Explanation of quantized-Hall-resistance plateaus in heterojunction inversion layers

G. A. Baraff and D. C. Tsui
Bell Laboratories, Murray Hill, New Jersey 07974
(Received 9 March 1981; revised manuscript received 12 June 1981)

A self-consistent calculation of n/B (inversion layer carrier density divided by magnetic field strength) vs $1/B$ exhibits quantized values over finite ranges of B (plateaus), just as seen in the Hall-resistance measurements of Tsui and Gossard. Electrons here in the inversion layer come from the ionized donors which, because of band bending, have a continuous energy density of states. These states fill or empty as the energy of Landau levels sweeps past them, producing the plateaus.

Quantization of the Hall resistance of the two-dimensional electron gas (2DEG) has recently been observed in the inversion layers of Si MOSFET's (metal-oxide-semiconductor field-effect transistors)[1] and GaAs-$Al_xGa_{1-x}As$ heterojunctions.[2] This new quantum effect has its origin in the fact that magnetic quantization splits the energy spectrum of the 2DEG into discrete Landau levels, each having a degeneracy $\beta = eB/h$. (Here, B is the magnetic field perpendicular to the 2DEG, h/e is the flux quantum, and we neglect spin and valley degeneracies.) When an integral number N_0 of Landau levels are filled, scattering cannot take place, owing to the presence of a gap between the filled and empty Landau levels. Under this condition the diagonal conductivity σ_{xx} vanishes and the off-diagonal Hall conductivity $\sigma_{xy} = ne/B$. Since $n = N_0\beta$ is the total number of electrons needed to fill the N_0 Landau levels, $\sigma_{xy} = N_0 e^2/h$, and the Hall resistance is $\rho_{xy} = h/N_0 e^2$. The Hall resistance has this special value only when there is no partially occupied Landau level in the system, i.e., when there is no Landau level at the Fermi energy.

This zero-order explanation is incomplete: It conceals two unresolved and unrelated problems. First, why does σ_{xy} equal $N_0 e^2/h$ to such high accuracy in the presence of, e.g., electron-electron interactions, edge effects, potential fluctuations, etc.? Several theoretical models deal with this aspect of the problem.[3–6] The present work does not. The second problem, which we address here, is why the quantized value of σ_{xy} is observed to persist over finite ranges of n or B, i.e., to exhibit a plateau. It is clear that for $\sigma_{xy} = ne/B$ to exhibit a plateau as a function of B, n must vary with B; i.e., there must be a reservoir to supply particles to the 2DEG.

In this Communication, we show that for the heterojunctions where the 2DEG is supposedly isolated, there are ionized donors at the right distance to serve as a reservoir. Our calculation of the potential at GaAs-$Al_xGa_{1-x}As$ heterojunctions, such as those studied in Ref. 2, shows that at doping concentrations appropriate for producing a 2DEG, the potential barrier on the n side of the junction is so low and thin that tunneling between the ionized donors and the inversion layer is rapid. These two systems are in equilibrium with each other even at low temperature. The number of carriers in the inversion layer then depends on B as the energy of the Landau levels sweeps past that of the donors, and it must be calculated self-consistently. This we do, first using Hartree theory. The calculated n/B exhibits the finite plateaus seen in the experiments.

The most important property of the reservoir influencing the width of the plateau is its spatial distance from the 2DEG. This is because motion of a Landau level relative to the Fermi energy E_F occurs for two reasons: Firstly, its magnetic energy $(N_0 + \frac{1}{2})\hbar\omega_c$ relative to the potential in the 2DEG rises with B. Secondly, the potential in the 2DEG relative to E_F also changes with B because the electrons needed to keep the levels full transfer in or out from the reservoir which establishes E_F. The more distant is that reservoir, the more rapidly does the potential in the 2DEG change for each electron transferred. When there is a Landau level *at* E_F, then an increase in B raises its magnetic energy and causes it to empty, transferring electrons back to the reservoir and lowering thereby the potential in the 2DEG. This potential lowering *holds* the partially filled level exactly at E_F, until the emptying is complete, whereupon the empty level rises above E_F. There are now no partially filled Landau levels and a new plateau starts. An increase of B now increases the number of carriers each filled Landau level can hold, and carriers transfer into the 2DEG from the reservoirs. The potential in the 2DEG rises, the magnetic energy of each Landau level also arises, and the topmost filled Landau level reaches E_F, which puts an end to the plateau. The cycle can then repeat. When one assumes that the ionized donors provide the only reservoir to the heterojunctions, the plateau widths, for reasonable values of the material parameters, are about 22% less than what is observed in the heterojunctions.

