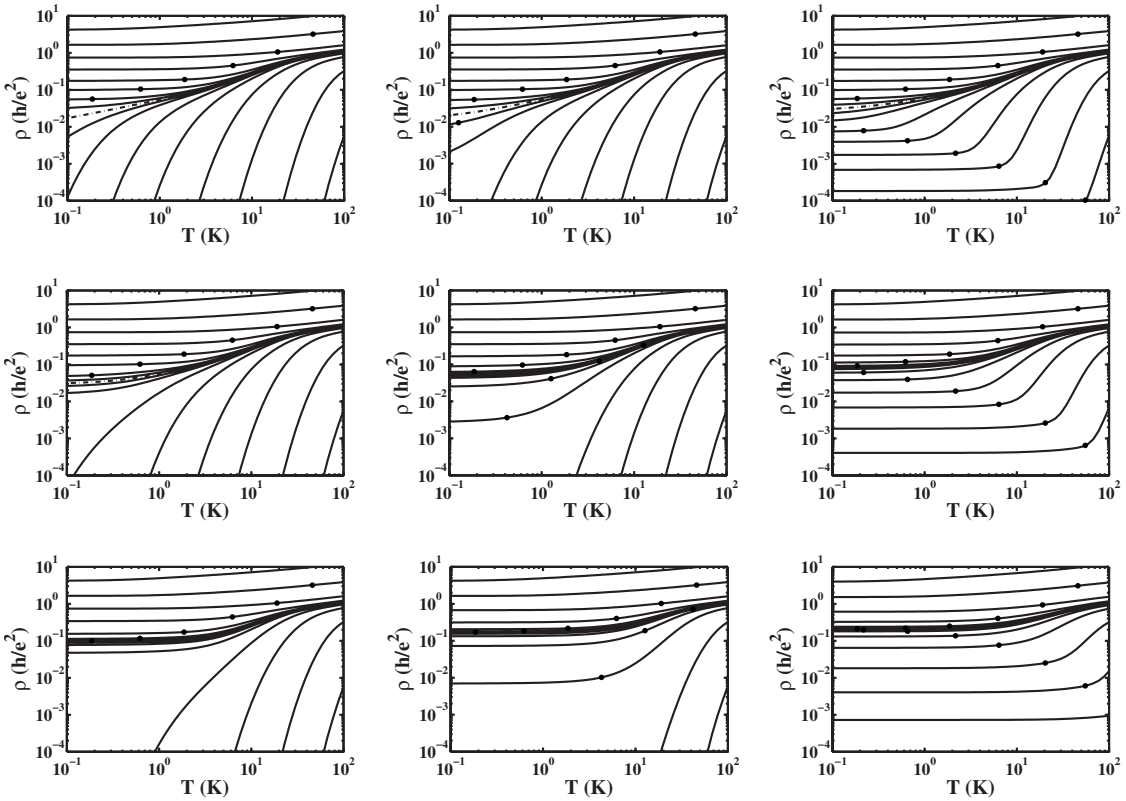

FIG. 6. Resistivity $\rho(T,n_s)$ for different types of trap energy broadening functions and different width ΔE_T. The first picture column shows uniform distribution with energy width of $\Delta E_T=0.03,0.3,3$ meV top-down, the second column a normal distribution, and the third a Lorentz distribution for same energy widths. For individual curves n_s is constant. The critical density $n_{sc}=10^{11}$ cm^{-2} for $\varepsilon_{TsCs}\simeq 95.7$ meV is indicated by the dashed line in each figure. The lines above n_{sc} show $n_s=0.10,0.30,0.70,0.90,0.97,0.99,0.997,0.999\times n_{sc}$, the lines below represent $n_s=1.001,1.003,1.01,1.03,1.1,1.3,2.0,4.0,10\times n_{sc}$ top-down, for a 3D trap density of $N_T=n_T/Z_{\max}=10^{18}$ cm^{-3}.

a wide distribution of trap energies does not mean that the disorder in the electronic system is larger. It just means that the charging of the trap states, which leads to scattering, is distributed over a wider gate voltage and temperature range. When finally all traps states are charged, the scattering potential at the electron position in the inversion layer is the same as if the trap energies would not be distributed over some energy range. The broadening should have no direct influence on additional quantum corrections which lead to Anderson localization at low temperatures.

VII. CONCLUSIONS

In this work, we have performed numerical calculations within the dipole trap model for Si-MOS structures. Originally this model was proposed by Altshuler and Maslov with several approximations, in order to get analytical solutions. Due to our numerical treatment, we could eliminate several approximations. We describe the potential inside the insulator by its detailed spatial dependence instead of the parabolic saddle-point approximation, we fix the trap state energy relative to the conduction-band edge instead to the electronic ground state inside the 2D layer and we have taken into account the detailed change in the chemical potential in the two-dimensional electron layer with respect to the bulk material, which seems to be more realistic than the two cases in the original treatment.

