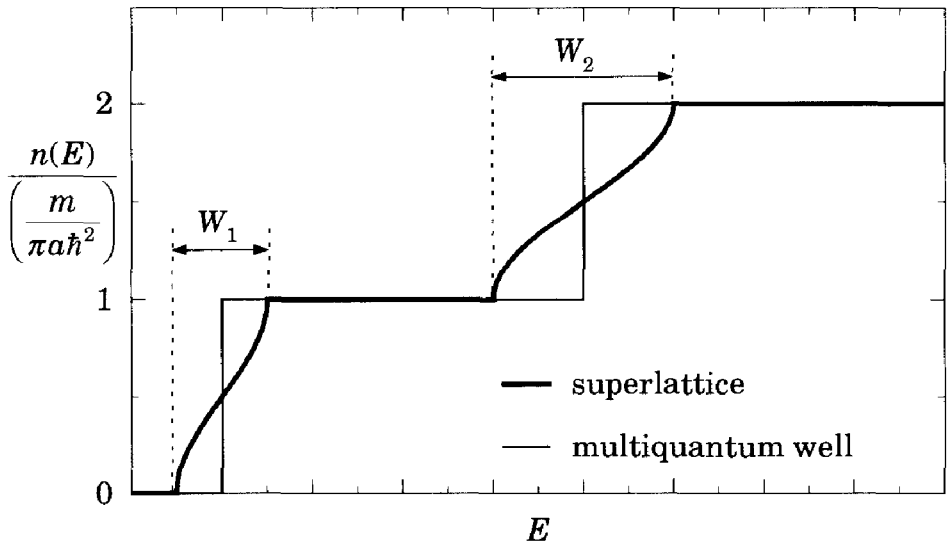

FIGURE 5.20. Density of states for a superlattice (thick line) and a multiquantum well (thin line) scaled to the same value in the plateau, $m/\pi a\hbar^2$. The two lowest bands have widths W_1 and W_2.

multiquantum well into a superlattice, broadens the sharp step in the density of states into an arcsine of width W. The bottom of each subband has a parabolic density of states, the general result for a three-dimensional system, but the superlattice is anisotropic. The effective mass for motion perpendicular to the direction of growth, m in the preceding formulas, is that for freely propagating electrons in the host, while that for motion along the superlattice depends on W.

5.7 Coherent Transport with Many Channels

The one-dimensional systems studied earlier in this chapter were purely one-dimensional in the sense that only one subband was occupied. Such systems are analogous to electromagnetic wave guides in which only one mode can propagate. Further subbands become occupied if the Fermi level of a quasi-one-dimensional system is raised, just as further modes can propagate in a wave guide as the frequency of operation is raised. In this section we shall first extend the theory of conduction to describe a scattering centre or 'sample' between two such quasi-one-dimensional systems or 'leads'. An example is a narrow constriction between two wider leads, which shows a quantized conductance. This is more complicated to describe than a purely one-dimensional system because reflection and transmission can now occur between different subbands or modes. The next step is to study samples that have more than two leads attached to them. The sample can now be a much more general object, and the most significant application will be to the quantum Hall effect in a sample with many leads. The theory remains restricted to coherent transport, so in

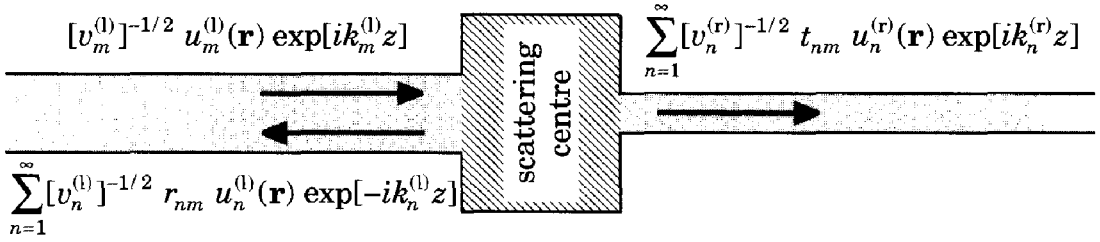

FIGURE 5.21. Coherent transport through a system with two leads, each with many propagating states.

practice the sample must be small enough that inelastic scattering is unlikely, and we shall treat only linear response and low temperature.

5.7.1 TWO LEADS WITH MANY CHANNELS: THE QUANTIZED CONDUCTANCE

The first extension beyond a strictly one-dimensional system is shown in Figure 5.21. There are two leads with a scattering centre between them, as in earlier sections, but each lead now has several subbands arising from the transverse states (Section 4.8). These subbands are also called *modes* or *channels*. It is essential that the transverse potential of each lead remains constant along its length, although its precise form is unimportant. This is part of the definition of a perfect lead, and any regions where the potential changes must be included as part of the scattering centre. The wave function within a lead takes the form

$$\psi(\mathbf{R}) = \sum_n v_n^{-1/2}[A_n \exp(ik_n z) + B_n \exp(-ik_n z)]u_n(\mathbf{r}). \tag{5.94}$$

This generalizes equation (5.1). Transverse states are labelled by n (which should really be two labels), with wave functions $u_n(\mathbf{r})$ and energies ε_n. The total energy $E = \varepsilon_n + \hbar^2 k_n^2/2m$, so states propagate if $E > \varepsilon_n$ and decay otherwise. The factor with the velocity v_n of each mode normalizes the states by flux rather than density as in equation (5.48). The two leads are not in general identical, so the energies of the one-dimensional subbands are different and so are the number of propagating states in each, N_{left} and N_{right}.