Other effects can alter the width of the plateaus. For example, a mobility edge within the broadened Landau levels can provide a reservoir of localized states. These would be an especially useful reservoir because, being located at the 2DEG, its emptying and filling would *not* alter the potential in the 2DEG. We shall not complicate the model by considering this reservoir even though it may well be present.

Another plausible process which can affect the plateau width is the dependence of the self-energy of the topmost filled Landau level on its occupation. This will change the rate at which that level empties when the field is increased. We postulate that the self-energy of the electrons in the partially filled Landau level, although influenced by the number of electrons in the lower filled levels, depends strongly on the degree to which the topmost level is filled, becoming more negative as that level fills. This is reminiscent of the explanation of the high-field oscillatory exchange enhancement of the g factor, and indeed, the self-energy used by Ando and Uemura[7] can be regarded as showing such a dependence. We propose a simple one-parameter form to express this dependence on filling. Including this self-energy in our calculation enhances the plateau width.

We now describe the model and the calculations. The heterojunction is depicted in Fig. 1. On the GaAs side $(x > 0)$, the net acceptor concentration N_A is low. When the junction is in equilibrium, the Fermi energy E_F is near the bulk GaAs valence band edge. Band bending fills some acceptors, establishing an immobile charge density N_A (in units of the electron charge $-e$) for $0 < x < L_A$. In the inversion layer, the mobile charge density is $\rho_I(x)$. The areal density of inversion charge, $Q_I = \int_0^\infty \rho_I(x)\,dx$, is determined directly from the Shubnikov–de Haas periodicity.

At the interface $(x = 0)$, the conduction band discontinuity Δ is known.[8] There can also be true localized states at the interface, such as broken bonds, whose energy is tied to the band structure and not to the Landau levels. These states are not describable

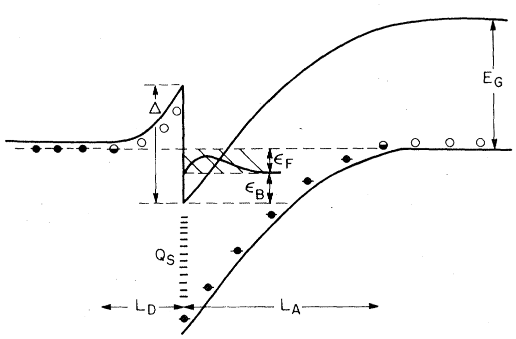

FIG. 1. Diagram of the heterojunction at equilibrium in zero magnetic field.

within the effective-mass formalism used for the inversion-layer carriers, and it is not correct to regard them as having been removed from the Landau levels. Nonetheless, they can provide a reservoir of electrons to supply carriers to the inversion layer and thus produce plateaus. Not knowing their energy density, we ignore their reservoir consequences here by assuming that they all lie beneath the inversion layer, and are always filled, giving rise to a charge density $Q_S\delta(x)$. Q_S must be adjusted for compatibility with the other measured parameters, since their values were, in fact, physically determined by Q_S.

On the $Al_xGa_{1-x}As$ side $(x < 0)$, the donor concentration N_D is high. It is nominally uniform up to a distance L from the interface, and zero closer than that. L is typically 0 to 100 Å in modulation-doped samples.[9] When the junction is in equilibrium, E_F is near the bulk $Al_xGa_{1-x}As$ conduction-band edge. Band bending empties some donors, establishing a charge density $-N_D$ for $-L_D < x < -L$. When we ignore the small difference between the dielectric constant ϵ in the two semiconductors, the following conditions (the depletion layer approximation) relate charge densities and energies

$$(4\pi e^2/\epsilon)[\tfrac{1}{2}N_D(L_d^2 - L^2) + Q_I X_I + \tfrac{1}{2}N_A L_A^2] = \Delta + E_G \ , \tag{1}$$