According to our calculations, the metallic regime at high electron densities n_s where the resistivity is decreasing toward lower temperature, is strongly developed. Also a critical density n_{sc} can be identified with a characteristic temperature dependence in between the two different regimes. For electron densities $n_s<n_{sc}$, the resistivity curves saturate toward low temperatures and approach a constant value. No insulting behavior occurs in the sense that ρ increases strongly toward low temperature for the taken temperature dependence of the chemical potential. Such an increase could be caused by a different temperature dependence of the chemical potential μ caused by additional influences.[33] Also the quantum corrections in the weak and strong localization regime—which are neglected here—would increase the resistance ρ at low electron densities and low temperatures. The most important fact is that also in our treatment we get clearly two distinctly different regimes.

In addition, we have generalized the dipole trap model by dropping the assumptions that the trap states are homogeneously distributed inside the oxide layer and that the energy distribution is δ-like. A narrow spatial distribution of the trap states near the oxide-semiconductor interface limits the number of charged states at high temperatures and thus gives an upper limit for the increase in the resistivity ρ as well. This

leads to a good agreement with experimental observations at higher temperatures. The energetic broadening of the trap states on the other hand leads to a finite amount of unoccupied and thus charged states in cases where otherwise all states would lie below the chemical potential μ and the number of charged trap states would go to zero if k_BT approaches zero. Thus for high electron densities with metallic behavior the resistivity will not infinitely decrease toward lower temperature, but saturate at a finite value, in agreement with experiments on Si-MOS structures. Depending on the size and on the form of the energetic broadening, the transition between the two regimes will be more or less smeared out, explaining the difference between samples with different quality.

As shown by Althuler and Maslov, the effect of a magnetic field could be taken into account by the Zeeman splitting of the trap states with spin $\pm 1/2$ which turns the metallic behavior into an insulating one. We did not include magnetic field effects in our calculations but an according energetic shift of the trap states has to lead to very similar effects in our refined model as well.

We also like to mention that for low electron densities care has to be taken for the dipole scattering model. It is assumed that the electrons in the two-dimensional layer shield the potential of the charged trap states and thus form a dipole field which is responsible for the scattering. At very low electron densities, this screening becomes weaker and the scattering will finally increase so that the resistivity should be higher in this regime. These effects have not been taken into account in the frame of the current work, as we like to present the main effects due to charging of trap states.

Altogether, our detailed numerical calculations within the dipole trap model show that a pronounced metallic state can be caused by trap states at an appropriate energy level inside the oxide of Si-MOS structures. At low electron densities, a different regime is observed, for which the resistivity saturates at low temperatures if the temperature slope of the chemical potential is negative or strongly increases for a positive slope. For the realistic assumptions of energetic broadening and narrow spatial distribution near the oxide-semiconductor interface, the behavior is in close agreement with experimental observations.

ACKNOWLEDGMENTS

Work was supported by the Austrian Science Foundation (FWF) under Project No. P16160. We thank our former colleague A. Prinz to perform first calculations within the Altshuler-Maslov trap model and D. L. Maslov for several hints.

APPENDIX A: TRAP MODEL DETAILS

We give here some details of the AM trap model together with our improvements of the calculations. In the absence of an external magnetic field, the number of positively charged trap states can be calculated with a modified Fermi-Dirac occupation function by

$$p_+(Z) = \frac{1}{\frac{1}{2}\exp\left[-\frac{E_T(Z)-\mu}{k_BT}\right]+1}, \tag{A1}$$

where the minus sign in the exponent comes from the fact that we count empty (positively charged) states and the factor 1/2 from the two possible spin orientation if the trap is empty. p_+ is determined by $E_T(Z)-\mu$ (as indicated by the vertical lines in Fig. 2) and the temperature.

AM assume that a positive charged trap is screened by the electrons in the 2DEG and the trap forms together with that image charge a dipole. For the transport scattering cross section σ_t of this dipole they found classically

$$\sigma_t(\varepsilon, Z) = c_\sigma \varepsilon^{-1/3} Z^{2/3}, \tag{A2}$$

where $c_\sigma = 2.74(e^2/8\pi\epsilon^*\epsilon_0)^{1/3}$ and Z is the distance between the trap and the oxide semiconductor interface, ε is the (kinetic) energy of the scattered electron relative to the ground-state energy of the inversion layer E_0,

$$\varepsilon = E - E_0, \tag{A3}$$

and ϵ^* is an effective dielectric constant $\epsilon^* = (\epsilon_{\text{ins}} + \epsilon_{\text{sc}})/2$ with ϵ_{sc} the dielectric constant of the semiconductor.