Inject a wave from the left purely in mode m. The scattering centre mixes the different modes so the scattered wave has contributions from all outgoing modes on both sides. The wave functions in the left and right leads take the form

$$\begin{aligned}\psi_{\text{left}}(\mathbf{R}) &= [v_m^{(l)}]^{-1/2}u_m^{(l)}(\mathbf{r})\exp(ik_m^{(l)}z) \\ &\quad + \sum_{n=1}^{\infty}[v_n^{(l)}]^{-1/2}r_{nm}u_n^{(l)}(\mathbf{r})\exp(-ik_n^{(l)}z), \\ \psi_{\text{right}}(\mathbf{R}) &= \sum_{n=1}^{\infty}[v_n^{(r)}]^{-1/2}t_{nm}u_n^{(r)}(\mathbf{r})\exp(ik_n^{(r)}z).\end{aligned} \tag{5.95}$$

The sums run over *all* values of n, not just the propagating states; the decaying states are essential to complete the wave function. There are now arrays of reflection and transmission coefficients rather than the simple numbers in the strictly one-dimensional case.

The S- and T-matrices defined in Section 5.2 can be extended to many modes. They are essential for detailed calculations but contain much information that may not be needed afterwards. A simpler matrix t can be constructed from the coefficients t_{nm} in equation (5.95), giving the transmission amplitude for an electron incident from the left in mode m to be transmitted on the right in mode n. We restrict t to propagating states, giving dimensions of $N_{\text{right}} \times N_{\text{left}}$.

One reason for using t is that it contains sufficient information to find the conductance. The derivation follows equation (5.58), which gave $G = (2e^2/h)T = (2e^2/h)|t|^2$ for one dimension. Consider electrons injected in a given mode m. Those that emerge in mode n make a contribution $(2e^2/h)|t_{nm}|^2$ to the conductance. The velocity of different modes is taken into account by the normalization and does not clutter this result. The total conductance is found by summing over all input and output modes:

$$G = \frac{2e^2}{h} \sum_m \sum_n |t_{nm}|^2. \tag{5.96}$$

A vital feature is that the sum is over *intensities* rather than *amplitudes*: it is assumed that there is no phase coherence between electrons injected in different modes, so that we can simply add the contributions to the current and not worry about interference between them. It is also implicit that each mode has the same Fermi level, yet another demand on the leads. The conductance can be written in a more compact form by using the Hermitian-conjugate matrix of t, defined by $(\mathrm{t}^\dagger)_{mn} = (t_{nm})^*$. Then

$$\begin{aligned} G &= \frac{2e^2}{h} \sum_{m,n} t_{nm} t_{nm}^* = \frac{2e^2}{h} \sum_{m,n} (\mathrm{t})_{nm} (\mathrm{t}^\dagger)_{mn} \\ &= \frac{2e^2}{h} \sum_n (\mathrm{tt}^\dagger)_{nn} = \frac{2e^2}{h} \mathrm{Tr}\,(\mathrm{tt}^\dagger) = \frac{2e^2}{h} \mathrm{Tr}\,(\mathrm{t}^\dagger \mathrm{t}). \end{aligned} \tag{5.97}$$

This is the form in which the result is usually quoted, where 'Tr' is the trace of the matrix (the sum of its diagonal elements). Note that the product $\mathrm{tt}^\dagger$ is square, although neither t nor $\mathrm{t}^\dagger$ need be, and that the two expressions for the trace are equal, although $\mathrm{tt}^\dagger$ and $\mathrm{t}^\dagger \mathrm{t}$ may not be the same size.

This result can be used to calculate the conductance of a short wire or constriction, the *quantum point contact*. A typical structure is illustrated in Figure 5.22(a). Two gates shaped like opposed fingers on the surface of a heterostructure are negatively biassed to deplete the 2DEG underneath them. The remaining electrons are forced to travel through the gap between the gates, which behaves like a short quasi-one-dimensional system. The broad regions of 2DEG on either side act as the 'leads'.

