The effect of interface roughness on the transport properties of a two-dimensional electron gas in a Si-MOSFET

1994 J. Phys.: Condens. Matter 6 2713

(http://iopscience.iop.org/0953-8984/6/14/009)

J. Phys.: Condens. Matter 6 (1994) 2713–2726. Printed in the UK

The effect of interface roughness on the transport properties of a two-dimensional electron gas in a Si-MOSFET

X Zianni and P N Butcher
University of Warwick, Physics Department, Coventry CV4 7AL, UK

Received 15 December 1993

Abstract. The mobility and thermopower of a 2DEG are calculated in order to explain measurements in two Si-MOSFETs. In one of the samples at $T < 2$ K and for electron densities $N_s > 10^{16}$ m^{-2}, positive thermopowers have been measured. The change of sign in the thermopower is attributed to the dominance of interface roughness scattering. Simultaneous consideration of two transport properties indicates the inadequacy of the current theory of interface roughness to explain electron scattering by interface irregularities in one of the samples. At $T > 2$ K, thermopower is explained by phonon drag and good agreement is found between theory and experiment.

1. Introduction

The mobility of a two-dimensional electron gas (2DEG) has been extensively studied both theoretically and experimentally because it reflects the dominant scattering processes. In the temperature range considered here ($T < 7$ K), the mobility of the 2DEG is limited in Si-MOSFETs by ionized impurities and interface roughness. The thermoelectric properties of electrons in low-dimensional structures have also attracted the interest of researchers. Diffusion thermopower reflects the energy dependence of the relaxation time, and is determined not only by the magnitude of the scattering but also by details concerning the distribution of scatterers and their type. Phonon-drag thermopower involves the electron–phonon interaction and, in the present system, is dominant above a temperature in the order of 2 K.

Electron mobility and thermopower have so far been examined independently. In this paper, calculations of both transport properties are presented in order to improve the understanding of the behaviour of 2D structures. The need to do this is shown in a special case in which the thermopower acquires positive values when $T < 2$ K.

The electron mobility and thermopower of two Si-MOSFETs have been measured by Gallagher and Oxley [1, 2]. The main difference between the two samples is the acceptor concentration N_A in the Si, which is two orders of magnitude higher in sample B ($N_A = 1.35 \times 10^{21}$ m^{-3}) than in sample A ($N_A = 1.5 \times 10^{19}$ m^{-3}). Since the samples were fabricated independently, they are also expected to have different ionized impurity concentrations and distributions and different interface quality. The mobility increases with electron density N_s to a maximum value and then decreases with increasing N_s. The peak mobility is 1.26 m^2 V^{-1} s^{-1} for sample A and 1.53 m^2 V^{-1} s^{-1} for sample B at $T = 1$ K. Sample A exhibits higher thermopower magnitudes than sample B. It is particularly notable that positive thermopower values have been measured in sample B at $T < 1.5$ K

0953-8984/94/142713+14$19.50

and $N_s \sim 10^{16}$ m^{-2} whereas, in sample A at even higher electron densities, and for all temperatures, the thermopower remains negative.

The positive thermopower has been attributed to a positive diffusion thermopower, which dominates when $T < 2$ K in a Si-MOSFET [1]. Positive thermopower has also been observed in other systems [3], in which it is attributed to second-subband occupation. In samples A and B only the ground subband is thought to be occupied. As supporting evidence we note that in a Si-MOSFET, with $N_{\mathrm{depl}} = 1.55 \times 10^{15}$ m^{-2} and $N_A = 1.65 \times 10^{21}$ m^{-3} (i.e. similar to the boron doping level in sample B), higher subbands become occupied only when $N_s > 3 \times 10^{16}$ m^{-2} [4]. Furthermore, the subband separation decreases with decreasing N_{depl}, and in sample A the second subband would become occupied well before it does in sample B. There has been an attempt to explain the positive thermopower using background impurity scattering [5], and measurements on sample A seem to support this idea. However, the present calculations do not support the theory proposed by Karavolas *et al* [5].

The basic theoretical formalism used in the calculations is presented in sections 2 and 3. In section 4 the dominant scattering mechanisms are discussed, the results are presented in section 5 and conclusions are drawn in section 6.

2. Ionized impurity scattering

The problem of elastic electron scattering by ionized impurities in the electric quantum limit has been formulated by Stern and Howard [6] and is treated here following their approach. We assume that the impurity charge is $\pm e$.

The transport relaxation rate for an electron with energy ε, in the Born approximation, is [4]

$$\frac{1}{\tau(\varepsilon)} = \frac{2\pi}{\hbar} \int \mathrm{d}z\, N_i(z) \sum_q \left[\frac{e^2 F(q,z)}{2\epsilon_0 \kappa_{sc} q \epsilon(q,T)}\right]^2 (1-\cos\theta)\delta(\varepsilon(k)-\varepsilon(k+q)). \tag{1}$$

In this equation $\varepsilon(k) = \hbar^2 k^2/2m^*$, θ is the scattering angle, ϵ_0 is the permittivity of free space, κ_{sc} is the static dielectric constant of Si, k is the electron wavevector and q is the change of electron wavevector due to scattering. $N_i(z)$ is the impurity density with i indicating the location of the ionized impurities, inside or outside of the inversion layer. The temperature-dependent dielectric function $\epsilon(q,T)$ is calculated in the random-phase approximation [7]. The form factor $F(q,z)$ can be expressed in terms of the variational wavefunction as follows [4].