$$N_D(L_D - L) = Q_I + N_A L_A + Q_S \ , \tag{2}$$

where

$$X_I = \int_0^\infty x\rho_I(x)\,dx \Big/ Q_I \ .$$

For one occupied subband, $Q_I = \int_0^{\epsilon_F} \rho(\epsilon)\,d\epsilon$. ϵ_F is the filling energy and $\rho(\epsilon)$ is the transverse density of states. Taking spin degeneracy, but not spin energy, into account

$$\rho(\epsilon) = \rho\hbar\omega_c \sum_n \delta(\epsilon - (n - \tfrac{1}{2})\hbar\omega_c) \ ,$$

$$\omega_c = eB/m^* \ ,$$

and

$$\rho = (2\pi)^{-1}(2m^*/\hbar^2) \ .$$

If there is no magnetic field, then $\rho(\epsilon) = \rho$, a constant. Thus

$$Q_I = \begin{cases} \rho\epsilon_F, \quad \text{no field} & \text{(3a)} \\ N\rho\hbar\omega_c, \quad (N - \tfrac{1}{2})\hbar\omega_c < \epsilon_F < (N + \tfrac{1}{2})\hbar\omega_c & \text{(3b)} \\ (N + \alpha)\rho\hbar\omega_c, \quad (N + \tfrac{1}{2})\hbar\omega_c = \epsilon_F \ . & \text{(3c)} \end{cases}$$

Equation (3c) applies when level N is *at* E_F. In such case, the fractional filling α lies between 1 and zero.

Since ϵ_F is determined by E_F in bulk,

$$(4\pi e^2/\epsilon)N_D(L_D^2-L^2)/2=\Delta-\epsilon_B-\epsilon_F \quad , \tag{4}$$

ϵ_B and X_I depend in detail on the potential $\phi(x)$ (see below), but they are typically 100 mV and 100 Å. Treating them for the the moment as known, we solve Eqs. (1), (2), (3a), and (4) to determine Q_S, L_A, ϵ_F, and L_D. If necessary, we recompute $\phi(x)$, recompute ϵ_B and X_I, and iterate. Choosing $Q_I=4.2\times10^{11}/\text{cm}^2$, $L=100$ Å, $N_D=7\times10^{17}/\text{cm}^3$, $N_A=10^{15}/\text{cm}^3$, and $\Delta=315$ mV gives $Q_S=3.8\times10^{11}/\text{cm}^2$, and $L_A=1.5\times10^4$ Å.

The barrier to tunneling is roughly parabolic, with $\Delta-\epsilon_B\approx245$ mV and $L_D=236$ Å. Using $m=0.068m_0$, we estimate the equilibration time between the two systems to be about 10^{-5} sec.

Now we discuss the effect of B. The sample is cooled and the semiconductors become insulating. As a result, acceptors do not adjust, $Q_A=N_AL_A$ is frozen at the value just found, the Fermi energy in the bulk, GaAs is irrelevant, and we ignore Eq. (1). The donors and the inversion layer are still in communication, however, and the L_D, Q_I, and ϵ_F are still related. We first consider magnetic fields where there is a partially filled Landau level. Then Eqs. (2), (3c), and (4) must be solved for Q_I, L_D, ϵ_F, and $(N+\alpha)$. Combining these four equations so as to eliminate Q_I, ϵ_F, and L_D, and putting $y\equiv\epsilon_F^0/\hbar\omega_c$ and $Q_A^*=Q_A^0+Q_S^0$, we obtain

$$A(N+\alpha)^2+(B+C)(N+\alpha)y+C(\tfrac{1}{2}-\alpha)y-Dy^2=0 \quad , \tag{5a}$$

$$A=(Q_I^0)^2 \quad , \tag{5b}$$

$$B=2Q_I^0(Q_A^*+N_DL) \quad , \tag{5c}$$

$$C=(2\pi e^2/\epsilon)^{-1}N_D\epsilon_F^0 \quad , \tag{5d}$$

$$D=(2\pi e^2/\epsilon)^{-1}N_D(\Delta-\epsilon_B)-Q_A^*(Q_A^*+2N_DL) \quad . \tag{5e}$$

The superscript zero refers to $B=0$ values found above.