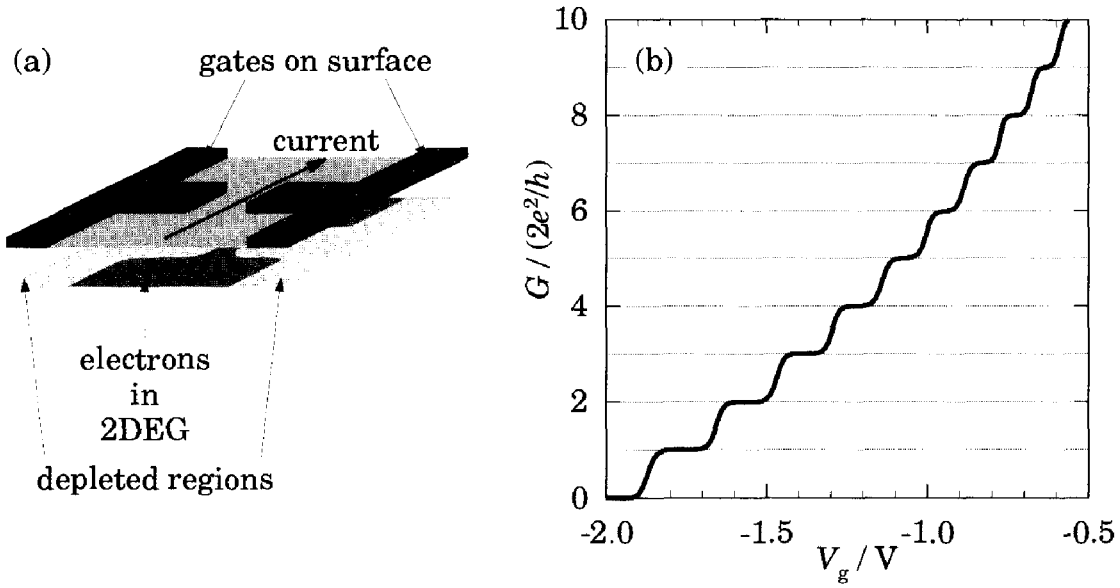

FIGURE 5.22. (a) Layout of a typical quantum point contact, a short constriction defined by patterned metal gates on the surface of a heterostructure containing a 2DEG. (b) Calculated conductance $G(V_g)$ as a function of gate voltage V_g. [From Nixon, Davies, and Baranger (1991).]

If the gates are short, the potential underneath them looks like the saddle shown in Figure 5.23(a). Because the potential varies so smoothly, it is possible to use the *adiabatic approximation*. The idea is to write the wave function of each mode in the separable form

$$\phi_n(\mathbf{R}) \approx u_n(\mathbf{r};\, z)[v_n(z)]^{-1/2}\{A_n(z)\exp[ik_n(z)z] + B_n(z)\exp[-ik_n(z)z]\}. \tag{5.98}$$

The wave function and energy in the transverse potential at z, $u_n(\mathbf{r};\, z)$ and $\varepsilon_n(z)$, are calculated as though this potential were constant along the wire. The wave number follows from $E = \varepsilon_n(z) + \hbar^2 k_n^2(z)/2m$. A mode may be propagating in one region and decaying in another. This approximation is related to the WKB method (Section 7.4) and is applicable only if the transverse potential changes slowly along z. It may also be possible to neglect scattering from one mode to another, in which case the amplitudes $A_n(z)$ and $B_n(z)$ may be calculated independently for each mode. The matrix t then becomes diagonal.

The energy $\varepsilon_n(z)$ of each subband varies with longitudinal position z through the constriction, rising to a broad peak in the middle (Figure 5.23(b)). Many states propagate while they are far from the constriction, but their wave number becomes imaginary as they approach the saddle point and see an apparent barrier when $\varepsilon_n(z) > E$. Such an electron (modes 2 and 3 in Figure 5.23(b)) may tunnel through the barrier but most of the amplitude is reflected unless the apparent barrier is low. Only electrons in the lowest N_{trans} modes propagate throughout the constriction (mode 1 in Figure 5.23(b)). Thus t has diagonal entries of nearly unity for the lowest N_{trans} modes, which are transmitted, and small values for the others. Equation (5.97)

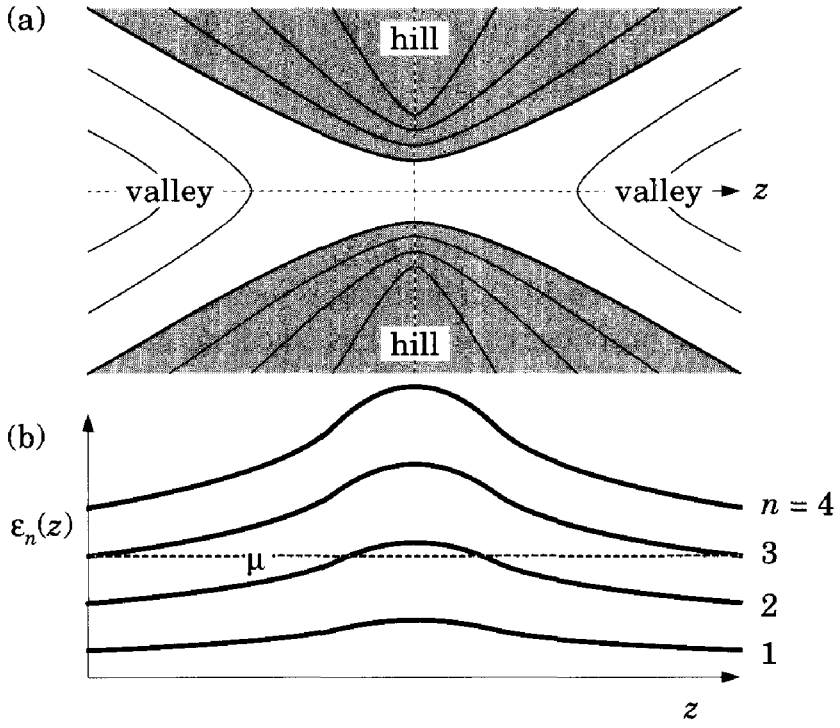

FIGURE 5.23. (a) Contours of a smooth saddle-point potential used to demonstrate the quantized conductance. The thicker line shows where the Fermi level meets the potential energy and the shaded region has high potential energy. (b) Energies $\varepsilon_n(z)$ of the transverse modes as a function of their position z.