The Drude formula together with the Boltzmann equation in relaxation-time approximation yields the resistivity $d\rho(Z)$ caused by the charged traps within the layer $[Z, Z+dZ]$. By integrating these contributions over the whole oxide $[0, D]$ one is able to express ρ in terms of an effective electron energy $\bar{\varepsilon}$ as used by AM (see Ref. 26). Furthermore $\varepsilon_{FE0} = E_F - E_0$ is the Fermi energy, $\mu_{E_0} = \mu - E_0$ is the chemical potential, each relative to the ground state energy of the inversion layer, and N_T is the 3D density of traps.

The 3D density of charged traps $N_T^+(Z)$ it is given by

$$N_T^+(Z) = N_T(Z) p_+(Z). \tag{A4}$$

The dipole cross section σ_t can be expressed with the help of an effective averaged energy,[26]

$$\bar{\varepsilon} = \varepsilon_{FE0}\left[\int_0^\infty \frac{1}{4k_BT}\left(\frac{\varepsilon}{\varepsilon_{FE0}}\right)^{5/6} \cosh^{-2}\left(\frac{\varepsilon - \mu_{E_0}}{2k_BT}\right) d\varepsilon\right]^{-6}, \tag{A5}$$

and the effective distance $\bar{Z}$ defined by

$$\bar{Z}^{2/3} = \frac{\int_0^D N_T^+(Z) Z^{2/3} dZ}{\int_0^D N_T^+(Z) dZ} = \frac{\int_0^D N_T^+(Z) Z^{2/3} dZ}{n_T^+} \tag{A6}$$

as

$$\sigma_t(\bar{\varepsilon}, \bar{Z}) = c_\sigma \bar{\varepsilon}^{-1/3} \bar{Z}^{2/3}. \tag{A7}$$

Here n_T^+ is the 2D density of charged traps.