For remote impurities ($z < 0$):

$$F(q,z) = P_0 \mathrm{e}^{qz}. \tag{2}$$

For background impurities ($z > 0$):

$$F(q,z) = [P(z) + \delta_k P_0 \mathrm{e}^{-qz}] \tag{3}$$

where

$$P(z) = \begin{cases} [b^3/(b-q)^3][\mathrm{e}^{-qz} - (\alpha_0 + \alpha_1 z + \alpha_2 z^2)\mathrm{e}^{-bz}] & q \neq b \\ \frac{1}{8}[1 + 2bz + 2b^2z^2 + \frac{4}{3}b^3z^3]\mathrm{e}^{-bz} & q = b \end{cases} \tag{4}$$

with

$$\alpha_0 = \frac{2q(3b^2+q^2)}{(b+q)^3)} \tag{5}$$

$$\alpha_1 = \frac{4bq(b-q)}{(b+q)^2} \tag{6}$$

$$\alpha_2 = \frac{q(b-q)^2}{b+q} \tag{7}$$

$$\delta_k = \frac{\kappa_{sc} - \kappa_{ins}}{\kappa_{sc} + \kappa_{ins}} \tag{8}$$

and

$$P_0 = \frac{b^3}{(b+q)^3}. \tag{9}$$

Here, b is the Fang and Howard variational wavefunction parameter [8]:

$$b = \frac{b_{SH}}{\beta} \tag{10}$$

where

$$b_{SH}^3 = \frac{12e^2 m_z^*}{\varepsilon_0 \kappa_{sc} \hbar^2}(N_{depl} + \tfrac{11}{32}N_s) \tag{11}$$

$$\beta = \delta^{1/2}\cosh(\tfrac{1}{3}\cosh^{-1}(4\delta^{-3/2})) \tag{12}$$

and

$$\delta = \frac{3}{8}\frac{m_z^*(\kappa_{sc} - \kappa_{ins})e^2}{4\pi\varepsilon_0\hbar^2 b_{SH}(\kappa_{sc} + \kappa_{ins})\kappa_{sc}}. \tag{13}$$

Neglecting the image term, b becomes equal to b_{SH}. The electron mobility is given by

$$\mu = \frac{e\tau}{m^*}. \tag{14}$$

Since the form factor depends on the position of the charged centres, the dependence of the mobility on the electron density also does so. Numerical calculations show that the variation is the slowest when the impurities are on the interface and the steepest when they are distributed in the insulator.

2.1. *Interface roughness scattering*

Any deviation from the perfect interface causes electron scattering. Because of the incomplete knowledge of the interface structure, the roughness is treated in a simple model described in detail by Ando *et al* [4]. The relaxation time is given by

$$\frac{1}{\tau} = \frac{2\pi}{\hbar}\sum_q \pi \left|\frac{\Delta\Lambda\Gamma(q)}{\epsilon(q,T)}\right|^2 e^{-q^2\Lambda^2/4}(1-\cos\theta)\delta(\varepsilon(k) - \varepsilon(k+q)) \tag{15}$$

where (assuming a Gaussian form for the correlation of the surface roughness) we obtain for the Fourier transform of the correlation function of $\Delta(r)$

$$\langle|\Delta_q|^2\rangle = \pi\Delta^2\Lambda^2 e^{-q^2\Lambda^2/4} \tag{16}$$

where Δ is the average displacement of the interface and Λ is the lateral correlation length.

The matrix element is

$$\Gamma(q) = \gamma(q) + \gamma_{im}(q). \tag{17}$$

The first term in equation (17) relates to the effective electric field that arises from the positive charges at the gate electrode and those in the depletion layer. The second term

is the image term, which accounts for the effect of the different dielectric constants of the materials across the interface. We refer the reader to [8] for detailed expressions for the functions in equation (17). It is assumed that scattering rates can be added (Matthiesen's rule). The deviations from this rule with variation of T have been found (see [9], [10]) to be negligible below 10 K and for electron densities that are not too low. Hence, the total relaxation rate is given by

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}. \tag{18}$$

3. Thermopower

The current and, correspondingly, the transport coefficients arise from two distinct processes: electron diffusion and phonon-drag of the electrons. The phonon drag contribution to σ is usually negligible [11] but phonon drag plays a very important role in the thermopower. The thermopower resulting from these two processes is described here.

3.1. Diffusion thermopower

Diffusion thermopower is given by Mott's formula [12]:

$$S^{\mathrm{d}} = -\frac{\pi^2 k_{\mathrm{B}}^2 T}{3e\sigma(\varepsilon_{\mathrm{F}})} \frac{\mathrm{d}\sigma(\varepsilon)}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=\varepsilon_{\mathrm{F}}}. \tag{19}$$

It is convenient to introduce the parameter

$$p = \varepsilon \frac{\mathrm{d}\ln\tau}{\mathrm{d}\varepsilon}\Big|_{\varepsilon_{\mathrm{F}}}. \tag{20}$$

The diffusion thermopower can be written:

$$S^{\mathrm{d}} = -\frac{\pi^2 k_{\mathrm{B}}}{3e} \frac{k_{\mathrm{B}} T}{\varepsilon_{\mathrm{F}}} (p+1) \tag{21}$$

We see that S^{d} is negative when $p > -1$ whereas it becomes positive for $p < -1$.

3.2. Phonon drag

A formula for the phonon-drag contribution S_{g} to the thermopower has been derived in [13], in which the coupled electron and phonon Boltzmann equations are linearized and solved in the relaxation-time approximation at low temperatures where only acoustic phonons are important. If both longitudinal (LA) and transverse (TA) acoustic phonons are considered, a phonon-drag formula can be written for each mode s as follows:

$$S_{\mathrm{gs}}^0 = -\frac{2C_{\mathrm{g}}\tau_{\mathrm{F}}}{\sigma_0} \int\int \mathrm{d}q\, \mathrm{d}q_z\, q^2 F(Q) \lambda^0(Q) \times \int_0^\infty \mathrm{d}u\, f(u^2+\gamma)[1 - f(u^2+\gamma+\hbar\omega_Q)]. \tag{22}$$