shows that the conductance is given by

$$G = \frac{2e^2}{h}\mathrm{Tr}\,(t^\dagger t) \approx \frac{2e^2}{h}N_{\mathrm{trans}}\,. \tag{5.99}$$

This is the quantized conductance. The value of N_{trans} can be changed by altering the width and depth of the constriction, usually by varying the bias V_g on the gates. Thus a plot of $G(V_g)$ should give a steplike curve, with G jumping by $2e^2/h$ whenever another mode is allowed to propagate through the saddle point. An example is shown in Figure 5.22; this is a simulation but some experiments are even better! The rounding of the steps is due to tunnelling through the saddle point. Raising the temperature has a similar effect.

Adiabatic propagation of electrons is not a necessary condition for the quantized conductance. Scattering between modes will have no effect provided that only forward scattering occurs so the direction of the electron is preserved (although the magnitude of its velocity must change). Any backward scattering, on the other hand, will be disastrous and the constriction must be kept short to avoid this, typically below 1 μm. Sharp features in the potential also produce structure in $G(V_g)$. This sensitivity to the details of the potential means that the quantization of the conductance is not particularly accurate, and a result within 10% of $(2e^2/h)N_{\mathrm{trans}}$ is good. This contrasts strongly with the quantum Hall effect, where the Hall conductance takes the same value but in a good sample is in perfect agreement with the value of

$2e^2/h$ deduced from other high-precision experiments. It will be explained with a similar formalism in Section 6.6.

5.7.2 SYSTEMS WITH MANY LEADS

The next step is to allow for an arbitrary number of leads N_{leads}. The general form of the geometry, with some specific applications, is shown in Figure 5.24. Typically some leads are used to inject currents, whereas others measure voltages, and the leads are often called current or voltage probes. It is assumed that *voltage probes* are connected to ideal voltmeters, which draw no current. Case (c) is important as it is commonly used in practice to measure resistance. A current is passed along the straight path and the resulting voltage is measured between the two side arms. This *four-probe* configuration is preferred because it is insensitive to the resistance of the contacts. Alternatively the voltage can be measured between the two contacts used to pass the current, giving a *two-probe* measurement. We shall see that the results can be startlingly different.

Although S-matrices are needed for detailed analysis, we shall continue with the t-matrix defined in the previous section. Use m and n to label the leads and α and β to label the propagating modes within each lead; N_m modes propagate in lead m. As usual the leads must have constant cross-section, and states should be normalized to constant flux. An inward-flowing current in a lead is defined to be positive.

Consider a particular lead and mode, (m, α), say. We are interested only in deviations from equilibrium so the excess current that impinges on the sample from this lead is given, as in Section 5.4.1, by $I_{m\alpha}^{\text{inc}} = (-2e/h)\delta\mu_m$. The change in Fermi energy $\delta\mu_m = -eV_m$, where V_m is the applied voltage, so $I_{m\alpha}^{\text{inc}} = (2e^2/h)V_m$. This is the same for all N_m modes in this lead.

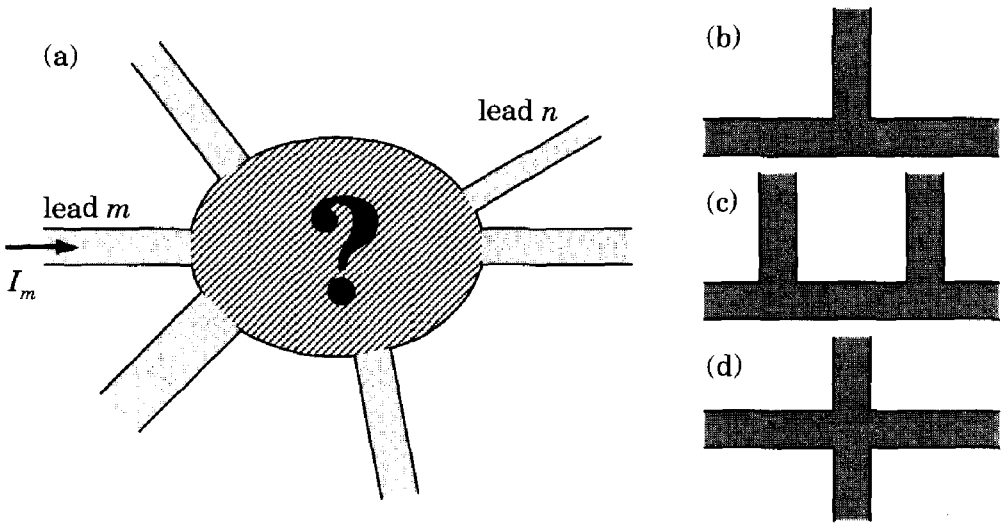

FIGURE 5.24. Geometry of a sample for coherent transport with many leads. The general case is shown in (a) with specific examples of (b) a T-junction, (c) four-terminal measurement of longitudinal resistance, and (d) a microscopic Hall bar.