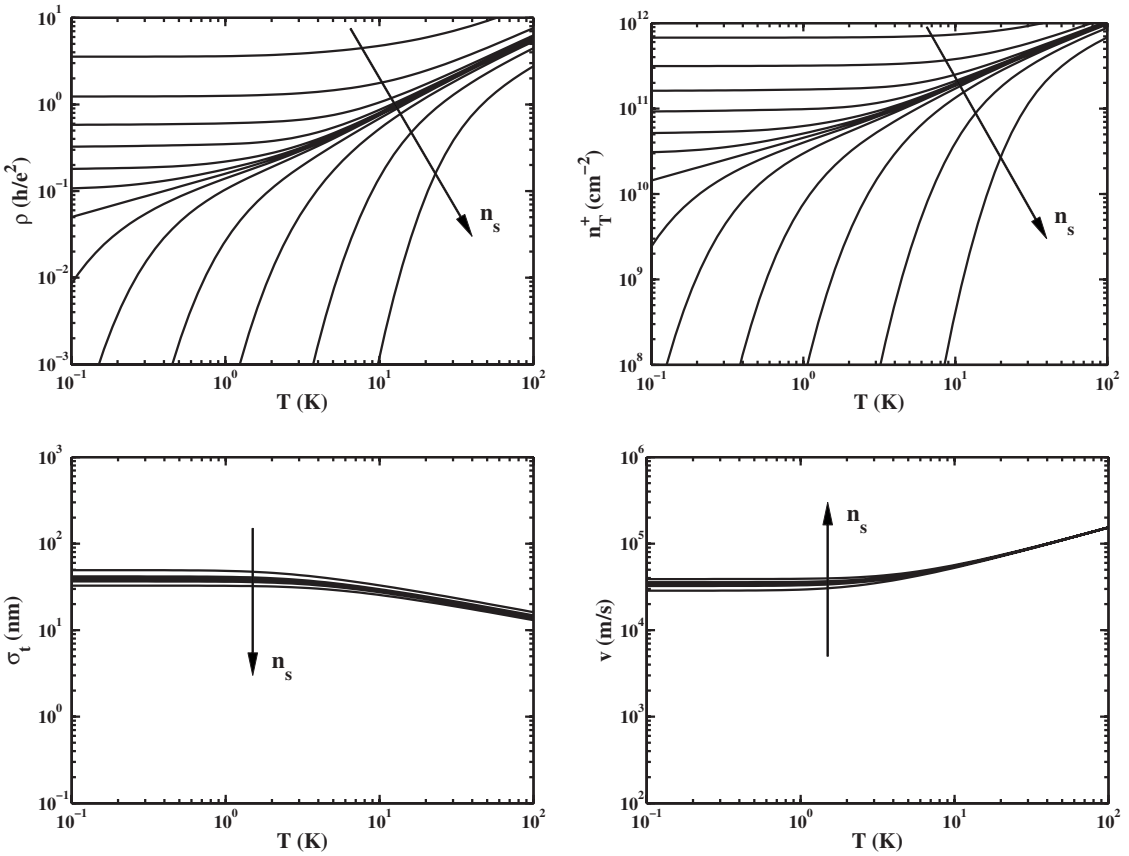

FIG. 7. (a) Temperature dependence of the resistivity $\rho \propto \sigma_t(\bar{\varepsilon},\bar{Z}) n_T^+ v(\bar{\varepsilon})$ and the corresponding contributions of (b) 2D density of charged traps n_T^+, (c) effective scattering cross section $\sigma_t(\bar{\varepsilon},\bar{Z})$, and (d) effective electron velocity $v(\bar{\varepsilon})$ for electron densities with logarithmic spacing of n_s away from n_{sc} with $n_s = [0.70, 0.90, 0.97, 0.99, 0.997, 0.999, 1.00, 1.001, 1.003, 1.01, 1.03, 1.1, 1.3] \times n_{sc}$ in top-down order for curves (a), (b), and (c) and bottom-up for (d). All subfigures plotted over same number of magnitudes to illustrate relative importance in contribution to ρ.

The integral in the numerator and that in the denominator of $\bar{Z}^{2/3}$ can be treated in quite the same way, so we define

$$\Omega_j \equiv \int_0^D N_T^+(Z) Z^j dZ = N_T \int_0^D \frac{Z^j dZ}{\frac{1}{2}\exp\left[-\frac{E_T(Z)-\mu}{k_B T}\right]+1}. \tag{A8}$$

In the last step, we followed AM and assumed that the trap density is constant within the oxide, respectively, in the region where $p_+(Z)$ does not vanish. Now we can write

$$n_T^+ = \Omega_0, \tag{A9}$$

$$\sigma_t(\bar{\varepsilon},\bar{Z}) = c_\sigma \bar{\varepsilon}^{-1/3} \bar{Z}^{2/3} = c_\sigma \bar{\varepsilon}^{-1/3} \frac{\Omega_{2/3}}{\Omega_0}. \tag{A10}$$

In order to calculate the resistivity

$$\rho = m_c \sigma_t(\bar{\varepsilon},\bar{Z}) n_T^+ v(\bar{\varepsilon})/n_s e^2 \tag{A11}$$

[as in Eq. (8)] the knowledge of n_T^+ is not necessary, as n_T^+ cancels out with the denominator of $\bar{Z}^{2/3}$ within $\sigma_t(\bar{\varepsilon},\bar{Z})$ and indeed without defining the integrals Ω_j explicitly, AM used an equation corresponding to

$$\rho = \frac{\sqrt{2m_c} c_\sigma \bar{\varepsilon}^{1/6} \Omega_{2/3}}{n_s e^2}, \tag{A12}$$

where m_c is the conductivity mass of the free electrons within the inversion layer.

Our main interest holds for the dependence of the metal-insulator transition on the electron density n_s, i.e., the temperature behavior of ρ as a function of n_s. In this context n_T^+ is very useful in order to see that it contributes the main variations to the resistivity $\rho(n_s,T)$ whereas $\sigma_t(\bar{\varepsilon},\bar{Z})$ and $v(\bar{\varepsilon})$ show only weak dependence on n_s and T, as illustrated in Fig. 7. The benefit of Eq. (A11) in comparison to the form of Eq. (A12) used by AM is, that the physical meaning of the terms becomes immediately clear.

It might appear strange that on the one hand side n_T^+ provides the main contribution to ρ, on the other hand it cancels out in another representation of ρ. This is resolved as the power of the effective mean distance $\bar{Z}^{2/3}$ is expressed by the ratio $\Omega_{2/3}/\Omega_0$ and is a factor inside σ_t showing only small variation with n_s. But $n_T^+ = \Omega_0$ can be canceled in the product

$\sigma_t \cdot n_T^+ = \Omega_{2/3}/\Omega_0 \cdot \Omega_0 = \Omega_{2/3}$ and what remains is just $\Omega_{2/3}$. So the calculation of n_T^+ is then not necessary but the strong n_s dependence remains within the term $\Omega_{2/3}$, whose meaning does not correspond to the usual contributions to and interpretations of $\rho(n_s,T)$.

In order to get an analytical solution for the integrals which corresponds to $\Omega_{2/3}$, AM expanded the electrostatic energy $\varepsilon_T(Z)$ into a Taylor series about the point Z_m where it reaches its maximum ε_m. This procedure is called saddle-point approximation.