Ignoring for the moment the Q_I dependence of ϵ_B, we see that Eq. (5) exhibits the essential features of the solution. We regard $(N+\alpha)$ as a parameter, and consider the resulting values of

$$Q_I/\rho\hbar\omega_c=(N+\alpha) \quad , \tag{6a}$$

$$\epsilon_F^0/\hbar\omega_c=y \quad . \tag{6b}$$

For $(N+\alpha)$ noninteger, α is uniquely determined, and for this value of α, there is one positive solution y. For $(N+\alpha)=N_0$ an integer, two values of α, $\alpha=0$ and 1, must be used, and for each, there is a positive solution y. Values of y between the two do not satisfy Eq. (5) because they do not correspond to fields where there is a partially filled level at the surface. For them, there are exactly N_0 filled Landau levels, and Eq. (3b) applies. These are the plateaus where ρ_{xy} would be quantized.

Now consider the variation of ϵ_B. Ultimately, ϵ_B depends on Q_A and Q_I, but in the present situation, Q_A is fixed. If the fractional variation in Q_I is not great, we can expand to first order in Q_I to obtain

$$\epsilon_B=[1+(m-1)G]\epsilon_B^0 \quad , \tag{7a}$$

$$G=\left[\frac{Q_I}{\epsilon_B}\frac{d\epsilon_B}{dQ_I}\right]^0 \quad , \tag{7b}$$

$$m=Q_I/Q_I^0=(N+\alpha)/y \quad . \tag{7c}$$

Although the evaluation of ϵ_B^0 and of G requires, in principle, the self-consistent solution of the Schrödinger equation, a simple analytic approximation, used first by Fang and Howard[10] and studied systematically by Stern, has been shown to be quite good.[11] One notes that for the wave function $\psi_b(z)=(b^3/2)^{1/2}ze^{-bz/2}$, the charge density $N_A+\psi_b(z)^2Q_I$, and the resulting $\phi(z)$, depend on b. The expectation value of the Hamiltonian taken with respect to the trial wave function $\psi_\beta(z)=(\beta^3/2)^{1/2}ze^{-\beta z/2}$ depends on b and β, and is to be minimized by varying β. Then the whole procedure is made self-consistent by setting $\beta=b$. The resulting equation determines β, and one obtains ϵ_B, X_I, and G as analytic functions of Q_I and Q_A.

Inserting Eq. (7) into Eq. (5e) allows us to separate out the equilibrium part, and we have

$$D=D^0-[(N+\alpha)/y-1]F \quad (F\equiv CG\epsilon_B^0/\epsilon_F^0) \quad .$$

By making use of the $B=0$ solution, we show that $D^0=A+B+C$. Using this, we find that for $\alpha=\frac{1}{2}$, the solution is $y=N+\alpha$, and $Q_I=Q_I^0$. For general values of α, we write $y=N+\alpha+K$, and to first order in $K/(N+\alpha)$, we find

$$K=(\tfrac{1}{2}-\alpha)\frac{C}{[2A+B+C+F-(\frac{1}{2}-\alpha)C/(N+\alpha)]}$$

This is single valued when $(N+\alpha)$ is not integer and has two values, because there are two values of α when $(N+\alpha)=N_0$. The plateau width, expressed as a fraction of the Shubnikov–de Haas period is $C/(2A+B+C+F)$. It arises solely because of the donors acting as a reservior.

The results obtained to this point are qualitatively identical to what is shown in Fig. 2. The slight difference arises because we have not yet considered the self-energy Σ of an electron in the topmost Landau level. Σ depends on N, on α, on Q_I, and on $\hbar\omega_c$, but since these last two are determined by N and α, we write $\Sigma=\Sigma_N(\alpha)$. To include its effect, we modify Eq. (4), replacing ϵ_B by $\epsilon_B+\Sigma$. The solutions are modified in form somewhat: D^0, as modified, is still equal to $A+B+C$, and $\epsilon_B-\epsilon_B^0$ is aug-