Of the electrons incident from this mode and lead, those entering into mode β of lead n have amplitude $t_{n\beta,m\alpha}$. This contributes a current $-I^{\text{inc}}_{m\alpha}|t_{n\beta,m\alpha}|^2$, which is negative because it flows outwards. Some electrons are reflected back down lead m, not necessarily in their original mode, with reflection amplitudes $r_{m\beta,m\alpha}$. Since $t_{n\beta,m\alpha}$ is defined only for $n \neq m$ and $r_{n\beta,m\alpha}$ exists only for $n = m$, one often defines $t_{m\beta,m\alpha} = r_{m\beta,m\alpha}$.

The total current in lead n due to electrons injected from a different lead m is given by summing over all modes in the two leads, as in equation (5.96). Thus

$$I_{nm} = -\frac{2e^2}{h} V_m \sum_{\beta=1}^{N_n} \sum_{\alpha=1}^{N_m} |t_{n\beta,m\alpha}|^2. \tag{5.100}$$

All expressions for currents involve sums such as this over the modes in both leads. It is like the trace in equation (5.97), although there are now extra subscripts on the transmission coefficients to label the leads. We can clarify the notation by introducing still more transmission and reflection coefficients which absorb these traces:

$$T_{nm} = \sum_{\beta=1}^{N_n} \sum_{\alpha=1}^{N_m} |t_{n\beta,m\alpha}|^2, \qquad R_m = \sum_{\beta=1}^{N_m} \sum_{\alpha=1}^{N_m} |r_{m\beta,m\alpha}|^2. \tag{5.101}$$

These coefficients can be greater than unity, unlike the simple coefficients in one-dimensional systems. For example, T_{nm} reaches the smaller of N_m and N_n for perfect transmission. Again, one often sets $T_{mm} = R_m$ to treat the transmission and reflection coefficients on an equal footing.

The net current injected into lead m is given by the incident current less the reflected current. The total incident current, summed over all propagating modes, is $(2e^2/h)N_m V_m$, so the net current is

$$I_{mm} = (2e^2/h)(N_m - R_m)V_m. \tag{5.102}$$

Conservation of current requires that this be equal to the total current injected from lead m that leaves the sample through other leads, that is, $I_{mm} = \sum_{n,n\neq m} I_{nm}$. This leads to a sum rule on the transmission coefficients,

$$R_m + \sum_{n,n\neq m} T_{nm} = N_m. \tag{5.103}$$

This is an obvious generalization of $R + T = 1$.

We have now calculated the currents due to electrons injected from lead m. All that remains is to sum the contributions from all leads. Lead n causes a current $-(2e^2/h)T_{mn}V_n$ in lead m, so the total current in lead m is

$$I_m = \frac{2e^2}{h}\left[(N_m - R_m)V_m - \sum_{n,n\neq m} T_{mn}V_n\right]. \tag{5.104}$$

This the Landauer–Büttiker formula for the conductance of a system with many leads. It can also be written in terms of a square conductance matrix whose dimension is given by the number of leads:

$$I_m = \sum_n G_{mn} V_n , \qquad G_{mn} = \frac{2e^2}{h}[(N_m - R_m)\delta_{mn} - T_{mn}]. \tag{5.105}$$

This matrix must be treated with some care. The condition (5.103) for current conservation means that each column n of G_{mn} sums to zero. This shows immediately that the determinant vanishes and the matrix is singular. Another condition comes from the requirement that no current should flow if all voltages are equal, which means that

$$\sum_n G_{mn} = 0 = (N_m - R_m) - \sum_{n,n\neq m} T_{mn}. \tag{5.106}$$

Thus each row m of the conductance matrix must also sum to zero. The fact that both the rows and columns sum to zero leads to the relation

$$\sum_n T_{mn} = \sum_n T_{nm}. \tag{5.107}$$

These conditions can be used to rewrite the current, equation (5.104), in a number of ways. Replacing $(N_m - R_m)$ using the condition (5.103) for current conservation leads to

$$I_m = \frac{2e^2}{h} \sum_{n,n\neq m} (T_{nm} V_m - T_{mn} V_n). \tag{5.108}$$

The diagonal term T_{nn} is not needed. Replacing the same term using the 'row' sum rule (5.106) gives

$$I_m = \frac{2e^2}{h} \sum_{n,n\neq m} (T_{mn} V_m - T_{mn} V_n) = \frac{2e^2}{h} \sum_{n,n\neq m} T_{mn}(V_m - V_n). \tag{5.109}$$

This shows explicitly that only differences between applied voltages are significant.

Note that equation (5.107) does not imply that the conductance matrix must be symmetric, except in the case of only two leads. However, time-reversal symmetry makes the matrix symmetric in the absence of a magnetic field. This symmetry is broken by a magnetic field, which is important in the analysis of the Hall effect.