Here, τ_{F} is the relaxation time at the Fermi level,

$$C_{\mathrm{g}} = \frac{|e|\lambda_{\mathrm{p}} g_{\mathrm{v}} (2m^*)^{1/2}}{32\pi^3 \hbar k_{\mathrm{B}} T^2 \rho} \tag{23}$$

$$F(Q) = \frac{Q|Z_{11}(q_z)|^2}{\varepsilon^2(Q,T)} \frac{1 - \mathrm{e}^{-\hbar\omega_Q/k_{\mathrm{B}}T}}{\sinh^2(\hbar\omega_Q/2k_{\mathrm{B}}T)} E_{\mathrm{s}}^2(Q) \tag{24}$$

and

$$\lambda^0(Q) = 1 + p\frac{1}{2}\frac{\hbar\omega_Q}{\varepsilon_F}\left[1 + \frac{\hbar\omega_Q}{\hbar^2 q^2/2m^*}\right]. \tag{25}$$

The factor $\lambda^0(Q)$ is the result of the consideration of an energy-dependent relaxation time in the integrand of the phonon-drag formula. The valley degeneracy g_v in equation (23) is two for Si and the in-plane effective mass m^* is $0.1905m_e$, where m_e is the free electron mass; $\hbar\omega_{Qs}$ is the phonon energy for wave vector Q and mode s, $f(k)$ is the is the Fermi distribution function, ρ is the density of the material (2.39 g cm^{-3} for Si), and λ_p is the phonon mean free path, which is derived from measured values of the bulk thermal conductivity [16]. At low temperatures boundary scattering dominates the phonon–phonon interaction and λ_p is mainly determined by the sample dimensions. The overlap integral Z_{11} is for the variatonal wavefunction is [4]

$$|Z_{11}(q_z)|^2 = b^6/(b^2 + q_z^2)^3 \tag{26}$$

and $E_s(Q)$ is the effective deformation coupling function for mode s and is given by

$$E_s^2(Q) = \begin{cases} \Xi_u^2(q_z^2/Q^2 + D)^2 & \text{LA} \\ \Xi_u^2(q^2 q_z^2/Q^4) & \text{TA} \end{cases} \tag{27}$$

Here, Ξ_u is the deformation potential for pure shear strain and $D = \Xi_d/\Xi_u$, with Ξ_d denoting the deformation potential for pure dilation. The values used in the calculations for Si are $\Xi_u = 9.0$ eV and $\Xi_d = -6.0$ eV [4]. Finally, the u-integration in equation (22) accounts for the spread of electron energies involved in phonon drag with

$$u^2 = \varepsilon(k) - \gamma \tag{28}$$

where

$$\gamma = \frac{(\hbar\omega_Q - \hbar^2 q^2/2m^*)^2}{4(\hbar^2 q^2/2m^*)}. \tag{29}$$

4. Dominant scattering mechanisms

Two mechanisms are known to contribute to the scattering of electrons in inversion layers in Si at low temperatures: charged impurity and interface roughness scattering. Scattering associated with Coulomb centres near the plane of the 2DEG can be separated into contributions from the depletion layer, the oxide charge and the interface charge. As a result of image forces, sodium ions (probably the main contribution to the interface charge) are attracted towards and remain close to the interface [17, 18].

The effectiveness of each scattering potential can be seen from the dependence on scattering angle of the integrand in the formula for the corresponding relaxation rate. Typical examples are shown in figure 1 for interface roughness (IFR) and charged impurity scattering when the charges are in the background (BI) of the electron gas, when they are located on the interface (IF) and when they are on the other side of the interface (RI). The curves have been normalized so that each one integrates to unity between the limits zero and π. It can be seen that interface roughness primarily scatters the electrons at large angles. For Coulombic scattering, the angular distribution is more extended and depends strongly on the position of the charged centres with respect to the electron gas. The electrons are scattered more effectively (i.e. at larger angles) the closer the impurities are to the electrons. In a zero-temperature treatment of screening, the calculated relaxation times are singular at

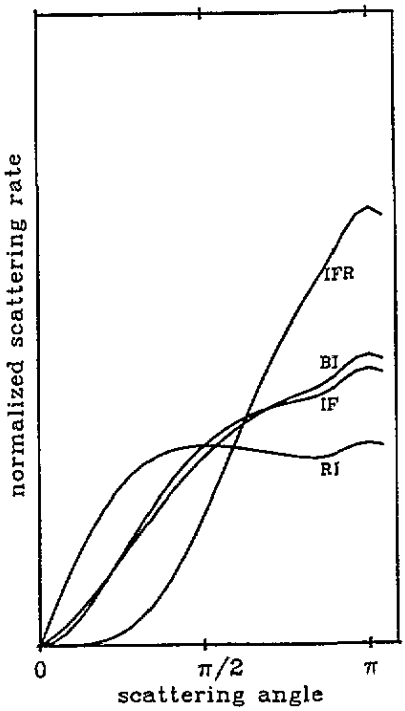

Figure 1. The normalized scattering rate at the Fermi level versus scattering angle for scattering by interface roughness (IFR), background (BI), interface (IF) and remote (RI) ionized impurities.

ε_F (see figure 2), and so calculated values of p [5] cannot be trusted. The temperature dependence of the screening is thus important for the energy dependence of the relaxation time τ around ε_F, and it is consequently crucial for the determination of the diffusion thermopower as equations (19) and (20) show. Figure 2 shows typical results. We see that the image potential reduces the scattering in all cases except for background impurities.

