IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

A calculation of the effect of screening on phonon drag thermopower in a Si MOSFET

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1989 J. Phys.: Condens. Matter 1 1261

(http://iopscience.iop.org/0953-8984/1/7/010)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 171.66.16.90
The article was downloaded on 10/05/2010 at 17:47

Please note that terms and conditions apply.

J. Phys.: Condens. Matter **1** (1989) 1261–1273. Printed in the UK

# A calculation of the effect of screening on phonon drag thermopower in a Si MOSFET

M J Smith and P N Butcher
Department of Physics, University of Warwick, Coventry CV4 7AL, UK

Received 4 August 1988

**Abstract.** Phonon drag thermopower $S_g$ calculations are presented for a quasi-2D electron gas at the (100) plane in a Si MOSFET and compared with previous theory and experiment. Without screening $-S_g$ is improved to a factor of approximately 16 times larger than experiment (from approximately 35). Multi-sub-band screening is considered and its importance assessed in the experimental range (2–6 K and electron densities to $10^{16}\,m^{-2}$) but a single sub-band approximation is found to be sufficient and gives thermopowers just approximately 40% different from experiment. Further improvement is made by adopting the Fang and Howard variational wavefunction and better material parameters. The best agreement is found to be about 5% at temperatures in the mid-range (4 K) and for the highest electron densities. The same excellent qualitative agreement with experiment is retained including the peak in $-S_g/T^3$ whose presence is explained by enhanced phonon absorption around the Kohn resonance ($q \simeq 2k_F$). The peak positions move up in temperature (by about 0.75 K) because the dominant $q$-value increases and reduces the effect of screening on $S_g$ at higher $T$. In GaAs/GaAlAs heterojunctions $S_g$ is much less sensitive to screening effects because the screening constant is much smaller in GaAs than it is in Si.

## 1. Introduction

Both experiment (Gallagher *et al* 1987, Fletcher *et al* 1986, 1988, Ruf *et al* 1988) and theory (Cantrell and Butcher 1987a, b) confirm the dominance at liquid helium temperatures of the phonon drag contribution $S_g$ to the thermopower $S$ of a quasi-2D electron gas (Q2DEG) in inversion layers and at heterojunctions. In Si MOSFETS (Gallagher *et al* 1987) and in GaAs/GaAlAs heterojunctions (Fletcher *et al* 1986, 1988, Ruf *et al* 1988) measured thermopowers much greater than the value $S_d$ expected from metallic electron diffusion processes alone (e.g. Blatt 1968) have been reported. Non-linear temperature dependence consistent with a large $S_g$ contribution is found. This was confirmed by Cantrell and Butcher (1987b) whose $S_g$ calculations give good qualitative agreement with experiment, particularly for Si in which a peak in $-S/T^3$ against temperature $T$ was noted by Gallagher *et al* (1987). Their predicted linear dependence of $S_g$ on the boundary scattering limited phonon mean free path $L$ (typically <1 mm) has since been confirmed in an elegant experiment by Fletcher *et al* (1988). Recently $S_d$ and $S_g$ terms have been isolated in $S$ as $aT$ and $bT^3$ respectively (Ruf *et al* 1988) and the $T^3$ contribution attributed to phonon drag found to dominate in the range 0.6–2.5 K. Thus experimental consensus is being reached. However, the accompanying theory whilst qualitatively accurate is quantitatively unsatisfactory in that the calculated values for $S_g$ are larger than those seen experimentally: by a factor of 35 in the MOSFET and by a factor of 2 in the heterojunction.

Here we present the results of screened $S_g$ calculations for a Q2DEG at the (100) surface in a Si MOSFET. We have taken the formalism of Cantrell and Butcher (1987a) and find that when screening of the electron–phonon interaction is introduced the quantitative agreement becomes extremely good whilst the excellent qualitative agreement is preserved. In § 2 we briefly review the unscreened $S_g$ formula. We find that even without screening this formula actually predicts thermopowers in better agreement with experiment than previously thought. This conclusion is confirmed by using an alternative, more direct, method of performing the numerical integration. In § 3 we show that screening is important in the calculation of $S_g$ and explain why it is more important in Si than it is in GaAs by considering a pure 2DEG. These considerations explain why the unscreened calculations were quantitatively much more accurate for a GaAs/GaAlAs heterojunction than for a Si MOSFET. In § 4 we explore the application of quasi-2D screening and review the multi-sub-band screening (MSS) theory of Mori and Ando (1979) in the infinite square-well (ISW) model used by Cantrell and Butcher (1987b). We present the results of this study by defining an effective MSS dielectric function for screening the matrix elements of a bare potential. We compare this with a single-sub-band approximation (SSA) and find for surface electron densities $n_0 < 10^{16}\,\mathrm{m}^{-2}$ and ISW widths $\delta < 100$ Å (the experimental range of Gallagher *et al* (1987)) that the SSA is adequate. Outside this range more sub-bands are required even in the electric quantum limit because in principle all the sub-band wavefunctions are needed to describe the redistribution of charge across the conducting channel.

The SSA is used in § 5 to screen the electron–phonon interaction in the $S_g$ formula and the results show considerable quantitative improvement over the unscreened $S_g$ values. We go beyond the ISW model by adopting the Fang and Howard variational wavefunction discussed in the review by Ando *et al* (1982) for each $n_0$ and use better values for the material parameters to calculate $S_g$ values within 30% of the experimental data. The best quantitative agreement is found at the higher experimental electron density and temperature values and is better than 5%. Finally we comment on the temperature dependence of $S_g$ in a Q2DEG in comparison to the $T^3$ dependence which is expected for metallic conduction, see for example, Blatt (1968).

## 2. Unscreened $S_g$ calculations

It is important to review the derivation of the unscreened $S_g$ formula both to see how screening may be introduced into it and so that our alternative numerical evaluation of the unscreened formula can be described. However, we outline only the points which are significant for the present paper and refer the reader to Cantrell and Butcher (1987a) for details.