It is straightforward to verify that equation (5.104) agrees with our earlier results for systems with two probes. The next most complicated case is a sample with three leads, as in Figure 5.25(a). Let lead 3 be a voltage probe connected to an ideal voltmeter, which draws no current, so $I_3 = 0$. A current I flows into lead 2 and out of lead 1, so $I_1 = -I$ and $I_2 = I$. Finally, set $V_1 = 0$ as the reference

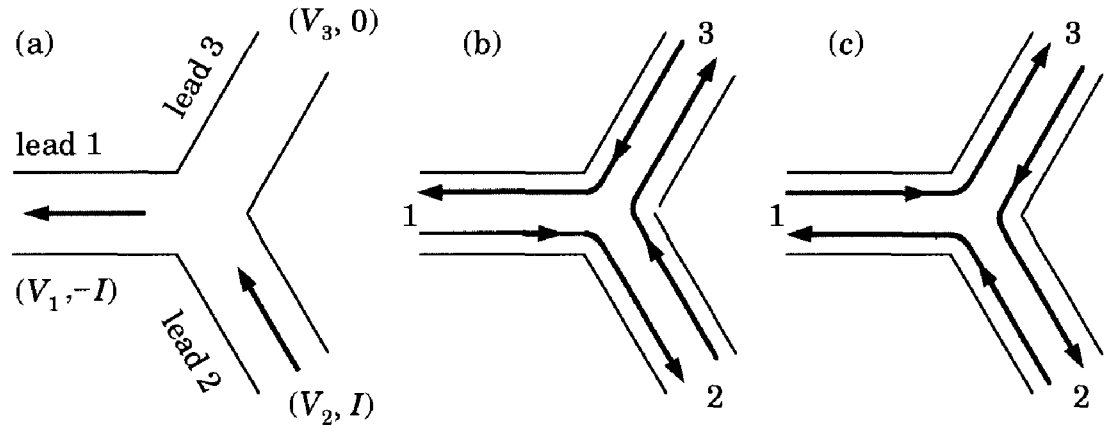

FIGURE 5.25. A sample with three leads to illustrate the multiprobe formula. A current I flows into lead 3 and out of lead 1; no current passes down lead 2, which is used purely to measure voltage. The two figures on the right show the transmission coefficients in large positive and negative magnetic fields when the device acts as a 'circulator'.

voltage. Equation (5.108) becomes

$$I_1 = -I = \frac{2e^2}{h}[-T_{12}V_2 - T_{13}V_3]$$

$$I_2 = I = \frac{2e^2}{h}[(T_{12} + T_{32})V_2 - T_{23}V_3]$$

$$I_3 = 0 = \frac{2e^2}{h}[-T_{32}V_2 + (T_{13} + T_{23})V_3]. \tag{5.110}$$

Adding the three equations gives zero on both sides, confirming that one is redundant. Solution of these equations shows that the two-probe conductance of the system between leads 1 and 2 is

$$\frac{I}{V_2} = \frac{2e^2}{h}\left(T_{12} + \frac{T_{13}T_{32}}{T_{13} + T_{23}}\right). \tag{5.111}$$

There are two contributions to the conductance. The first is due to those electrons that go directly from lead 1 to lead 2, as expected. The second is indirect and arises from electrons that go from lead 1 into 3. This is a voltage probe and carries no net current, so the flow must be balanced by an equal and opposite current that divides between leads 1 and 2 in the ratio of their transmission coefficients. Another useful result is the potential measured in lead 3:

$$\frac{V_3}{V_2} = \frac{T_{32}}{T_{13} + T_{23}} = \frac{T_{32}}{T_{31} + T_{32}}. \tag{5.112}$$

This closely resembles a potential divider. The second form of the denominator follows from the rows-and-columns sum rule (5.107).

To check that these make sense, suppose that the leads each support only one transverse mode and that the structure has threefold symmetry. In the absence of

a magnetic field the transmission coefficients between different leads are identical. The highest value of T_{12} permitted by the usual conservation and symmetry laws is $\frac{4}{9}$ and in this case we find

$$\frac{I}{V_2} = \frac{2e^2}{h}\left(\frac{4}{9} + \frac{2}{9}\right) = \frac{2}{3}\frac{2e^2}{h}, \qquad \frac{V_3}{V_2} = \frac{1}{2}. \tag{5.113}$$

The presence of the strongly coupled voltage probe, which reflects some of the electrons despite drawing no net current, has reduced the conductance below its value of $2e^2/h$ for a perfect system with only two leads. The ratio $V_3/V_2 = \frac{1}{2}$ is just what we would expect for a classical potential divider and follows from symmetry.

Now apply a large magnetic field. It is possible, as we shall see in Section 6.5, to arrange that the electrons are all forced to go down the lead to their right. In microwaves this would be a 'circulator'. Then $T_{12} = T_{23} = T_{31} = 1$, the others vanish, and we get

$$\frac{I}{V_2} = \frac{2e^2}{h}(1+0) = \frac{2e^2}{h}, \qquad \frac{V_2}{V_1} = 0. \tag{5.114}$$

Reversing the field gives $T_{21} = T_{32} = T_{13} = 1$ and

$$\frac{I}{V_2} = \frac{2e^2}{h}(0+1) = \frac{2e^2}{h}, \qquad \frac{V_2}{V_1} = 1. \tag{5.115}$$

The behaviour of the voltage probe is quite different for the two directions of magnetic field, although the conductance measured between two probes is not affected by the direction of the magnetic field (a general result for a two-probe measurement, mentioned earlier).