Usually, the charged acceptors in bulk Si play little role as scatterers except at extremely small electron concentrations. For a bulk doping of 10^{22} m^{-3}, there are only $\sim 10^{14}$ m^{-2} acceptors within 100 Å from the interface in most samples. Background impurity scattering has been found to be negligible even for sample B (the corresponding relaxation time is of the order of 100 ps) in contrast to the predictions of [5, 19]. Gallagher's experiments also indicate that background impurities do not dominate at high N_s, since the measured mobilities of sample B (high N_A) are bigger than those of sample A (low N_A). Another important difference between our calculations and those presented in [5, 19] is the electron-density dependence of the mobility limited by background impurities. We find the same behaviour for background impurities as for the other ionized impurities. Our findings are supported by similar calculations presented in [20]. Furthermore, for background impurity scattering p is found here to be only slightly negative (see figure 3). It is therefore impossible to attribute positive thermopower to background impurity scattering. In [5] a value of $p = -3$ is calculated, which leads the authors to a contrary conclusion. Our calculations do not support the results given in [5], although they are based on the same theory. It seems that the authors of [5] may have been misled by a numerical error.

Our calculated values of p are shown in figure 3 for both samples when all the scattering

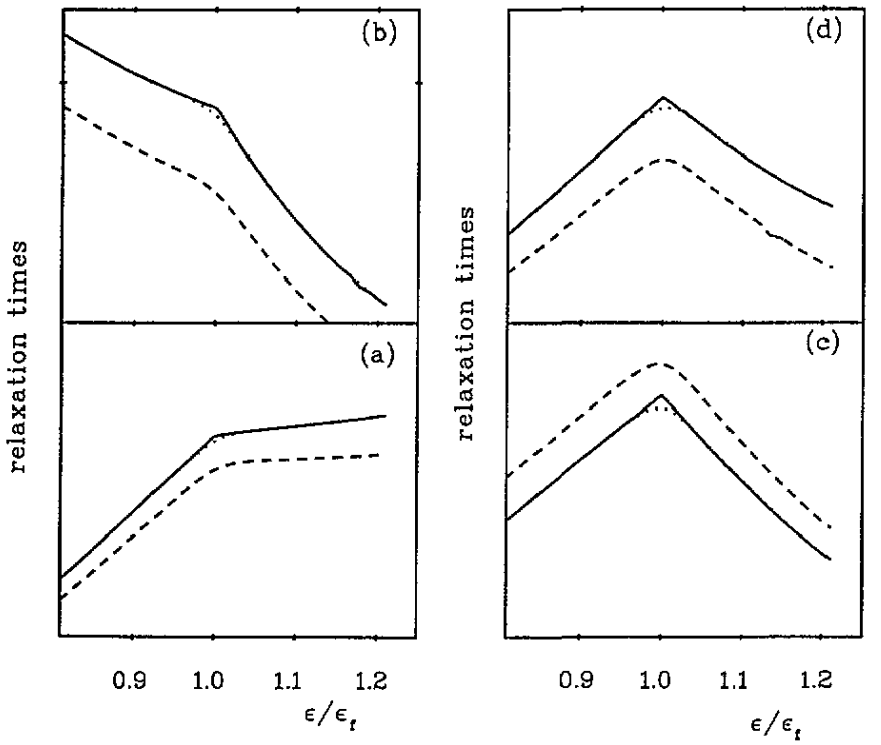

Figure 2. The electron relaxation times (linear axis scale) limited by remote impurities (a), interface roughness (b), background impurities (c) and interface impurities (d) are plotted against $\varepsilon/\varepsilon_F$ when image effects are neglected (dashed lines), when temperature broadening of screening is neglected (solid lines) and when all corrections are included (dotted lines).

mechanisms are considered. By comparing figure 3(b) and (d), it can be seen that the value of p for each scattering mechanism is only slightly affected by N_A. For interface roughness scattering, p is very sensitive to the parameters Λ and Δ. The sign of p is determined by the effectiveness of the scattering potential and it is positive (negative) for small- (large-) angle-dominated scattering mechanisms. It has been already mentioned that the further from the electrons the impurities are, the more small-angle scattering dominates. It is therefore not surprising that p is positive for remote impurities, slightly negative for interface ionized impurities and even more negative for background impurities. Strongly negative values of p are found for interface roughness scattering, in agreement with the fact that short-range roughness causes large angle scattering (see figure 1).

The concentration of Coulomb centres at and near the interface in the oxide layer has little effect on p. It primarily modifies the strength of the process as a proportionality constant. This explains why the calculated values of p for remote impurities and impurities on the interface are almost the same in both samples irrespective of the different impurity concentrations (figure 3 (b) and (d)). The cause of the small changes in p for the individual scattering processes is the different acceptor concentration of the two samples, since N_A influences the electron distribution as shown in equations (10) and (11). Nevertheless, all the ionized impurity concentrations are important in determining the value of the total p, since they determine the weight of each scattering process in the total scattering rate according to equation (18). Finally, for interface roughness the strength of scattering depends on the