Phonon drag arises from the net flux of phonon momentum in a temperature gradient when there is an electron–phonon interaction and hence a net momentum transfer to the electrons. In the Q2DEG systems of interest to us this interaction is weak but dominates the electron behaviour as evidenced by $S_g$. In the calculation of $S_g$ it is assumed that the phonons have a 3D character and interact with quasi-2D electrons which are free in an $xy$ plane of area $A$ and confined in the $z$ direction, with some characteristic width $\delta$, to a set of sub-bands labelled by an index $\alpha$. Then $(\alpha, \boldsymbol{k})$ labels the electron state with wavevector $\boldsymbol{k}$ in sub-band $\alpha$, with wavefunction:

$$\psi_{\alpha,\boldsymbol{k}}(\boldsymbol{R}) = A^{-1/2}\,\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\varphi_\alpha(z) \tag{2.1}$$

where $\boldsymbol{R}$ is $(\boldsymbol{r}, z)$ and $\boldsymbol{r}$ is $(x, y)$. To calculate $S$ we note that the thermoelectric field in the

direction of a temperature gradient $\nabla T$ in the $xy$ plane, is given at zero current density by:

$$E = S\nabla T. \tag{2.2}$$

When the EMF is reduced to zero by allowing current to flow the 2D thermoelectric current density due to the temperature gradient is

$$J = -\sigma S \nabla T \tag{2.3}$$

in which $\sigma$ is the 2D conductivity. $J$ can be obtained by evaluating the 3D electric current density from (2.1) summing over electron states and integrating across the Q2DEG channel to give:

$$J = \frac{-2e}{A} \sum_{\alpha,k} f_\alpha(k) v_{\alpha,k}. \tag{2.4}$$

Here we account for spin by a factor of 2 and for the occupancy of state $(\alpha, k)$ by the distribution function $f_\alpha(k)$; $v_{\alpha,k}$ is the corresponding electron velocity. $S$ is obtained by comparing (2.3) and (2.4) and $S_g$ is that part of $S$ arising from the departure of the phonon distribution from thermal equilibrium.

To obtain $f_\alpha(k)$ the coupled electron and phonon Boltzmann equations are linearised and solved in the relaxation time approximation at temperatures low enough to allow the neglect of all but acoustic phonons of wavevectors $Q = (q, q_z)$, with $q = (q_x, q_y)$, and energies $\hbar\omega_Q$. The result is proportional to a weighted sum over $\alpha$, $\beta$, $k$ and $Q$ of the transition rate $P_Q^a(\alpha, k; \beta, k + q)$ for the adsorption of a phonon $Q$ by an electron going from state $(\alpha, k)$ to state $(\beta, k + q)$. The electronic matrix element involved in the transition rate is evaluated by using the electron–LA phonon adsorption potential for the phonon:

$$U_Q(R) = \mathrm{i}E_1(\hbar/2\rho V\omega_Q)^{1/2} Q\, \mathrm{e}^{\mathrm{i}Q\cdot R} \tag{2.5}$$

in the Golden Rule. Here $E_1$ is a spherically symmetric acoustic phonon deformation potential and $\rho$ and $V$ are the density and volume of the material. Ground sub-band occupation only is assumed and hence all subband labels and summations over them are dropped. When the sums over $k$ and $Q$ are transformed to integrals the final result for $S_g$ from LA phonons can be written in the form:

$$S_g = -\frac{\hbar v_S E_1^2 m^* g_v L}{8(2\pi)^3 k_B T^2 n_0 e\rho} \times \int\int \frac{q^3 Q^2 G(Q) |\int \varphi^*(z) \mathrm{e}^{\mathrm{i}q_z z} \varphi(z)\, \mathrm{d}z|^2}{\sinh^2(\hbar\omega_Q/2k_B T)}\, \mathrm{d}q\, \mathrm{d}q_z \tag{2.6}$$

where $m^*$ is the constant electron effective mass assumed for motion in the $xy$ plane, $g_v$ is the valley degeneracy and $v_s$ is the sound velocity for LA phonons. In (2.6) $G(Q)$ is defined by:

$$G(Q) = \int_{-\pi}^{\pi} \delta\left(\frac{\hbar^2 q^2}{2m^*} + \frac{\hbar^2 k_F q \cos\theta}{m^*} - \hbar\omega_Q\right) \mathrm{d}\theta \tag{2.7}$$

and has the value:

$$G(Q) = \frac{2m^*}{\hbar^2 q k_F} (1 - \alpha(Q)^2)^{-1/2} \tag{2.8a}$$

where:

$$\alpha(\boldsymbol{Q}) = \frac{2m^*\omega_Q - \hbar q^2}{2\hbar q k_{\mathrm{F}}} \tag{2.8b}$$

when the argument of the delta function in (2.7) vanishes for some $\theta$ and is zero otherwise. The condition for a non-zero result is $|\alpha(\boldsymbol{Q})| \leqslant 1$, which defines the field of integration in (2.6).

The anisotropy of Si causes an additional complication since both LA and TA modes contribute to $S_{\mathrm{g}}$. To deal with this Cantrell and Butcher (1987b) follow Ridley (1982) by taking $E_1$ inside the integrals in (2.6) and replacing it by $\Xi_u(q_z^2/Q^2 + D)$ and by $\Xi_u q q_z/Q^2$ for LA and TA modes respectively. Here $\Xi_{\mathrm{u}}$ is the deformation potential for pure shear strain and $D = \Xi_{\mathrm{d}}/\Xi_{\mathrm{u}}$ with $\Xi_{\mathrm{d}}$ denoting the deformation potential for pure dilation. The resultant $S_{\mathrm{g}}$ is then the sum of two contributions of the form (2.6): one for the LA phonon mode and one for the TA phonon mode. In their evaluation of (2.6) Cantrell and Butcher then remove the divergence in the integrand at $|\alpha(\boldsymbol{Q})| = 1$ (and the restriction of the field of integration) by replacing the delta function in (2.7) by a Lorentzian with a width small enough to make $S_{\mathrm{g}}$ independent of its precise value. The values of $S_{\mathrm{g}}$ calculated in this way were about 35 times what is observed. When we repeat the same calculations using the same parameter values we find improved agreement with experiment in that the calculated results are now only about 16 times what is observed. We have confirmed these new unscreened results by performing the calculations using (2.8) as it stands and a direct, numerical integration routine which takes particular care of the inverse square-root singularity in $G(\boldsymbol{Q})$ at the boundary of the integration field. We find agreement with our calculations using a Lorentzian approximation to the $\delta$-function, to better than 5% and conclude that the calculations of Cantrell and Butcher (1987b) lack accuracy in the numerical integration.