An important feature is that the indirect and direct currents are not coherent with each other, because the indirect current involves electrons that emerge from a different lead (the voltage probe 3). This can provide a useful theoretical trick to simulate lack of full coherence in tunnelling, which is very difficult to treat more formally. One just couples an additional voltage probe to the sample where the incoherence is supposed to originate. This picture can be verified by suppressing direct transmission from 1 to 2 so that all current is forced to take the indirect path. Then equation (5.111) becomes

$$\frac{I}{V_2} = \frac{2e^2}{h}\frac{T_{13}T_{32}}{T_{13}+T_{23}}. \tag{5.116}$$

The symmetry $T_{23} = T_{32}$ reduces this to the classical formula for two resistors in series, contact 3 acting as the joint between them.

Finally, assume that the coupling to lead 3 is very weak. This might be the case in practice because we would like the voltage probe to disturb the system as little as possible. Then the direct current dominates equation (5.111), which depends only

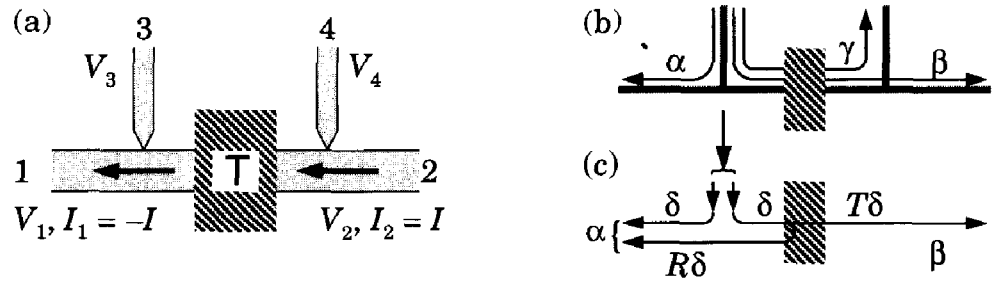

FIGURE 5.26. Two- and four-terminal measurements of the resistance of a tunnelling barrier. (a) Current is passed between probes 1 and 2; voltage can be measured either between these or between the weakly coupled probes 3 and 4. (b) Definition of the transmission coefficients coupling the voltage probes to the sample. (c) Derivation of the relation between the transmission coefficients α and β.

on T_{12} as we would hope. The voltage on probe 3, equation (5.112), unfortunately depends strongly on the ratio $T_{31} : T_{32}$ of the couplings to the two other leads. We would expect $V_3 = \frac{1}{2}V_2$ for a perfect structure (in the absence of a magnetic field). This requires the couplings to the two current leads to be equal, not very surprisingly. Any imperfections that break this symmetry will affect V_3.

Our final task is to derive a general formula for a four-probe resistance. The configuration is shown in Figure 5.26(a). Let current I enter through probe 2 and leave through probe 1; we wish to find the voltage between probes 4 and 3, which draw no current. There is a general notation $R_{mn,pq} = V_{pq}/I_{mn}$ for such quantities, where V_{pq} is the potential difference that appears between contacts p and q in response to a current between m and n. Thus we want the four-terminal resistance $R_{21,43}$. The full set of equations (5.109) is

$$\begin{pmatrix} T_{12}+T_{13}+T_{14} & -T_{12} & -T_{13} & -T_{14} \\ -T_{21} & T_{21}+T_{23}+T_{24} & -T_{23} & -T_{24} \\ -T_{31} & -T_{32} & T_{31}+T_{32}+T_{34} & -T_{34} \\ -T_{41} & -T_{42} & -T_{43} & T_{41}+T_{42}+T_{43} \end{pmatrix}$$

$$\times \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{pmatrix} = \frac{h}{2e^2}\begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix} = \frac{h}{2e^2}\begin{pmatrix} -I \\ I \\ 0 \\ 0 \end{pmatrix}. \tag{5.117}$$

We know that one of these equations is redundant, so drop that for I_1. We also know that only differences of voltages are significant, so we can set one to zero; V_3 is a convenient choice as we want V_{43}. We are then left with a 3×3 set of equations,

$$\begin{pmatrix} -T_{21} & T_{21}+T_{23}+T_{24} & -T_{24} \\ -T_{31} & -T_{32} & -T_{34} \\ -T_{41} & -T_{42} & T_{41}+T_{42}+T_{43} \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \\ V_4 \end{pmatrix} = \frac{h}{2e^2}\begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix}. \tag{5.118}$$

This can be solved simply using Cramer's rule or the like, since the matrix is now well behaved. The result is

$$R_{21,43} = \frac{h}{2e^2}\frac{T_{42}T_{31} - T_{41}T_{32}}{S}, \tag{5.119}$$

where S is the determinant of the 3×3 matrix in equation (5.118). A worry is that eliminating different variables from the original 4×4 set might give different forms of this result. Fortunately the numerator does not change, and the sum rule that requires that the rows and columns all sum to zero means that any 3×3 submatrix of the original matrix has the same determinant (although one has to watch the sign!). Another useful result is the two-probe resistance, deduced from the voltage between the current probes,

$$R_{21,21} = \frac{h}{2e^2}\frac{(T_{31} + T_{32} + T_{34})(T_{41} + T_{42} + T_{43}) - T_{34}T_{43}}{S}. \tag{5.120}$$

These results will be used in Section 6.6.1 to study the propagation of edge states in the quantum Hall effect.