We were able to bring the integrals into the form[34]

$$\Omega_j = N_T Z_m^{j+3/2} \sqrt{\frac{k_B T}{\varepsilon_D D}} \times \underbrace{\sqrt{\pi}}_{\Gamma\left(\frac{1}{2}\right)} \mathcal{F}_{-1/2}\left(\ln 2 + \frac{\varepsilon_{TsE0} + \varepsilon_m - \mu_{E_0}}{k_B T}\right), \tag{A13}$$

which are now expressed by Fermi-Dirac integrals[35]

$$\mathcal{F}_k(\eta) = \frac{1}{\Gamma(k+1)} \int_0^\infty \frac{\mathcal{Z}^k d\mathcal{Z}}{\exp(\mathcal{Z}-\eta)+1}, \tag{A14}$$

where Γ is the gamma function.

On the right-hand side of Eq. (A13), j appears only in the exponent of Z_m so within the saddle-point approximation Eq. (A6) simplifies to

$$\bar{Z}^{2/3} = \frac{\Omega_{2/3}}{\Omega_0} \simeq Z_m^{2/3}. \tag{A15}$$

APPENDIX B: CHEMICAL POTENTIAL

AM described two scenarios for the temperature behavior of the chemical potential: (A) the chemical potential of the 2DEG and of the Si substrate coincide. (B) The 2DEG is disconnected form the substrate. For the case (A) they assumed that the temperature behavior in the 2DEG is the same as in the bulk. However, they did not take into account that the chemical potential in the 2DEG is measured against the ground-state energy E_0 and in the bulk against the conduction- or valence-band edge, i.e., they assumed E_0 and the band bending to be fixed.

For (B) AM used an equation analogous to

$$\mu_{E_0} = k_B T \ln\left[\exp\left(\frac{\varepsilon_{FE0}}{k_B T}\right) - 1\right]. \tag{B1}$$

If only one subband of the inversion layer is occupied (quantum limit), the Fermi energy relative to its ground-state energy is given by[28]

$$\varepsilon_{FE0} = \frac{2\pi\hbar^2 n_s}{g_s g_{v2D} m_{d2D}}, \tag{B2}$$

where g_s, g_{v2D}, and m_{d2D} are the spin degeneracy, the valley degeneracy (for the 2DEG), and the density-of-states mass (2D), respectively. The two equations above can be derived from

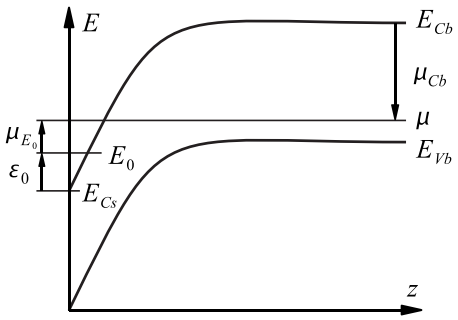

FIG. 8. Band bending and notation on semiconductor side of the OS interface for thermal equilibrium between 2D layer and bulk. The electronic ground-state energy of the inversion layer E_0, the CB edge E_C, its values at the interface E_{Cs} and in the bulk E_{Cb}, the valence-band edge E_V, and the chemical potential μ are shown schematically, also the ground-state energy relative to the CB edge ε_{0Cs}, the chemical potential relative to CB edge in the bulk μ_{Cb} and relative to the ground-state energy μ_{E_0} are noted.

$$n_s = \underbrace{\frac{g_s g_{v2D} m_{d2D}}{2\pi\hbar^2}}_{\text{2D density of states}} \int_{E_0}^\infty \underbrace{\frac{1}{\exp\left(\frac{E-\mu}{k_B T}\right)+1}}_{\text{Fermi-Dirac distribution}} dE. \tag{B3}$$

Our assumptions are shown schematically in Fig. 8 (semiconductor side of the OS interface) and are as follows. In thermal equilibrium, there is a single chemical potential μ throughout the structure. For a certain temperature T, μ is determined in the bulk by the (residual) doping density (giving $\mu_{Cb} = \mu - E_{Cb}$, details are shown elsewhere).[34] In the inversion layer the position of μ relative to E_0 (i.e., $\mu_{\varepsilon_{0Cs}}$) follows directly from the 2D density n_s.

The band bending $E_{Cb} - E_{Cs}$ adjusts so that

$$E_{Cb} - E_{Cs} = \varepsilon_{0Cs} + \mu_{E_0} - \mu_{Cb}. \tag{B4}$$

If the band bending increases, the quantum well gets narrower and the ground state energy E_0 increases relative to the CB edge. Thus $\varepsilon_{0Cs} = E_0 - E_{Cs}$ itself is a function of the band bending, and is determined here by a self-consistent calculation (fix point iteration).

