CHAPTER 2

Quantum Point Contacts

H. van Houten and C. W. J. Beenakker

PHILIPS RESEARCH LABORATORIES
EINDHOVEN THE NETHERLANDS

B J. van Wees

DEPARTMENT OF APPLIED PHYSICS
DELFT UNIVERSITY OF TECHNOLOGY
DELFT THE NETHERLANDS

I INTRODUCTION 9
II SPLIT-GATE QUANTUM POINT CONTACTS 13
III BALLISTIC QUANTUM TRANSPORT 17
1 *Introduction* 17
2 *Conductance Quantization of a Quantum Point Contact* 19
3 *Magnetic Depopulation of Subbands* 33
4 *Magnetic Suppression of Backscattering at a Point Contact* 39
5 *Electron Beam Collimation and Point Contacts in Series* 45
6 *Coherent Electron Focusing* 56
7 *Breakdown of the Conductance Quantization and Hot Electron Focusing* 64
IV ADIABATIC TRANSPORT IN THE QUANTUM HALL EFFECT REGIME 81
8 *Introduction* 81
9 *Anomalous Quantum Hall Effect* 84
10 *Anomalous Shubnikov–de Haas Effect* 94
11 *Aharonov Bohm Oscillations and Inter-Edge Channel Tunneling* 99
ACKNOWLEDGMENTS 105
REFERENCES 106

I. Introduction

The subject of this chapter is quasi-one-dimensional quantum transport Only a few years ago, a prevalent feeling was that there is a "limited purpose in elaborating on playful one-dimensional models" for quantum transport [1] This situation has changed drastically since the realization of the quantum point contact, which now offers ample opportunity to study transport problems of textbook simplicity in the solid state Interestingly, many of the phenomena

ISBN 0 12 752135 6

treated in this chapter were not anticipated theoretically, even though the were understood rapidly after their experimental discovery.

In this chapter, we review the experimental and theoretical work by th Philips–Delft collaboration on electrical transport through quantum poir contacts. These are short and narrow constrictions in a two-dimensiona electron gas (2DEG), with a width of the order of the Fermi wave length λ_1 Throughout our presentation, we distinguish between ballistic and adiabati transport. *Ballistic quantum transport* takes place in low magnetic fields, fc which Landau level quantization is unimportant and the Fermi wavelengt ($\lambda_F \approx 40$ nm) governs the quantization. In stronger fields in the quantu Hall effect regime, the Landau-level quantization dominates, characterize by the magnetic length ($l_m \equiv (\hbar/eB)^{1/2} \approx 10$ nm at $B = 5$ *T*). In the latter re gime, inter-Landau-level scattering can be suppressed and *adiabatic quantu transport* may be realized. Because of the high mobility, elastic impurit scattering and inelastic scattering are of secondary importance in the ballisti and adiabatic transport regimes. Scattering is determined instead by th geometry of the sample boundary. The concept of a mean free path thus lose much of its meaning, and serves only as an indication of the length scale o which ballistic transport can be realized. (The transport mean free path i weak magnetic fields is about 10 μm in wide 2DEG regions.) Fully adiabati transport in strong magnetic fields has been demonstrated over a shor distance of the order of a μm, but may be possible on longer length scales Separate and more detailed introductions to these two transport regimes ar given in Part III (which is concerned with ballistic quantum transport) an Part IV (where adiabatic quantum transport is discussed). The following i intended only to convey the flavor of the subject, and to give an elementar introduction to some of the essential characteristics.

The common starting point for the structures investigated is the degenerat two-dimensional electron gas (2DEG), present at the interface between GaA and $Al_xGa_{1-x}As$ layers in a heterostructure. (Experimental details are given ir Part II; for general reviews of the 2DEG, see Refs. 2 and 3.) The electrons ar confined in the GaAs by a potential well at the interface with the AlGaAs which results from the repulsive barrier due to the conduction band offse between the two semiconductors (about 0.3 eV), and from the attractive electrostatic potential due to the positively charged ionized donors in the AlGaAs layer. The electrons thus are confined in a direction normal to the interface, but free to move along the interface. This implies that a two-dimensional subband is associated with each discrete confinement level in the well. Usually, the potential well is sufficiently narrow (about 10 nm) that only a single two-dimensional subband is occupied, and the density of states is strictly two-dimensional. At low temperatures, these states are occupied up to

the Fermi energy, $E_F \approx 10$ meV. Additional confinement occurs in a lateral direction if a narrow channel is defined electrostatically in the 2DEG. This leads to the formation of one-dimensional subbands, characterized by free motion in a single direction.