Determining the variational parameter $b$ in the Fang and Howard wavefunction (see § 5) provides an estimate of the correct ISW width to take in the calculations. Equating the position expectation value $\langle z \rangle$ in the two models gives $\delta \simeq 6/b$ where $b$ is determined by known parameters. We find that $\delta = 80$ Å is a much more representative value than the rough estimate of 20 Å used in the original calculations. $S_{\mathrm{g}}$ is a decreasing function of $\delta$ (Cantrell and Butcher 1987b) but changing $\delta$ from 20 Å to 80 Å only reduces $-S_{\mathrm{g}}$ by 10% because $S_{\mathrm{g}}$ is only weakly dependent on $\delta$ in the regime under discussion. For this reason it also seems unlikely that the use of a more realistic wavefunction than that for an ISW model would do much to bring about a qualitative reconciliation of theory and experiment at this stage. We therefore turn to the central concern of this paper (which is to introduce screening of the electron–phonon interaction) by considering the screening problem in the ISW model used previously.

## 3. Screening: general considerations

Static linear screening theory seeks to obtain the screened potential $V(\boldsymbol{R})$ by accounting for the redistribution of free charge in response to a weak bare potential $V^{\mathrm{b}}(\boldsymbol{R})$. In our case $V^{\mathrm{b}}(\boldsymbol{R}) = U_Q(\boldsymbol{R})$ in (2.5). To calculate $V(\boldsymbol{R})$ the perturbation in the electron density is best described by an expansion in the set of unperturbed electron states. For a homogeneous 3D free-electron system $V$ may then be related to $V^{\mathrm{b}}$ through the Lindhardt dielectric function (see for example, Ehrenreich and Cohen (1959)) in Fourier space $\varepsilon(\boldsymbol{Q})$:

$$\tilde{V}(\boldsymbol{Q}) = \tilde{V}^{\mathrm{b}}(\boldsymbol{Q})/\varepsilon(\boldsymbol{Q}) \tag{3.1}$$

where $\tilde{V}(\boldsymbol{Q})$ and $\tilde{V}^{\rm b}(\boldsymbol{Q})$ are the 3D Fourier transforms of $V(\boldsymbol{R})$ and $V^{\rm b}(\boldsymbol{R})$ respectively. However, this simple result is only obtained because of the complete translational invariance of the homogeneous system which is lost in the Q2DEG. In the strict 2D limit in which the 2DEG is confined to $z = 0$ in a 3D host of relative permittivity $\kappa$ we have a similar relationship between the 2D Fourier transforms of the bare and screened potentials on the plane $z = 0$. Thus $\boldsymbol{Q}$ is replaced by $\boldsymbol{q}$ in (3.1) with (Stern 1967):

$$\varepsilon(q) = 1 - V^{\rm c}\Pi(\boldsymbol{q}) = 1 + (q_s/q)\xi(q) \tag{3.2a}$$

where

$$\begin{aligned} \xi(q) &= 1 & q &\leqslant 2k_{\rm F} \\ &= 1 - (1 - (2k_{\rm F}/q)^2)^{1/2} & q &> 2k_{\rm F}. \end{aligned} \tag{3.2b}$$

Here $V^{\rm c} = {\rm e}^2/2\kappa\varepsilon_0 q$ is the 2D Fourier transform of the Coulomb interaction $e^2/4\pi\kappa\varepsilon_0 R$ on the plane $z = 0$. $\Pi(\boldsymbol{q})$ is the 2D polarisability defined in (3.10) with $\mu = \lambda = 1$ and the screening constant:

$$q_s = \frac{g_{\rm v} e^2 m^*}{2\pi\kappa\varepsilon_0\hbar^2}. \tag{3.3}$$

The matrix element whose square enters into the calculation of $P^{\rm a}_{Q}(\alpha, \boldsymbol{k}; \beta, \boldsymbol{k} + \boldsymbol{q})$ and $S_{\rm g}$ is $\langle\beta, \boldsymbol{k} + \boldsymbol{q}|U^{\rm a}_{Q}(\boldsymbol{R})|\alpha, \boldsymbol{k}\rangle$ with $U^{\rm a}_{Q}(\boldsymbol{R})$ given by (2.5). An estimate of the effect of screening this matrix element may be gained by considering the 2D limit for $U^{\rm a}_{Q}(\boldsymbol{R})$ by putting $z = 0$. Then the 2D Fourier transform of the bare potential is clearly given by:

$$\tilde{V}^{\rm b}(\boldsymbol{q}) = {\rm i}E_1(\hbar/2\rho V\omega_Q)^{1/2}Q. \tag{3.4}$$

The dominant contributions to $S_{\rm g}$ come from near $q = 2k_{\rm F}$. We therefore estimate the effect of screening on $S_{\rm g}$ by evaluating $\varepsilon^{-2}(q)$ from (3.2) when $q = 2k_{\rm F}$ with $n_0 = 9.8 \times 10^{15}\ {\rm m}^{-2}$ (the largest value in the data of Gallagher *et al* 1987). We find that $S_{\rm g}$ can be expected to be reduced by a factor of approximately 0.05. For GaAs the corresponding reduction factor is only 0.5. On the basis of these elementary considerations we expect screening to play a dominant role in determining the absolute magnitude of $S_{\rm g}$ in the MOSFET data as suggested by Gallagher *et al* (1987) and a much less important but still significant role in the heterojunction data because of the relative values of $q_{\rm s}$ in Si and GaAs.

In quasi-2D the finite well width must be allowed for. Translational invariance is retained only in the $xy$ plane. Moreover, since the $z$ dependence of the perturbed electron density is conveniently expanded in the set $\varphi_\alpha(z)$, all the sub-bands are involved in the screening whether they are occupied or not. These ideas lead to the result of Siggia and Kwok (1970) that:

$$\tilde{V}^{\rm b}_{\alpha\beta}(\boldsymbol{q}) = \sum_{\mu,\lambda} \varepsilon_{\alpha\beta\mu\lambda}(\boldsymbol{q})\tilde{V}_{\mu\lambda}(\boldsymbol{q}) \tag{3.5}$$

where

$$\tilde{V}_{\alpha\beta}(\boldsymbol{q}) = \langle\alpha, \boldsymbol{k} + \boldsymbol{q}|V(\boldsymbol{R})|\beta, \boldsymbol{k}\rangle \tag{3.6}$$

and is equal to the matrix element between sub-bands $\alpha$ and $\beta$ of the 2D Fourier transform of $V(\boldsymbol{R})$ with $z$ held fixed. The quantity $\tilde{V}^{\rm b}_{\alpha\beta}(\boldsymbol{q})$ has the same interpretation in terms of $V^{\rm b}(\boldsymbol{R})$. The bare and screened matrix elements are related by (3.5) which may be written in the more transparent form:

$$\tilde{V}^{\rm b} = \varepsilon\tilde{V} \tag{3.7}$$

where $\tilde{V}^{\mathrm{b}}$ and $\tilde{V}$ are vectors whose elements are the matrix elements in (3.5) in some chosen order and $\varepsilon$ is the dielectric superoperator matrix. Thus, taking account of only two sub-bands, 1 and 2, we have:

$$\begin{bmatrix} V^{\mathrm{b}}_{11} \\ V^{\mathrm{b}}_{12} \\ V^{\mathrm{b}}_{21} \\ V^{\mathrm{b}}_{22} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1111} & \varepsilon_{1112} & \varepsilon_{1121} & \varepsilon_{1122} \\ \varepsilon_{1211} & \varepsilon_{1212} & \varepsilon_{1221} & \varepsilon_{1222} \\ \varepsilon_{2111} & \varepsilon_{2112} & \varepsilon_{2121} & \varepsilon_{2122} \\ \varepsilon_{2211} & \varepsilon_{2212} & \varepsilon_{2221} & \varepsilon_{2222} \end{bmatrix} \begin{bmatrix} V_{11} \\ V_{12} \\ V_{21} \\ V_{22} \end{bmatrix}. \tag{3.8}$$

The $\boldsymbol{q}$ dependence of all elements has been left as understood for clarity and for $n$ sub-bands $\varepsilon$ is thus an $n^2 \times n^2$ matrix whose elements are:

$$\varepsilon_{\alpha\beta\mu\lambda}(\boldsymbol{q}) = \delta_{\mu,\alpha}\delta_{\lambda,\beta} - V^{\mathrm{c}}F_{\alpha\beta\mu\lambda}(\boldsymbol{q})\Pi_{\mu\lambda}(\boldsymbol{q}). \tag{3.9}$$

In (3.9) the multi-sub-band polarisability is:

$$\Pi_{\mu\lambda}(\boldsymbol{q}) = \lim_{\eta\to 0}\frac{2g_{\mathrm{v}}}{A}\sum_{k}\frac{f_\lambda(\boldsymbol{k}) - f_\mu(\boldsymbol{k}+\boldsymbol{q})}{\varepsilon_\lambda(\boldsymbol{k}) - \varepsilon_\mu(\boldsymbol{k}+\boldsymbol{q}) + \mathrm{i}\eta} \tag{3.10}$$

and is given in the Appendix. We note here that $\Pi_{\mu\lambda}(\boldsymbol{q}) = 0$ for $\mu, \lambda > 1$ for the case of interest to us in which only the ground sub-band, with $\mu = 1$, is occupied. The form factor is given by:

$$V^{\mathrm{c}}F_{\alpha\beta\mu\lambda}(\boldsymbol{q}) = \int \tilde{g}(\boldsymbol{q}, z, z')\varphi^*_\alpha(z)\varphi_\beta(z)\varphi_\mu(z')\varphi^*_\lambda(z')\,\mathrm{d}z\,\mathrm{d}z' \tag{3.11}$$

where $\tilde{g}(\boldsymbol{q}, z, z')$ is the 2D Fourier component of the net Coulomb interaction between electrons at $z$ and $z'$. This quantity satisfies:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left[\kappa(z)\frac{\mathrm{d}}{\mathrm{d}z}\tilde{g}(\boldsymbol{q}, z, z')\right] - q^2\kappa(z)\tilde{g}(\boldsymbol{q}, z, z') = -\frac{e^2}{\varepsilon_0}\delta(z - z') \tag{3.12}$$

in the general case and is just $V^{\mathrm{c}}\,\mathrm{e}^{-q|z-z'|}$ when $\kappa(z)$ is a constant $\kappa$. In a MOSFET in which the Q2DEG resides in Si of relative permittivity $\kappa_2$ at the interface with $SiO_2$ of relative permittivity $\kappa_1$, we set $z = 0$ at the junction so that:

$$\kappa(z) = \begin{cases} \kappa_1 & z < 0 \\ \kappa_2 & z \geqslant 0. \end{cases} \tag{3.13}$$

In this case:

$$\tilde{g}(\boldsymbol{q}, z, z') = \tfrac{1}{2}V^{\mathrm{c}}\{(1 + \kappa_1/\kappa_2)\,\mathrm{e}^{-q|z-z'|} + (1 - \kappa_1/\kappa_2)\,\mathrm{e}^{-q|z+z'|}\} \tag{3.14}$$

for $z \geqslant 0$, which is the standard result when the average of $\kappa_1$ and $\kappa_2$ replaces $\kappa$ in $V^{\mathrm{c}}$ (Ando *et al* 1982).

The case of the pure 2DEG is treated by letting $\delta \to 0$. In this limit the denominators of all the $\Pi_{\mu\lambda}(\boldsymbol{q})$ become very large except for $\Pi_{11}(q)$ and only the case $\alpha = \beta = 1$ is of interest. Hence only $\varepsilon_{1111}(\boldsymbol{q})$ is significant. Moreover, the form factor $F_{1111}(\boldsymbol{q})$ is the average of $\tilde{g}(q, z, z')/V^{\mathrm{c}}$ over the electron positions in the ground sub-band and tends to unity in the 2D limit. Thus the 2D result embodied in (3.2) and (3.3) is recovered. In quasi-2D when $\delta$ remains very small $\Pi_{11}(\boldsymbol{q})$ still dominates but $F_{1111}(\boldsymbol{q}) \leqslant 1$. Retaining only those terms with $\mu = \lambda = 1$ in (3.5) and evaluating $F_{1111}(\boldsymbol{q})$ gives the SSA to MSS with the quasi-2D dielectric function:

$$\varepsilon_s(\boldsymbol{q}) = 1 - V^{\mathrm{c}}\Pi(\boldsymbol{q})F_{1111}(\boldsymbol{q}) = 1 + q_s\xi(\boldsymbol{q})F_{1111}(\boldsymbol{q})/q. \tag{3.15}$$

The SSA is commonly used (Ando *et al* 1982) because of its simplicity. However, it will break down when $\delta$ is sufficiently large. In the next section we estimate the error involved in using it in the cases of interest to us by making calculations for the ISW model. Fortunately they turn out to be small so that the major complications inherent in the MSS equation (3.5) may be avoided in our calculations. The quantities required are given in the Appendix.