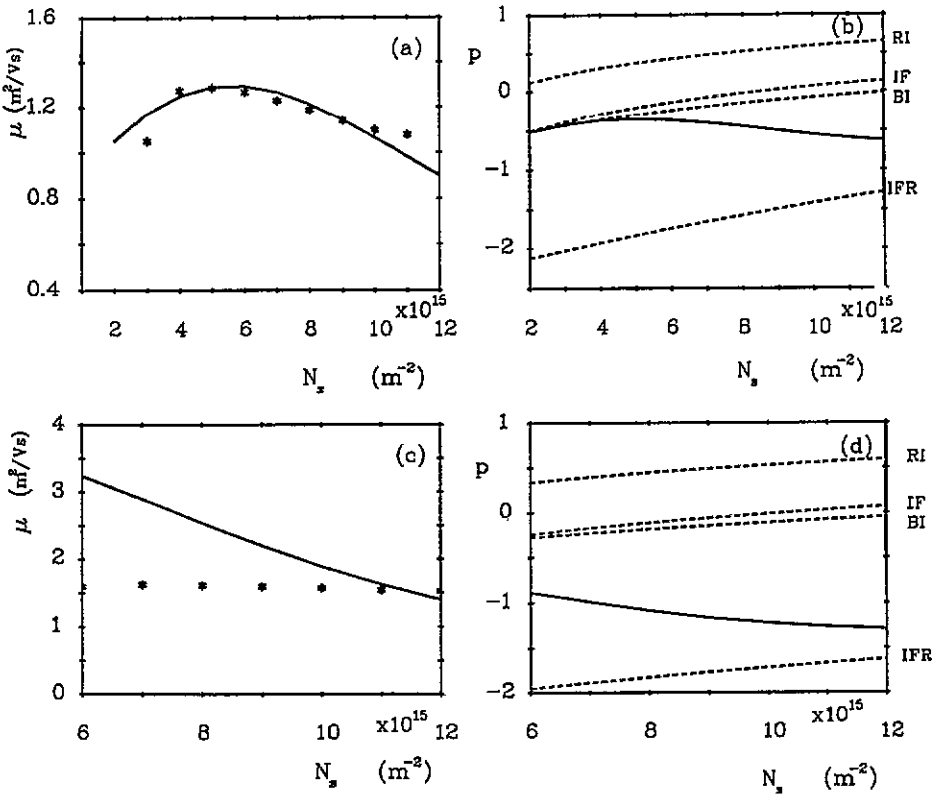

Figure 3. The electron-density variations of the calculated electron mobilities (solid curves) are compared with those measured at $T = 1$ K for (a) sample A; $N_{IF} = 8 \times 10^{14}$ m^{-2}, $\Lambda = 35$ Å and $\Delta = 7$ Å and (c) sample B; $N_{IF} = 2 \times 10^{14}$ m^{-2}, $\Lambda = 20$ Å and $\Delta = 9$ Å. The calculated p-values for the individual scattering processes and their sum (solid line) are shown for sample A in (b) and for sample B in (d).

values of both Λ and Δ whereas p depends only on the value of Λ.

5. Calculated transport coefficients

5.1. $T < 2$ K

In order to calculate the relaxation times related to each scattering mechanism, the impurity density $N_i(z)$ and the interface roughness parameters, Λ and Δ, are needed, as can be seen in equations (1) and (15). These parameters determine the weight of each scattering mechanism in limiting the electron mobility. Their magnitudes are not accurately controlled in the experiments and they can only be estimated by trying to fit the experimental data. After a systematic search, it is concluded that the experiments can be satisfactorily explained by ionized impurities on the interface and interface roughness scattering.

Calculated mobilities together with the experimental data are shown in figure 4 for both samples. The concentration of interface impurities is chosen to be $N_{IF} = 7.5 \times 10^{14}$ m^{-2} for sample A and 5.4×10^{14} m^{-2} for sample B. The interface roughness parameters are estimated to be $\Lambda = 82$ Å and $\Delta = 6$ Å for sample A and $\Lambda = 76$ Å and $\Delta = 4$ Å for sample B. At low electron concentrations the mobility decreases rapidly because of the onset of

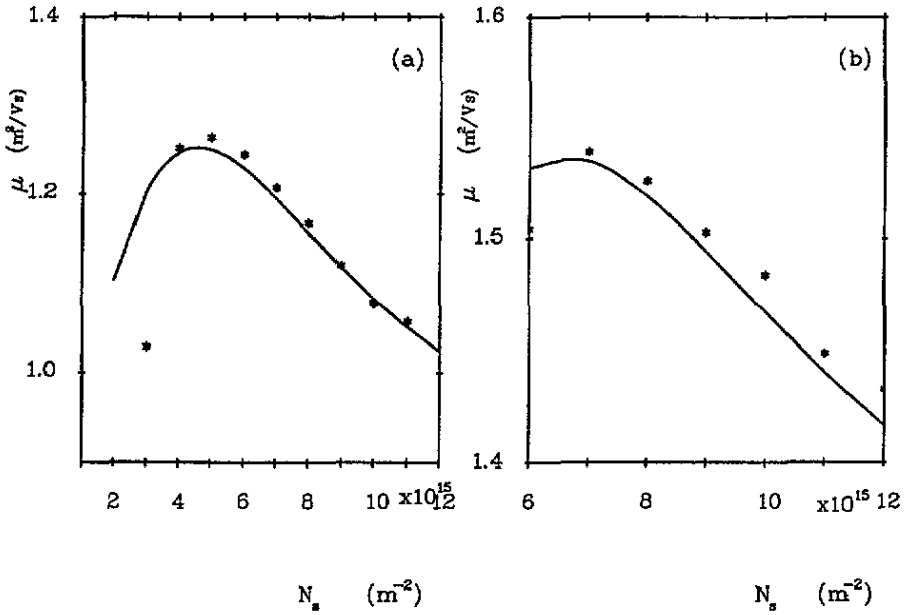

Figure 4. The electron-density variations of the calculated electron mobilities (solid curves) are compared with those measured at $T = 1$ K for (a) sample A; $N_{IF} = 7.5 \times 10^{14}$ m^{-2}, $\Lambda = 82$ Å and $\Delta = 6$ Å and (b) sample B; $N_{IF} = 5.4 \times 10^{14}$ m^{-2}, $\Lambda = 76$ Å and $\Delta = 4$ Å.

thermally activated hopping in which effects of disorder are important. These effects have been neglected here because they only affect a restricted region of low electron densities well below $N_s \sim 10^{16}$ m^{-2}, where the change of sign of the thermopower is observed.

Although the measured mobilities are very well explained by the above parameters, both sets of parameters give high values of thermopower when $T < 2$ K so that diffusion dominates. Moreover, no positive thermopower is found for sample B. This is because a positive value of p arises with the assumed parameters. The observed thermopower magnitudes require a negative p. Positive values of p are associated with the assumed long-range roughness, characterized by $\Lambda > k_F^{-1}$. In this case the screening decreases the effect of the roughness, and as the electron gas becomes more dense, the carriers 'see' less roughness. They are then less effectively scattered and p is therefore positive. When $\Lambda < k_F^{-1}$, the electrons 'feel' the interface irregularities to a greater extent and, as their density increases, they move closer to the interface and are scattered more effectively, so that p is negative. We conclude that samples A and B both have a small lateral correlation length for the surface roughness.