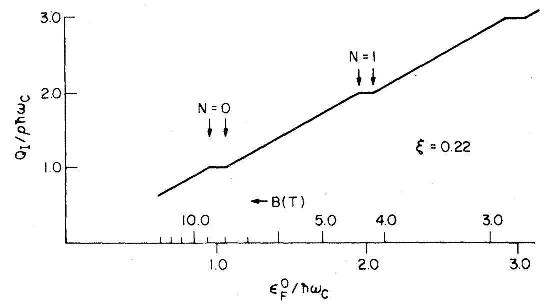

FIG. 2. $Q_I/\rho\hbar\omega_c$ vs $\epsilon_F^0/\hbar\omega_c$, calculated for sample 1, Ref. 2, using parameters given in the text.

mented by $\delta\Sigma(\alpha) = \Sigma_N(\alpha) - \Sigma^0$. Σ^0 is the self-energy of an electron at the Fermi energy of a 2DEG at $B = 0$.

Although $\delta\Sigma$ is not known *a priori*, we can make reasonable guesses about its form. Because we found that $\alpha = \frac{1}{2}$ gave the $B = 0$ value of density, we conjecture that $\delta\Sigma$ will vanish at $\alpha = \frac{1}{2}$. The difference $\delta\Sigma(0) - \delta\Sigma(1)$ is the energy associated with g-factor enhancement, and the appropriate scale of energy is $\hbar\omega_c$. A simple form which satisfies these conjectures, varies strongly with α (the electrons in a partially filled level should be quite polarizable), and fits easily into our method of computation, is $\delta\Sigma(\alpha) = (\frac{1}{2} - \alpha)\hbar\omega_c\xi$. By analogy with the g-factor situation, we expect the parameter ξ to be positive, and in a better theory, to depend on Landau level width.

The analysis is easily repeated with $\delta\Sigma$. The final result is that in calculating K, $(\frac{1}{2} - \alpha)$ is multiplied by the factor $(1 + \xi)$. In Fig. 2, we exhibit $(N + \alpha)$ vs y, calculated using $\xi = 0.22$ and using parameters describing sample 1 in Ref. 2. Field values shown correspond to having $\hbar\omega_c = \epsilon_F^0$ at $B = 8.4$ T. The arrows are at the edges of the σ_{xy} plateaus reported in Table II, Ref. 2. In the absence of ξ (our only fitting parameter), the plateaus are 22% narrower. Our calculations demonstrate that self-consistency alone can account for much of, and inclusion of self-energy the entirety of, the experimental results from GaAs-$Al_xGa_{1-x}As$ heterojunctions.

In Si MOSFET's, there is no known donor in the SiO_2 to act as an electron reservoir. The acceptors in Si are separated by too wide a depletion layer to be effective. Similarly, the gate electrode, which acts as a reservoir via the wires which connect it to the 2DEG, is so far away that the plateau it produces would have virtually no width. However, the interface states included in our calculation can be present in the energy range of interest and act as a reservoir. The localized states lying outside the mobility edges of each Landau level may also act as a reservoir. Clearly, all these effects must be properly taken into account before quantitative conclusions about mobility edges and localization can be drawn.

It is a pleasure to thank S. J. Allen, and H. L. Stormer, A. C. Gossard, and P. M. Platzman for useful comments.

[1]K. V-Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[2]D. C. Tsui and A. C. Gossard, Appl. Phys. Lett. 38, 550 (1981).
[3]D. C. Tsui and S. J. Allen Phys. Rev. B (in press).
[4]R. E. Prange, Phys. Rev. B 23, 4802 (1981).
[5]R. B. Laughlin, Phys. Rev. B 23, 5632 (1981).
[6]H. Aoki and T. Ando.
[7]T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 37, 1044 (1974).
[8]R. Dingle, in *Festkorperprobleme, Advances in Solid State Physics* (Pergamon-Vieweg, Braunschweig, 1975), Vol. XV, p. 21.
[9]R. Dingle, H. L. Stormer, A. C. Gossard, and W. Weigmann, Appl. Phys. Lett. 37, 665 (1978).
[10]F. F. Fang and W. E. Howard, Phys. Rev. Lett. 16, 797 (1966).
[11]F. Stern, Phys. Rev. B 5, 4891 (1972).