## 4. Screening: calculations

Using the matrix notation of (3.7) the MSS equation can be formally inverted so that (3.5) can be written in the form:

$$\tilde{V}_{\alpha\beta}(\boldsymbol{q}) = \sum_{\mu,\lambda} \varepsilon^{-1}_{\alpha\beta\mu\lambda}(\boldsymbol{q})\tilde{V}^{\mathrm{b}}_{\mu\lambda}(\boldsymbol{q}) \tag{4.1}$$

where $\varepsilon^{-1}_{\alpha\beta\mu\lambda}(\boldsymbol{q})$ is an element of the inverse of the dielectric superoperator matrix. By expanding the bare potential matrix elements $\tilde{V}^{\mathrm{b}}(\boldsymbol{q}, z)$ as a Fourier series in $q_z$ where

$$\tilde{V}^{\mathrm{b}}_{\mu\lambda}(\boldsymbol{q}) = \langle\mu|\tilde{V}^{\mathrm{b}}(\boldsymbol{q}, z)|\lambda\rangle = \sum_{q_z} \tilde{V}^{\mathrm{b}}(\boldsymbol{Q})\langle\mu|\,\mathrm{e}^{\mathrm{i}q_z z}|\lambda\rangle \tag{4.2}$$

and by similarly expanding $\tilde{V}(\boldsymbol{q}, z)$ we obtain:

$$\tilde{V}_{\alpha\beta}(\boldsymbol{q}) = \sum_{q_z} \tilde{V}(\boldsymbol{Q})\langle\alpha|\,\mathrm{e}^{\mathrm{i}q_z z}|\beta\rangle = \sum_{q_z} \frac{\tilde{V}^{\mathrm{b}}(\boldsymbol{Q})\langle\alpha|\,\mathrm{e}^{\mathrm{i}q_z z}|\beta\rangle}{\varepsilon_{\alpha\beta}(\boldsymbol{Q})}. \tag{4.3}$$

In (4.3) we have defined the effective MSS dielectric function for screening $\tilde{V}_{\alpha\beta}(\boldsymbol{q})$ as:

$$\varepsilon_{\alpha\beta}(\boldsymbol{Q}) = \langle\alpha|\,\mathrm{e}^{\mathrm{i}q_z z}|\beta\rangle \Big/ \sum_{\mu\lambda} \varepsilon^{-1}_{\alpha\beta\mu\lambda}(\boldsymbol{q})\langle\mu|\,\mathrm{e}^{\mathrm{i}q_z z}|\lambda\rangle. \tag{4.4}$$

Then (4.3) has a similar structure to (3.1) but relates to a particular matrix element of the potential and involves a sum over $q_z$. Equations (4.3) and (4.4) provide a convenient reformulation of the MSS equations. In particular we see that the SSA is obtained by ignoring the dependence of $\varepsilon_{\alpha\beta}(\boldsymbol{Q})$ on $\alpha$, $\beta$ and $q_z$, replacing it by $\varepsilon_{\mathrm{s}}(\boldsymbol{q})$ in (3.15) and taking it outside the summation sign.

The calculation of $S_{\mathrm{g}}$ involves $|\tilde{U}^{\mathrm{a}}_{11}(\boldsymbol{q})|^2$ and we see from (2.5) that only the particular phonon component $q_z$ is present on the RHS of (4.3). The factor:

$$f_{\alpha\beta}(\boldsymbol{Q}) = \varepsilon_{\mathrm{s}}(\boldsymbol{q})^2/|\varepsilon_{\alpha\beta}(\boldsymbol{Q})|^2. \tag{4.5}$$

then measures the change in the contribution to $S_{\mathrm{g}}$ from the phonon with wavevector $\boldsymbol{Q}$ effected by replacing the SSA by MSS.

We have made calculations of $f_{11}(\boldsymbol{Q})$ for $n = 1, 2, 3, \ldots$, sub-bands until convergence was achieved. In all cases $n \leqslant 5$ was sufficient and for some cases only three sub-bands were necessary. All the quantities $\varepsilon_{\mathrm{s}}(\boldsymbol{q})$, $\varepsilon_{\alpha\beta}(\boldsymbol{Q})$ and $f_{\alpha\beta}(\boldsymbol{Q})$ were calculated in the ISW for $\delta$ ranging from 1 to 300 Å with $n_0 = 10^{15}$, $10^{16}$ and $10^{17}\ \mathrm{m}^{-2}$ and $q$ and $q_z$ between 0 and $10\,k_{\mathrm{F}}$. At $n_0 = 10^{15}\ \mathrm{m}^{-2}$ we find that the SSA is good for all $\delta$. Even for large $\delta$ (200–300 Å) deviations from MSS > 1% are only just beginning to appear when $q_z > 4k_{\mathrm{F}}$. These deviations decrease for any $q_z$ with increasing $q$ but increase with increasing $q_z$ at given $q$. At $n_0 = 10^{16}\ \mathrm{m}^{-2}$ they remain $\leqslant$1% for all $q$, $q_z$ when $\delta \leqslant 50$ Å but for $\delta = 200$ and 300 Å the deviations increase to 10% even at $q_z \simeq k_{\mathrm{F}}$ for small $q$. Hence, as expected, the SSA is becoming inaccurate for large $\delta$, large $q_z$ and large $n_0$. However, in the regime

of interest to us ($n_0 < 10^{16}\,\mathrm{m}^{-2}$ and $\delta \simeq 20$–100 Å) the SSA is adequate because the dominant contributions to $S_g$ come from $|Q| < 2k_F$. This is confirmed even at the highest experimental value of $n_0$ ($9.8 \times 10^{15}\,\mathrm{m}^{-2}$) where MSS effects are largest: over the temperature range 2–6 K the difference between using the SSA and MSS in the calculation of $S_g$ is negligible when $\delta = 20$ Å and is only 0.2% when $\delta = 80$ Å. Tang (1988) has recently discussed MSS in GaAs/AlGaAs heterojunctions.

## 5. Screened $S_g$ calculations

We have evaluated the effects of including the SSA to MSS in the calculation of $S_g$ for the ISW model used by Cantrell and Butcher (1987b). The difference between theory and experiment comes down from a factor of approximately 16 too large to approximately 40% too small. This is a considerable improvement over the previous theoretical results and warrants closer treatment of the sub-band wavefunctions because the information discarded by taking an ISW of constant width may now be comparable with the difference between theory and experiment.