At low temperatures, the 2D electron system could indeed be decoupled from the bulk substrate without a common chemical potential for thermal equilibrium. For the case that n_s is then determined only by the applied gate voltage, which reflects case (B) of AM. But in nonthermal equilibrium, charges might also be stored in other parts of the system and a detailed description is difficult. We do not consider such a situation further but if n_s is determined solely by the gate voltage, the behavior should be very similar to the outcome of our calculations as the coupling with the depletion layer and the substrate causes just relative small changes in n_s.

For our calculations we assumed a {001} silicon surface plane. The parameters, which were used throughout this work are given in Appendix E. The detailed behavior of the chemical potential $\mu(T,n_s)$ is displayed in Figs. 9(a) and 9(b) and is important for the understanding of the behavior of

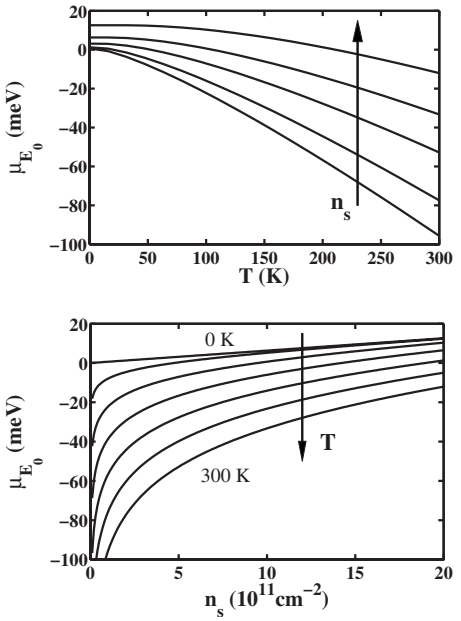

FIG. 9. Chemical potential μ relative to the ground-state energy E_0. (a) Curves with constant n_s, bottom up: $n_s=1,2,5,10,20 \times 10^{11}$ cm^{-2}. (b) Curves for constant T, top down: $T=0,50,100,150,200,250,300$ K.

$\rho(T,n_s)$. For constant n_s μ decreases with increasing T, and obviously increases with increasing n_s for constant T.

APPENDIX C: APPROXIMATION FOR LOW TEMPERATURES

For given values T and n_s, we calculate here the resistivity ρ with help of the Eqs. (A12) and (A13). Beside the explicit temperature dependence also the chemical potential μ_{E_0} and the effective electron energy $\bar{\varepsilon}$ are functions of T.

The integral in Eq. (A5) can be replaced by a Fermi-Dirac integral in that the cosh term is replaced by the first derivative of the Fermi distribution and then an integration by parts is performed. This results in

$$\bar{\varepsilon} = \varepsilon_{FE0}\left(\frac{\varepsilon_{FE0}}{k_B T}\right)^5 \left[\Gamma\left(\frac{11}{6}\right)\mathcal{F}_{-1/6}\left(\frac{\mu_{E_0}}{k_B T}\right)\right]^{-6}. \quad \text{(C1)}$$

For low temperatures, we can take advantage of the behavior of $\mu_{E_0}(T)$ and $\bar{\varepsilon}(T)$ for $T\to 0$. It can easily be shown that for constant n_s and therefore constant ε_{FE0} all derivatives vanish at this point. So $\mu_{E_0}(T)$ and $\bar{\varepsilon}(T)$ are very flat functions at $T\to 0$, for $k_B T \ll \varepsilon_{FE0}$ they can be approximated by $\mu_{E_0}(T)\simeq\varepsilon_{FE0}$ and $\bar{\varepsilon}(T)\simeq\varepsilon_{FE0}$. Further the Fermi-Dirac integral $\mathcal{F}_j(\eta)$ can be approximated[35] either by an expression with η^{j+1} or by $\exp\eta$ for $\eta\gg 0$ or $\eta\ll 0$, respectively.