An interesting application, of great historical importance, is to compare the resistance of a tunnelling barrier measured using two or four probes. The system is shown in Figure 5.26(a), where the sample is the barrier in the middle with transmission coefficient T. The voltage probes are assumed to be identical and very weakly coupled to the structure to cause minimum disturbance. Assume for simplicity that only one mode propagates through the structure.

We need the transmission coefficients, of which there are six assuming time-reversal invariance. The largest is $T_{12} = T_{21} = T$ due to the barrier. All others involve coupling to the voltage probes 3 and 4 and are small by assumption. The transmission coefficients from voltage probe 3 to the other probes are shown in Figure 5.26(b). Let $T_{31} = \alpha$ and $T_{32} = \beta$, which are both of order δ, say. The third coefficient $T_{34} = \gamma$ will be of order δ^2, as it describes propagation through both of the weakly transmitting voltage contacts. The coefficients for probe 4 are the same but for reflection symmetry. The determinant of the matrix in equation (5.118) is

$$S \approx \det \begin{vmatrix} -T & T & -\alpha \\ -\alpha & -\beta & -\gamma \\ -\beta & -\alpha & \alpha+\beta \end{vmatrix} \approx T(\alpha+\beta)^2, \tag{5.121}$$

retaining only terms to lowest order. The two-probe resistance (equation 5.120) becomes

$$R_{21,21} \approx \frac{h}{2e^2}\frac{(\alpha+\beta)^2 - \gamma^2}{S} \approx \frac{h}{2e^2}\frac{1}{T}; \qquad G_{2\text{-probe}} = \frac{2e^2}{h}T. \tag{5.122}$$

This is a familiar result. The four-probe resistance (equation 5.119) is

$$R_{21,43} \approx \frac{h}{2e^2}\frac{\alpha^2 - \beta^2}{S} \approx \frac{h}{2e^2}\left(\frac{\alpha-\beta}{\alpha+\beta}\right)\frac{1}{T}. \tag{5.123}$$

As in the three-lead example, the voltage between the weakly coupled probes depends on the ratio of their transmission coefficients in the two directions. Figure 5.26(c) shows how to determine this ratio. Current from probe 3 divides equally in the two directions when it leaves the voltage probe, with transmission coefficients δ. One branch of this current flows into lead 1 without impediment but the other branch encounters the barrier in the middle of the device. A fraction R is reflected by this and only a fraction T passes through to reach probe 2; we can ignore the small effect of probe 4, which it passes on the way. Thus $\alpha = (1+R)\delta$ and $\beta = T\delta$. This finally gives

$$R_{21,43} \approx \frac{h}{2e^2}\frac{R}{T}; \qquad G_{4\text{-probe}} = \frac{2e^2}{h}\frac{T}{R} = \frac{2e^2}{h}\frac{T}{1-T}. \tag{5.124}$$

This result is due to Landauer, and the difference between the two-probe and four-probe results was long a source of controversy. There is little difference between $G_{2\text{-probe}}$ and $G_{4\text{-probe}}$ for a weakly transmitting barrier but they disagree strongly as the barrier becomes more transparent. In the limiting case of $T = 1$ we have $G_{2\text{-probe}} = 2e^2/h$ but $G_{4\text{-probe}} = \infty$. What is the source of the difference?

If there is no barrier at all and $T = 1$, it seems clear that the distribution of electrons should be the same at all points within the wire and that a probe used to measure the voltage should return the same value at any point, giving $G_{4\text{-probe}} = \infty$. The difference with $G_{2\text{-probe}}$ is that the voltages reflect the Fermi levels of the *reservoirs*. There must be a difference between these Fermi levels in order to drive a current. The current that leaves a reservoir is proportional to the product of the density of states at the Fermi level, the Fermi velocity, and the difference in Fermi levels, and this product is finite. Thus a non-zero voltage must be applied simply to generate the non-equilibrium distribution of electrons needed to pass a current, even if that current is then transmitted perfectly to the other end of the sample. This extra voltage appears to be due to an extra contact resistance of $h/2e^2$ in series with the sample. It also reminds us that energy must be supplied to support conduction, even through a perfect wire, but leaves open the question of how this energy is dissipated.

5.8 Tunnelling in Heterostructures

We have assumed throughout the previous sections that the structures consist of a single material with a superposed potential. This is not strictly applicable to a heterostructure such as a barrier of AlGaAs surrounded by GaAs. In simple cases the changes needed are very similar to those for quantum wells, discussed in Section 4.9. For example, the wave function of a layered structure still separates into a transverse plane wave and a longitudinal part that can be treated with T-matrices, but the energy no longer separates totally and the effective height of a barrier depends on the transverse wave vector $\mathbf{k}$. The matching conditions at a heterointerface must