To study the consequences of this idea, the scattering parameters are now modified as follows. The concentration of interface impurities is estimated as $N_{IF} = 8 \times 10^{14}$ m^{-2} for sample A and 2×10^{14} m^{-2} for sample B. It is noted that these densities are in agreement with [17, 21], where the surface charge is estimated as $\sim 10^{14}$ m^{-2}. The interface roughness parameters are estimated as $\Lambda = 35$ Å, $\Delta = 7$ Å for sample A and $\Lambda = 20$ Å, $\Delta = 9$ Å for sample B. The calculated mobilities and thermopowers of the two samples are shown in figures 3 and 5. Quantitatively, the results are in a very good agreement with the data, with the exception of the mobility of sample B. Although the assumed height of the roughness (Δ) is relatively high in both samples, it can be considered acceptable, since similar values have been reported elsewhere [22]. The roughness is very sensitive to the conditions of

fabrication and microscopic investigation is the only reliable way to estimate its form and magnitude; the large values of Δ would not therefore be the main problem, if the mobility could be satisfactorily explained. Here, a much steeper decrease of the mobility with N_s is calculated than is measured. We suggest that the present isotropic Gaussian statistical theory of interface roughness is inadequate to describe the magnitude and the dependence of the mobility on N_s when we assume an energy dependence of the relaxation time compatible with the observed thermopower values.

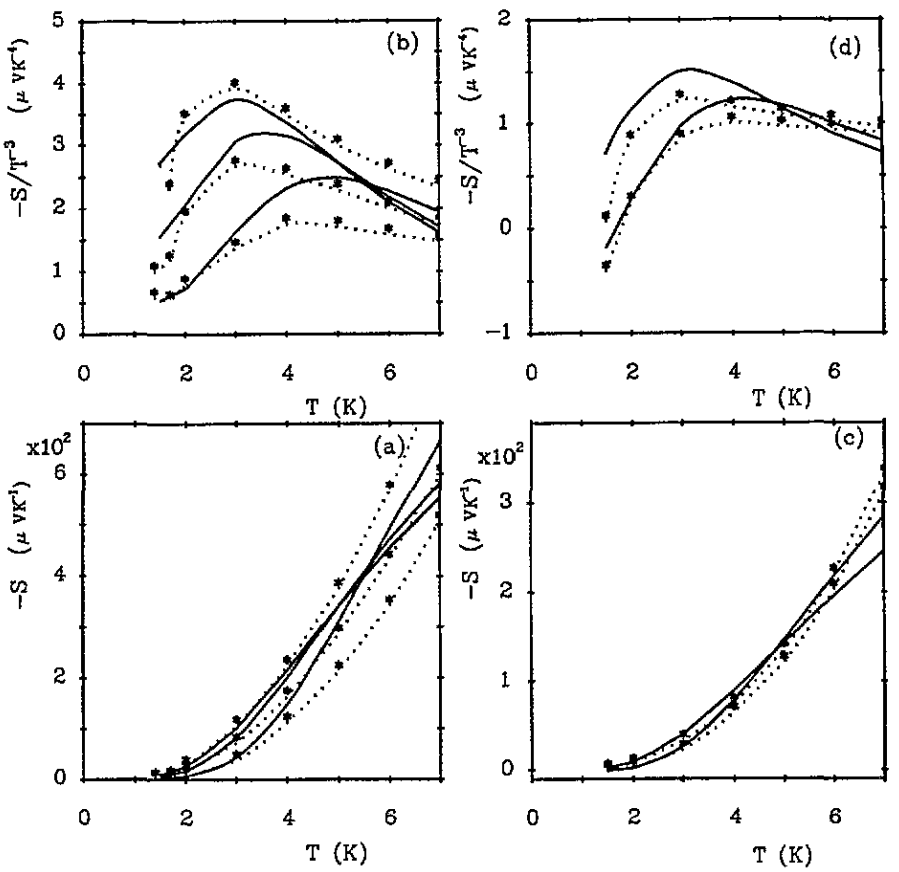

Figure 5. The temperature variations of $-S$ and $-S/T^3$ are plotted for different electron densities in (a, c) and (b, d) respectively for both samples. The solid lines are the calculated values and the asterisks the measured values; the curves through the asterisks are guides to the eye. From the uppermost to the lowest curve the corresponding electron densities are (i) for sample A: 8.01, 10.7, 20 $\times 10^{15}$ m^{-2} (a, b), and (ii) for sample B: 8.35, 14.1 $\times 10^{15}$ m^{-2} (c, d).

Goodnick *et al* [23] investigate the statistical properties of random fluctuations in the interface boundary using cross-sectional high-resolution transmission electron microscopy (HRTEM). The observed spectral properties of the roughness on normally prepared and intentionally roughened samples appear to be well characterized by an exponential decay in the autocovariance function rather than the Gaussian approximation. However, they find that the difference in the statistical properties of the interface does not lead to significant changes in the dependence of the mobility on N_s, particularly when $\Lambda < k_F^{-1}$ [24]. The parameters characterizing the roughness are only slightly different when either the Gaussian

or the exponential correlation is used. We also note that in this analysis the interface boundary is chosen as the last discernible lattice fringe corresponding to the periodicity of the Si. This procedure is somewhat arbitrary at many points, as an abrupt change from crystalline Si to non-crystalline SiO_2 is not always apparent. The analysis also neglects the contribution to the interface width of intermediate bonding states and the existence of a transition layer between Si and SiO_2. Stern [25] estimates the transition layer width to be of the order of 5 Å and suggests that electrons in the lowest subband for a (001) interface have up to 1% of their charge within the transition layer at large interface electric fields. Stern also supposes that scatterers in the transition layer can account for interface scattering, which has generally been attributed to surface roughness.