The Fang and Howard wavefunction

$$\varphi^{\mathrm{b}}(z) = (b^3/2)^{1/2} z\, \mathrm{e}^{-bz/2} \qquad z \geqslant 0 \tag{5.1}$$

still neglects electron penetration into the oxide $z < 0$ but the parameter $b$ may be determined for each $n_0$. This sub-band wavefunction should therefore provide a much better description of the dependence of $S_g$ upon $n_0$ (see Ando *et al* 1982). To determine $b$ a variational calculation is performed:

$$\frac{\mathrm{d}}{\mathrm{d}b} E(b) = 0 \tag{5.2a}$$

where

$$E(b) = \langle T_z \rangle + \langle V_{\mathrm{ext}}(z) \rangle + \tfrac{1}{2}\langle V_{\mathrm{s}}(z) \rangle + \langle V_{\mathrm{xc}}(z) \rangle \tag{5.2b}$$

and is the average many-body energy per electron which is the sum of contributions from the kinetic energy $T_z$, the external potential $V_{\mathrm{ext}}(z)$, the Hartree self-energy $V_{\mathrm{s}}(z)$, and the local density approximation to the exchange and correlation energy $V_{\mathrm{xc}}(z)$. The averages in (5.2*b*) are all evaluated using the wavefunction (5.1). The kinetic energy associated with motion in the $xy$ plane is ignored because it is independent of $b$. By neglecting the image, exchange–correlation, and quadratic term in the depletion energy we obtain the familiar analytic result:

$$b^3 = \frac{12 m_z^* \,\mathrm{e}^2}{\kappa_2 \varepsilon_0 \hbar^2} \left(N_{\mathrm{depl}} + \frac{11}{32} n_0\right). \tag{5.3}$$

In this equation $m_z^*$ is the effective mass in the $z$ direction and $N_{\mathrm{depl}}$ is the areal density of acceptors in Si. It is supposed that there are $N_{\mathrm{A}}$ acceptors per unit volume in the Si and that they are completely ionised over the depletion layer $0 < z < d$. Then we have:

$$N_{\mathrm{depl}} = N_{\mathrm{A}} d = (2\varphi_{\mathrm{d}} \kappa_2 \varepsilon_0 N_{\mathrm{A}}/e)^{1/2} \tag{5.4}$$

where the electrostatic band bending $\varphi_{\mathrm{d}}$ is given approximately by equating the electron energy $e\varphi_{\mathrm{d}}$ to the Si band gap (1.12 eV). We use equations (5.3) and (5.4), which are known to compare well with self-consistent calculations (Ando *et al* 1982), in what follows.

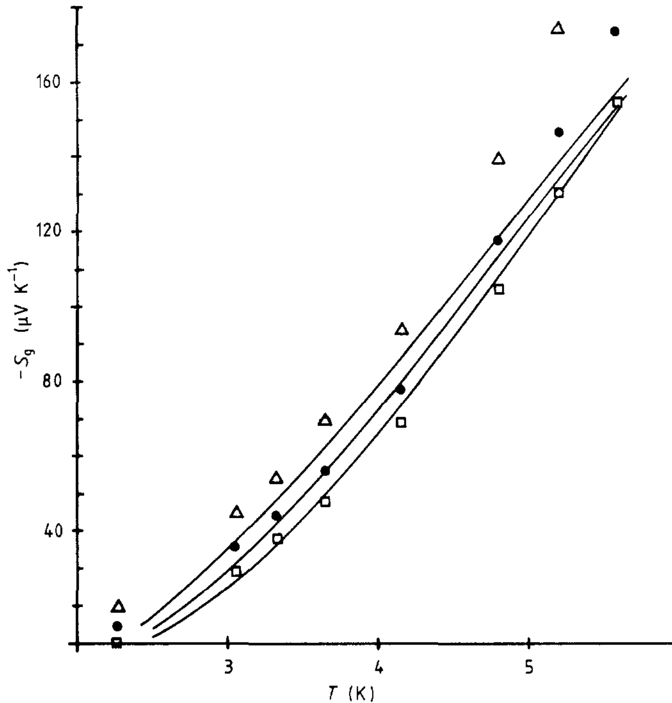


**Figure 1.** A plot of $-S_g$ against temperature for the three largest experimental values of electron density, full curves are theoretical results and the symbols are the experimental points of Gallagher *et al* (1987). Electron densities in units of $10^{15}$ m$^{-2}$ are 9.8 (□), 7.8 (●) and 6.1 (△).

Taking $b$ from (5.3) and (5.4) we have used $\varphi^{b}(z)$ to recalculate $S_g$. We find the screened variational thermopowers are between 5 and 30% from the measured values. When we use better material parameter values than those used by Cantrell and Butcher (1987b) we find that the agreement becomes very good indeed. We assume the commonly accepted values for $\Xi_u$ and $\Xi_d$ of 9.0 and $-6.0$ eV (Ando *et al* 1982); we take $v_L$ and $v_T$ from the standard reference (*Landolt–Börnstein, New Series* 1982) and average over the phonon modes of appropriate propagation and lattice displacement directions, to obtain $8.831 \times 10^3$ and $5.281 \times 10^3$ ms$^{-1}$ respectively, and we take $m^*$ at $0.1905m_0$ and $m_z^*$ at $0.916m_0$. The values taken by Cantrell and Butcher (1987b) were: 8.0, 1.6 eV; $8.5 \times 10^3$, $5.0 \times 10^3$ ms$^{-1}$ and $0.2m_0$ respectively. Plots of $-S_g$ against $T$ calculated in this way are compared with the experimental data points of Gallagher *et al* (1987) in figure 1 for the highest electron densities $n_0 = 9.8$, 7.8 and $6.1 \times 10^{15}$ ms$^{-1}$. Both the experimental and theoretical values of $-S_g$ increase when $n_0$ decreases. The agreement with experiment at 4 K is $-5\%$, $-4\%$ and 6.8% respectively.

Figure 2 shows plots of $-S_g/T^3$ against $T$ derived from those of figure 1. They exhibit the rather flat maxima found previously (Cantrell and Butcher 1987b, Gallagher *et al* 1987) which arise from the coincidence of the dominant $q$-value with the Kohn resonance at $2k_F$. However, the maxima are now all moved up in temperature by about 0.75 K as compared with those found in the original calculations. This is a direct result of introducing screening. As $T$ increases so does the dominant $q$-value and consequently the effectiveness of screening in reducing $S_g$ falls off. It would therefore appear that the close agreement between the experimental peak positions and those predicted by the