For $T\to 0$ the argument of the Fermi-Dirac integral in Eq. (A13) increases beyond any border. Which approximation for the Fermi-Dirac integral is applicable depends on the sign of $\varepsilon_{TsE0}+\varepsilon_m-\varepsilon_{FE0}$. (Here μ_{E_0} is replaced by ε_{FE0} as by definition $\mu_{E_0}\to\varepsilon_{FE0}$ for $T\to 0$.) This is conform with AMs definition of the transition point, $(\varepsilon_{TsE0}+\varepsilon_m-\varepsilon_{FE0})/k_B T=0$. Accordingly we define

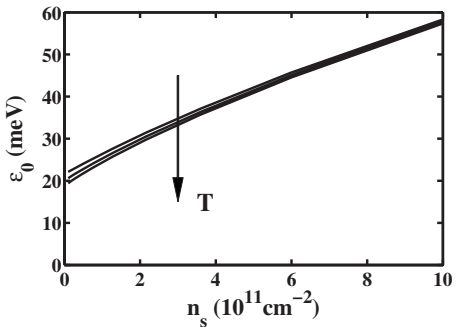

FIG. 10. Ground-state energy of the inversion layer ε_{0Cs} versus the density of electrons in the inversion layer n_s for three temperatures, top-down: $T=0,200,300$ K.

$$\varepsilon_{mF} = \varepsilon_{TsE0} + \varepsilon_m - \varepsilon_{FE0},$$

$$\varepsilon_{mF} > 0 \to \text{insulating},$$

$$\varepsilon_{mF} < 0 \to \text{metallic},$$

$$\varepsilon_{mF} = 0 \to \text{transition point}$$

and a critical density

$$n_{sc} = n_s|_{\varepsilon_{mF}=0}. \quad \text{(C2)}$$

Applying the appropriate approximations results in

$$\Omega_j \simeq \begin{cases} 2N_T Z_m^{j+3/2}\sqrt{\dfrac{k_B T\ln 2}{\varepsilon_D D}+\dfrac{\varepsilon_{mF}}{\varepsilon_D D}} & \text{for } \varepsilon_{mF}>0 \\ 2N_T Z_m^{j+3/2}\sqrt{\dfrac{k_B T\pi}{\varepsilon_D D}\exp\left(\dfrac{\varepsilon_{mF}}{k_B T}\right)} & \text{for } \varepsilon_{mF}<0 \\ \underbrace{\mathcal{F}_{-1/2}(\ln 2)}_{\simeq 0.891} N_T Z_m^{j+3/2}\sqrt{\dfrac{k_B T\pi}{\varepsilon_D D}} & \text{for } \varepsilon_{mF}=0. \end{cases} \quad \text{(C3)}$$

Please note an interesting behavior. When setting $\varepsilon_{mF}=0$ in the first two equations for $T>0$ they converge neither into each other nor into the third one. This apparent discrepancy can be understood, as $|\varepsilon_{mF}|$ gets smaller and smaller the maximum temperature where the approximations for the Fermi-Dirac integral are just applicable also gets smaller and smaller and finally vanishes for $\varepsilon_{mF}=0$. Indeed for $\varepsilon_{mF}=0$ and $T\to 0$ the three cases yield the same result, i.e., $\Omega_j=0$.

APPENDIX D: GROUND-STATE ENERGY OF THE INVERSION LAYER

The ground-state energy of the inversion layer is calculated with help of the Ritz variational principle. For convenience we introduce a new coordinate system $z=-Z$, i.e., the z axis is perpendicular to the OS interface, positive z values correspond with the semiconductor side. For the electrons in the inversion layer, the bent CB of the semiconductor together with the step at the interface builds the quantum well. Figure 10 shows the ground-state energy ε_{0Cs} versus the electron density n_s, where ε_{0Cs} decreases with decreasing n_s.

The details of the calculation are given in the following.

We use the Fang-Howard envelope wave function according to AFS,[31]

$$\varphi(z,b) = \begin{cases} \sqrt{\frac{b^3}{2}} z \exp\left(-\frac{bz}{2}\right) & \text{for } z \geq 0 \\ 0 & \text{for } z < 0 \end{cases} \quad \text{(D1)}$$

with a confining potential $U(z)$ approximated by contributions from the charged acceptors within the depletion layer, the interaction with all other electrons in the inversion layer, and the interaction with the image charges. For more details, see AFS (Ref. 28) and Ref. 34. The resulting ground-state energy is measured against the CB edge at the interface E_{Cs} as requested. With the Ritz variational method, an expression for b can be obtained. AFS neglect terms proportional to $\exp(-bz_d)$, β, and $\delta(b)b^{-3}$. We find a solution, where we neglected only $\delta(b)b^{-3}$, which leads to a slightly improved value for the ground-state energy.[34]

The density n_{depl} is also needed and can be calculated from the total band bending $e\phi_0 = E_{Cb} - E_{Cs}$ (b=bulk, s=surface). To obtain a solution one has to assume that in the depletion layer all acceptors are charged and that in the bulk there is charge neutrality. The boundary between the depletion layer and the bulk is not sharp, this is described by a term $-k_BT$ in the band bending.[36] In order to get the potential one has to solve the Poisson equation with the charge density which corresponds with the test wave function φ (for details see Ref. 34). The values used for the calculation are given in Appendix E.