The exact form of the power spectrum of the roughness is determined by the oxidation process and varies from sample to sample. Only a precise knowledge of the physics involved in oxide growth would permit a more detailed theoretical description. Scanning tunnelling microscopy might be able to reveal the importance of any anisotropy of the correlation [17]. The details of the surface structure are more important at high N_s and N_A, because in both cases the electrons are close to the interface. In sample A the electron's mean position $\langle z \rangle \sim 35$ Å and, in sample B, $\langle z \rangle \sim 30$ Å. It is possible that the carriers in sample B, being closer to the interface than those in sample A, and therefore having a bigger proportion of their wavefunctions in the transition layer, are consequently more sensitive to disorder in the neighbourhood of the interface.

Finally we note that p, as defined by equation (20), is appropriate in the calculation of S_d and it is not at all obvious that it coincides with $\mathrm{d}\ln\tau/\mathrm{d}\ln\varepsilon_F$, as has sometimes been assumed [29]. In sample B at high densities $\mathrm{d}\ln\tau/\mathrm{d}\ln\varepsilon_F \sim -0.2$ and cannot explain the positive values of thermopower.

5.2. $T > 2$ K

At temperatures above 2 K, phonon drag thermopower dominates over diffusion thermopower. Calculated and measured thermopower values are shown in figure 5 for the two samples. The phonon-drag calculations are based on Cantrell–Butcher theory [13]. Some improvement has been made by considering the temperature dependence of screening and the phonon mean free path (in sample A λ_p =1.08 mm, 0.92 mm at $T = 1.5$ K and 7 K respectively; in sample B λ_p =0.56 mm, 0.5 mm at $T = 1.5$ K and 7 K respectively) and the effect of thermal broadening of the Fermi function (see equation (22)).

Oxley provides plots of S^{-1} against N_s for a range of T [2]. The N_s^{-1} dependence of S_g predicted by simple models is supported by the experimental data, except at the highest electron densities. The curves intercept the N_s-axis at $N_{MIT} = 1.2 \times 10^{15}$ m^{-2} for sample A and $N_{MIT} = 1.75 \times 10^{15}$ m^{-2} for sample B. This point corresponds to the transition between strongly and weakly localized electron states [26, 27]. Since phonons are assumed to interact with free electrons, the appropriate electron concentration in the calculation of S_g is $N_{sfree} = N_s - N_{MIT}$. This correction causes an enhancement of $-S_g$, because the corresponding N_s is reduced and there is also a shift of the positions of the peak in $-S_g/T^3$ to lower T. The calculated peaks now lie very close to the measured ones. In figure 6 the results including this correction are compared with those obtained neglecting it.

Electron screening reduces the effectiveness of the momentum transfer from phonons to electrons. An extensive quantitative analysis is presented by Smith and Butcher [7] and it is shown that the calculated magnitudes of S_g are lowered up to 20 times by screening so that its inclusion in the theory makes possible a quantitative comparison with experiment. At electron concentrations greater than 10^{16} m^{-2}, the calculations seem to overestimate

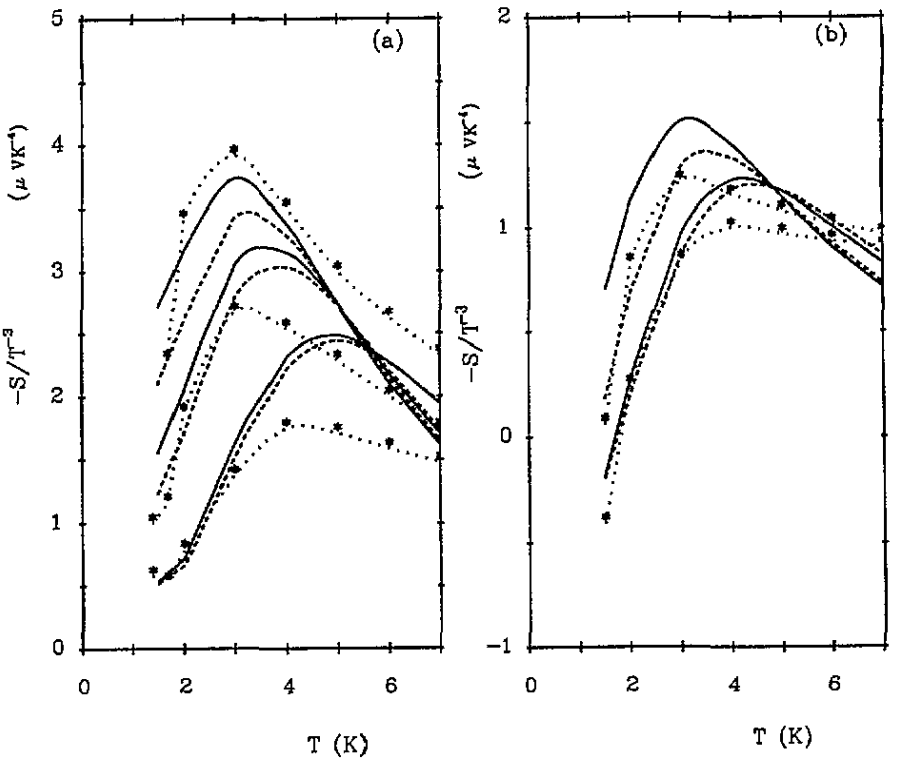

Figure 6. Similar plots to those in figure 5 (b) and (d) with the additional broken curves showing the calculated values when the metal–insulator transition in N_s is neglected.

$|S_g|$. One reason for this might be the single-subband approximation (SSA) in screening [7] which fails at high N_s. The subbands separation decreases with decreasing N_A and it seems possible that screening is underestimated by the SSA.