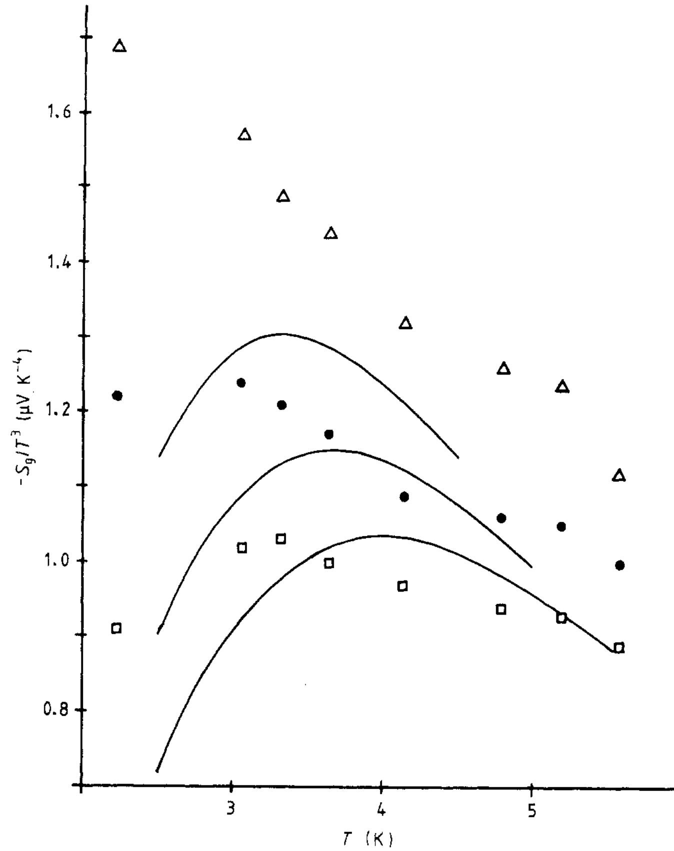


**Figure 2.** A plot of $-S_g/T^3$ against temperature for the same data in figure 1.

unscreened $S_g$ calculations was fortuitous. We note that our treatment of the 3D phonons as having spherically symmetric scalar deformation potentials and constant (averaged) sound velocities is a little cavalier. We await both better treatments of the phonons and further experimental data to resolve the discrepancy in the peak positions (in figure 2). Merely taking different $\Xi_u$, $\Xi_d$, $v_L$ and $v_T$ is insufficient. This affects $|S_g|$ and has little effect on the peak position compared with the 0.75 K displacement noted here. Treating $\varphi^b(z)$ more closely by retaining all the terms in (5.2*b*) is only slightly more inconvenient because (5.2*a*) is then solved for *b* numerically. However, the expected change in *b* is small and shifts the peak position only slightly. Halving or doubling *b* for example causes a shift $< 0.2$ K.

Finally, we note that peaks in both experimental and theoretical plots of $-S_g/T^3$ against $T$ confirm that there is no pure $T^3$ dependence in $S_g$ even at the highest (i.e. most metallic) electron densities. The underlying form is $T^3$ as pointed out by Gallagher *et al* (1987) but there is also the enhancement through the divergence in the unbroadened integrand of (2.6) around $q \simeq 2k_F$ superimposed. Clearly $S_g$ increases faster than $T^3$ below the peak and more slowly above. This may explain the smooth departure from linearity of $S/T$ taken as $a + bT^2$ and plotted against $T^2$, seen in a GaAs/GaAlAs heterojunction by Ruf *et al* (1988).

## 6. Conclusions

The unscreened $S_g$ formula of Cantrell and Butcher (1987*a*) leads to thermopowers approximately 16 times larger than experiment and not approximately 35 as previously thought. Introducing broadening to remove the singularity in the integrand of the $S_g$ formula has only a small quantitative effect. Screening has a profound effect, however, particularly in Si where the screening constant is much larger than in GaAs due to the valley degeneracy and larger effective mass. This explains why the unscreened results agree better with experiment in GaAs than in Si. The excellent qualitative agreement with experimental thermopower measurements is preserved in the case of Si when screening is introduced. In adopting single-sub-band screening, which has been shown to be sufficient in the experimental regime, the quantitative difference is just $-40\%$ even in the constant infinite square-well model adopted previously by Cantrell and Butcher (1987*b*). The agreement with experimental data is improved considerably by adopting a variational wavefunction and better values for the material parameters. The best agreement with experiment is found to be $\leqslant 5\%$ at temperatures in the middle of the experimental regime for the highest electron densities. A small shift to higher temperatures is noted in the peak position in graphs of $-S_g/T^3$ versus $T$, which should be considered further by an improved treatment of the phonons in the light of additional experimental results.

### Acknowledgments

The authors wish to thank B L Gallagher and D G Cantrell for helpful discussions.

## Appendix

The multi-sub-band polarisability $\Pi_{\mu\lambda}(\boldsymbol{q})$ is calculated directly from (3.10) which can be written as

$$\Pi_{\mu\lambda}(\boldsymbol{q}) = \pi_1(\omega) - \pi_2(\omega) \tag{A1}$$

where

$$\pi_1(\omega) = \frac{2g_v}{A} \lim_{\eta\to 0} \sum_{\boldsymbol{k}} \frac{f_\lambda(\boldsymbol{k})}{\varepsilon(\boldsymbol{k}) - \varepsilon(\boldsymbol{k}+\boldsymbol{q}) - \omega} \tag{A2}$$

and

$$\pi_2(\omega) = \frac{2g_v}{A} \lim_{\eta\to 0} \sum_{\boldsymbol{k}} \frac{f_\mu(\boldsymbol{k}+\boldsymbol{q})}{\varepsilon(\boldsymbol{k}) - \varepsilon(\boldsymbol{k}+\boldsymbol{q}) - \omega}$$

and $\omega$ is $\varepsilon_\mu - \varepsilon_\lambda - \mathrm{i}\eta$ with $\varepsilon_i$ the $i$th sub-band energy. We find $\pi_2(\omega)$ is $-\pi_1(-\omega)$ and $\pi_1(\omega)$ is evaluated by transforming to an integral over $\boldsymbol{k}$ up to $|\boldsymbol{k}| = k_F$ in the electric quantum limit. Here $f_\lambda(\boldsymbol{k})$ is zero for $\lambda > 1$ or $\lambda = 1$ and $|\boldsymbol{k}| > k_F$ and we have

$$\pi_1(\omega) = \pm \frac{g_v m^* \delta_{\lambda,1}}{2\pi} \left\{ \left[ \left(1 + \frac{2m^*\omega}{q^2}\right)^2 - \left(\frac{2k_F}{q}\right)^2 \right]^{1/2} - \left[ \left(1 + \frac{2m^*\omega}{q}\right)^2 \right]^{1/2} \right\} \tag{A3}$$