The ground-state energy ε_{0Cs} itself is a (small) part of the total band bending, this problem is solved by a fix point iteration using $\varepsilon_{0Cs}=0$ as start value.

APPENDIX E: PARAMETERS FOR Si-MOS STRUCTURE

The parameters given in Table I were used for the numerical calculations of the Si-MOS structure at a {001} silicon surface plane throughout the paper.

*gerhard.brunthaler@jku.at

[1] S. V. Kravchenko, G. V. Kravchenko, J. E. Furneaux, V. M. Pudalov, and M. D'Iorio, Phys. Rev. B **50**, 8039 (1994).
[2] S. V. Kravchenko, W. E. Mason, G. E. Bowker, J. E. Furneaux, V. M. Pudalov, and M. D'Iorio, Phys. Rev. B **51**, 7038 (1995).
[3] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).
[4] E. Abrahams, S. V. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. **73**, 251 (2001).
[5] M. Y. Simmons, A. R. Hamilton, M. Pepper, E. H. Linfield, P. D. Rose, D. A. Ritchie, A. K. Savchenko, and T. G. Griffiths, Phys. Rev. Lett. **80**, 1292 (1998).
[6] M. P. Lilly, J. L. Reno, J. A. Simmons, I. B. Spielman, J. P. Eisenstein, L. N. Pfeiffer, K. W. West, E. H. Hwang, and S. Das Sarma, Phys. Rev. Lett. **90**, 056806 (2003).
[7] V. Senz, T. Ihn, T. Heinzel, K. Ensslin, G. Dehlinger, D. Grützmacher, U. Gennser, E. H. Hwang, and S. D. Sarma, Physica E **13**, 723 (2002).
[8] S. J. Papadakis and M. Shayegan, Phys. Rev. B **57**, R15068 (1998).
[9] F. Stern, Phys. Rev. Lett. **44**, 1469 (1980).
[10] A. Gold and V. T. Dolgopolov, Phys. Rev. B **33**, 1076 (1986).
[11] S. Das Sarma, Phys. Rev. B **33**, 5401 (1986).
[12] A. Gold, J. Phys.: Condens. Matter **15**, 217 (2003).
[13] S. Das Sarma, M. P. Lilly, E. H. Hwang, L. N. Pfeiffer, K. W. West, and J. L. Reno, Phys. Rev. Lett. **94**, 136401 (2005).
[14] S. Das Sarma and E. Hwang, Solid State Commun. **135**, 579 (2005).
[15] A. M. Finkel'stein, Z. Phys. B: Condens. Matter **56**, 189 (1984).
[16] C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. B **30**, 527 (1984).
[17] A. Punnoose and A. M. Finkel'stein, Phys. Rev. Lett. **88**, 016802 (2001).
[18] A. Punnoose and A. M. Finkel'stein, Science **310**, 289 (2005).
[19] G. Zala, B. N. Narozhny, and I. L. Aleiner, Phys. Rev. B **64**, 214204 (2001).
[20] I. V. Gornyi and A. D. Mirlin, Phys. Rev. B **69**, 045313 (2004).
[21] S. V. Kravchenko and M. P. Sarachik, Rep. Prog. Phys. **67**, 1 (2004).
[22] A. A. Shashkin, Phys. Usp. **48**, 129 (2005).
[23] V. T. Dolgopolov, Low Temp. Phys. **33**, 98 (2007).
[24] F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).
[25] W. R. Clarke, C. E. Yasin, A. R. Hamilton, A. P. Micolich, M. Y. Simmons, K. Muraki, Y. Hirayama, M. Pepper, and D. A. Ritchie, Nat. Phys. **4**, 55 (2008).
[26] B. L. Altshuler and D. L. Maslov, Phys. Rev. Lett. **82**, 145 (1999).
[27] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley-Interscience, New York, 1981).
[28] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[29] T. Hori, *Gate Dielectrics and MOS ULSIs* (Springer-Verlag, Berlin, 1997).
[30] T. M. Klapwijk and S. Das Sarma, Solid State Commun. **110**, 581 (1999).
[31] F. F. Fang and W. E. Howard, Phys. Rev. Lett. **16**, 797 (1966).
[32] V. M. Pudalov (private communication).
[33] The insulting regime in the work of AM was obtained by a decreasing chemical potential toward lower temperatures as a result of more or less unrealistic assumptions for the coupling between inversion layer and bulk.
[34] T. Hörmann and G. Brunthaler (unpublished).
[35] J. S. Blakemore, *Semiconductor Statistics* (Courier Dove, New York, 1987).
[36] F. Stern, Phys. Rev. B **5**, 4891 (1972).