We also note that at high temperatures and low electron densities the calculations underestimate the thermopower. In Si, the lattice anisotropy in the electron–phonon interaction has been regarded as small and instead of considering each direction of the phonon wavevector Q separately in calculating the transition rate, as it has been done by Herring and Vogt [30], an implicit averaging over the azimuthial angle of Q has been adopted [15]. Proper consideration of the phonon anisotropy would broaden the phonon distribution and further enhance the electron–phonon interaction at high temperatures and low electron densities.

Calculations in which the temperature dependence of the screening, image potential corrections and the energy dependence of the electron relaxation time are omitted show that these only have only a small effect on S_g. What does become important, as T approaches 7 K, is the thermal tail on the Fermi function. The present calculations allow for non-degeneracy of the electron energy by properly dealing with the u-integration. At $T = 7$ K, $|S_g|$ is reduced by between 30 and 40% for N_s between 3.5 and 12.7×10^{15} m^{-2} [14]. Neglecting the non-degeneracy effect would overestimate the thermopower.

6. Conclusions

The main contribution of this work to the understanding of electron transport properties, when a single subband is occupied, is the simultaneous consideration of the mobility and the thermopower. Electron-mobility data have been commonly satisfactorily explained by scattering by ionized impurities and by interface irregularities [4]. A simple theory was considered adequate in modelling the interface roughness. This study shows that the quantitative success of such treatments often relies on the freedom used to estimate the weight of each scattering mechanism which is involved.

In almost all cases, and for both the systems examined, the mobility and the thermopower have been simultaneously explained satisfactorily. A serious discrepancy between theory and experiment is found only for one Si-MOSFET sample, in which a change of sign in the thermopower is observed at low temperatures. This change of sign is attributed to dominant interface roughness scattering. Here, the theory deviates from the measured N_s dependence of the mobility at high electron densities. It is concluded that this discrepancy is due to the failure of theory of interface roughness to accommodate effects caused by the peculiarities of the structure near the interface. The need for a more elaborate theory for interface roughness is indicated. Guidance from experiments can be obtained by a microscopic examination of Si/SiO_2 interfaces in samples that present positive thermopowers at low temperatures and electron densities low enough to exclude second subband occupation, combined with similar examinations of samples that do not show positive thermopower data.

Acknowledgments

The authors would like to thank John Oxley and Bryan Gallagher for giving them access to their data prior to publication and for many helpful discussions about its interpretation. This work has been supported by SERC, the Hirst Research Centre and the University of Warwick.

References

[1] Gallagher B L, Oxley J P, Galloway T, Smith M J and Butcher P N 1990 *J. Phys.: Condens. Matter* **2** 755
[2] Oxley J P 1991 *PhD Thesis* University of Nottingham
[3] Ruf C, Brummell M A, Gmellin E and Ploog K 1989 *Superlatt. Microstruct.* **6** 175
[4] Ando T, Fowler H and Stern F 1982 *Rev. Mod. Phys.* **54** 437
[5] Karavolas V K, Fromhold T M, Butcher P N, Mulimani B G, Gallagher B L and Oxley T P 1990 *J. Phys.: Condens. Matter* **2** 10401
[6] Stern F and Howard W E 1967 *Phys. Rev.* **163** 816
[7] Smith J and Butcher P N 1989 *J. Phys.: Condens. Matter* **1** 1261
[8] Ando T 1977 *J. Phys. Soc. Japan* **43** 1616
[9] Kawamura T and Das Sarma S 1990 *Phys. Rev.* B **42** 3725
[10] Stern F 1980 *Phys. Rev. Lett.* **44** 1469
[11] Blatt F J 1968 *Physics of Electronic Conduction in Solids* (New York: McGraw-Hill)
[12] Butcher P N 1993 *Physics of Low-Dimensional Semiconductor Structures* eds P N Butcher N H March and M Tosi (New York: Plenum) pp 95–176
[13] Cantrell D G and Butcher P N 1987 *J. Phys. C: Solid State Phys.* **20** 1985
[14] Smith M J 1989 *PhD thesis* University of Warwick
[15] Ridley B K 1982 *Quantum Processes in Semiconductors* (Oxford: Clarendon)
[16] Ashroft M N and Mermin N D 1976 *Solid State Physics* (New York: Holt, Saunders)
[17] Kruithof G H, Klapwijk T M and Bakker S 1991 *Phys. Rev.* B **43** 6642
[18] DiMaria D J 19777 *J. Appl. Phys.* **48** 5149
[19] Karavolas V K and Butcher P N 1991 *J. Phys.: Condens. Matter* **3** 2597

[20] Matsumoto Y and Uemura Y 1974 *Proc. 2nd Int. Conf. on Solid Surfaces (1974); Japan. J. Appl. Phys. Suppl.* 2 367
[21] Cheng Y C and Sullivan E A 1973 *Surf. Sci.* 34 717
[22] Gold A and Dolgopolov V T 1986 *Phys. Rev.* B 33 1076 and references therein
[23] Goodnick S M, Ferry D K and Wilmsen C W 1985 *Phys. Rev.* B 32 8171
[24] Fishmann G and Calecki D 1991 *Phys. Rev.* B 43 11581
[25] Stern F 1977 *Solid State Commun.* 21 163
[26] Kramer B, Bergmann G and Bruynserade Y 1985 *Localization, Interaction and Transport Phenomena* (Berlin: Springer)
[27] Nagaoma Y and Fukyama H 1982 *Anderson Localization* (Berlin: Springer)
[28] Smith M J and Butcher P N 1990 *J. Phys.: Condens. Matter* 2 2375
[29] Gallagher B L and Butcher P N 1992 *Handbook on Semiconductors* vol 1, ed T S Moss (Amsterdam: North-Holland) p 721–816
[30] Herring C and Vogt E 1956 *Phys. Rev.* **101** 944