Hence

$$\Pi(\boldsymbol{q}) = \frac{-g_v m^* \xi(q)}{\pi\hbar^2} \tag{A4}$$

is the 2DEG polarisability of (3.2). The sign is determined by the limit $q \to 0$ obtained by taking expansions of $f(\boldsymbol{k}+\boldsymbol{q})$ and $\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}$ about $f(\boldsymbol{k})$ and $\varepsilon(\boldsymbol{k})$. The final result for $\lambda = 1$ and $\mu > 1$ is then

$$\Pi_{\mu,1}(\boldsymbol{q}) = -\frac{g_v m^*}{2\pi\hbar^2}\left\{\left(1+\frac{2m^*\varepsilon'_\mu}{\hbar^2 q^2}\right)^2 - \left[\left(1+\frac{2m^*\varepsilon'_\mu}{\hbar^2 q^2}\right)^2 - \left(\frac{2k_F}{q}\right)^2\right]^{1/2}\right\} \tag{A5}$$

where $\varepsilon'_\mu$ is $\varepsilon_\mu$ measured from the ground sub-band energy $\varepsilon_1$. The quantity $\Pi_{\mu\lambda}(\boldsymbol{q})$ is symmetrical in $\mu$ and $\lambda$ and when both are greater than unity it vanishes when only the ground sub-band is occupied. The general result is given by Mori and Ando (1979).

The form factor $F_{\alpha\beta\mu\lambda}(\boldsymbol{q})$ is calculated from the definition (3.11) and the result (3.14) taking

$$\varphi_\alpha(z) = (2/\delta)^{1/2}\sin(\alpha\pi z/\delta) \qquad \alpha = 1, 2, 3 \ldots \tag{A6}$$

We find that

$$F_{\alpha\beta\mu\lambda}(\boldsymbol{q}) = F_1 + F_2 \tag{A7}$$

where

$$F_1 = \frac{2}{\delta^2}\left(1 - \frac{\kappa_1}{\kappa_2}\right) I(\alpha, \beta, \boldsymbol{q}) I(\mu, 1, \boldsymbol{q}) \tag{A8}$$

and

$$F_2 = \frac{q}{\delta^2}\left(1 + \frac{\kappa_1}{\kappa_2}\right)[G(p_{\alpha\beta}) - G(p'_{\alpha\beta})] \tag{A9}$$

in which

$$G(p_{\alpha\beta}) = \frac{1}{p_{\alpha\beta}^2 + q^2}\left[\tfrac{1}{2}\Delta_{\alpha\beta\mu\lambda} - I(\mu, 1, \boldsymbol{q}) - \cos(p_{\alpha\beta}\delta)\, \mathrm{e}^{-q\delta} I(\mu, 1, -q)\right] \tag{A10}$$

with

$$p_{\alpha\beta} = (\alpha - \beta)\pi/\delta \qquad p'_{\alpha\beta} = (\alpha + \beta)\pi/\delta$$

$$\Delta_{\alpha\beta\mu\lambda} = \delta_{\alpha-\beta,\mu-\lambda} + \delta_{\alpha-\beta,\mu+\lambda} + \delta_{\alpha-\beta,\mu+\lambda} + \delta_{\alpha-\beta,-\mu-\lambda}$$

$$I(\alpha, \beta, \boldsymbol{q}) = \tfrac{1}{2}[I'(p_{\alpha\beta}) - I'(p'_{\alpha\beta})]$$

and

$$I'(p_{\alpha\beta}) = \frac{q}{p_{\alpha\beta}^2 + q^2}[1 - \mathrm{e}^{-q\delta}\cos(p_{\alpha\beta}\delta)]. \tag{A11}$$

For the matrix element $\langle\alpha|\mathrm{e}^{\mathrm{i}q_z z}|\beta\rangle$ the result is

$$M_{\alpha\beta}(q_z) \equiv \int_0^\delta \varphi_\alpha^*(z)\, \mathrm{e}^{\mathrm{i}q_z z}\varphi_\beta(z)\, \mathrm{d}z = M^R_{\alpha\beta}(q_z) + \mathrm{i}M^I_{\alpha\beta}(q_z) \tag{A12}$$

where $M^R_{\alpha\beta}(q_z)$ and $M^I_{\alpha\beta}(q_z)$ are both real with

$$M^R_{\alpha\beta}(q_z) = m_2 \sin\theta_2 - m_1 \sin\theta_1 \tag{A13}$$

and

$$M^I_{\alpha\beta}(q_z) = m_2(1 - \cos\theta_2) - m_1(1 - \cos\theta) \tag{A14}$$

in which

$$m_1 = \frac{q_z^2}{\delta(p_{\alpha\beta}^2 - q_z^2)} \qquad m_2 = \frac{q^2}{\delta(p_{\alpha\beta}'^2 - q_z^2)} \tag{A15}$$

and

$$\theta_1 = (\alpha - \beta)\pi/\delta + q_z\delta \qquad \theta_2 = (\alpha + \beta)\pi/\delta + q_z\delta. \tag{A16}$$

The symbols $p_{\alpha\beta}$ and $p'_{\alpha\beta}$ are defined in (A11).

## References


Ando T, Fowler A B and Stern F 1982 *Rev. Mod. Phys.* **54** 437
Blatt F J 1968 *Physics of Electronic Conduction in Solids* (New York: McGraw-Hill)
Cantrell D G and Butcher P N 1987a *J. Phys. C: Solid State Phys.* **20** 1985
—— 1987b *J. Phys. C: Solid State Phys.* **20** 1993
Ehrenreich H and Cohen M H 1959 *Phys. Rev.* **115** 786
Fletcher R, D'Iorio M, Sachrajda A S, Stoner R, Foxon C and Harris J J 1988 *Phys. Rev.* B **37** 3137
Fletcher R, Maan J C, Ploog K and Weimann G 1986 *Phys. Rev.* B **33** 7122
Gallagher B L, Gibbings C J, Pepper M and Cantrell D G 1987 *Semicond. Sci Technol.* **2** 456
*Landolt–Börnstein, New Series* 1982 Group III, vol. 17a (Berlin: Springer)
Mori S and Ando T 1979 *Phys. Rev.* B **19** 6433
Ridley B K 1982 *Quantum Processes in Semiconductors* (Oxford: Clarendon)
Ruf C, Obloh H, Junge B, Gmelin E, Ploog K and Weimann G 1988 *Phys. Rev.* B **37** 6377
Siggia E D and Kwok P C 1970 *Phys. Rev.* B **2** 1024
Stern F 1967 *Phys. Rev. Lett.* **18** 546
Tang D S 1988 *Phys. Rev.* B **37** 8319