Throughout this chapter we will use a magnetic field perpendicular to the 2DEG as a tool to modify the nature of the quantum states. In a wide 2DEG, a perpendicular magnetic field eliminates the two degrees of freedom, and forms dispersionless Landau levels (which correspond classically to the motion of electrons in cyclotron orbits). One thus has a purely discrete density of states. Near the boundary of the 2DEG, the Landau levels transform into magnetic edge channels, which are free to move along the boundary, and correspond classically to skipping orbits. These edge channels have a one-dimensional dispersion (i.e., the energy depends on the momentum along the boundary). In this respect, they are similar to the one-dimensional subbands resulting from a purely electrostatic lateral confinement in a channel. Because of the one-dimensional dispersion law, both edge channels and one-dimensional subbands can be viewed as propagating modes in an electron waveguide. This similarity allows a unified description of the quantum Hall effect and of quantum-size effects in narrow conductors in the ballistic transport regime.

A really unequivocal and striking manifestation of a quantum-size effect on the conductance of a single narrow conductor came, paradoxically, with the experimental realization by the Delft–Philips collaboration[4] and by the Cambridge group[5] of the quantum point contact—a constriction that one would have expected to be too short for one-dimensional subbands to be well-developed. A major surprise was the nature of the quantum-size effect: The conductance of quantum point contacts is quantized in units of $2e^2/h$. This is reminiscent of the quantum Hall effect, but measured in the absence of a magnetic field. The basic reason for the *conductance quantization* (a fundamental cancellation of group velocity and density of states for quantum states with a one-dimensional dispersion law) already was appreciated in the original publications. More complete explanations came quickly thereafter, in which the mode-coupling with the wide 2DEG at the entrance and exit of the narrow constriction was treated explicitly. Rapid progress in the theoretical understanding of the conductance quantization, and of its subsequent ramifications, was facilitated by the availability of a formalism,[6,7] which turned out to be ideally suited for quasi-one-dimensional transport problems in the ballistic and adiabatic transport regimes. The Landauer–Büttiker formalism treats transport as a transmission problem for electrons at the Fermi level. The ohmic contacts are modeled as current injecting and collecting reservoirs, in which all inelastic scattering is thought to take place exclusively. As described by Büttiker in Chapter 4, the measured conductances then can be expressed as

treated in this chapter were not anticipated theoretically, even though they were understood rapidly after their experimental discovery.

In this chapter, we review the experimental and theoretical work by the Philips–Delft collaboration on electrical transport through quantum point contacts. These are short and narrow constrictions in a two-dimensional electron gas (2DEG), with a width of the order of the Fermi wave length λ_F. Throughout our presentation, we distinguish between ballistic and adiabatic transport. *Ballistic quantum transport* takes place in low magnetic fields, for which Landau level quantization is unimportant and the Fermi wavelength ($\lambda_F \approx 40$ nm) governs the quantization. In stronger fields in the quantum Hall effect regime, the Landau-level quantization dominates, characterized by the magnetic length ($l_m \equiv (\hbar/eB)^{1/2} \approx 10$ nm at $B = 5$ *T*). In the latter regime, inter-Landau-level scattering can be suppressed and *adiabatic quantum transport* may be realized. Because of the high mobility, elastic impurity scattering and inelastic scattering are of secondary importance in the ballistic and adiabatic transport regimes. Scattering is determined instead by the geometry of the sample boundary. The concept of a mean free path thus loses much of its meaning, and serves only as an indication of the length scale on which ballistic transport can be realized. (The transport mean free path in weak magnetic fields is about 10 μm in wide 2DEG regions.) Fully adiabatic transport in strong magnetic fields has been demonstrated over a short distance of the order of a μm, but may be possible on longer length scales. Separate and more detailed introductions to these two transport regimes are given in Part III (which is concerned with ballistic quantum transport) and Part IV (where adiabatic quantum transport is discussed). The following is intended only to convey the flavor of the subject, and to give an elementary introduction to some of the essential characteristics.

The common starting point for the structures investigated is the degenerate two-dimensional electron gas (2DEG), present at the interface between GaAs and $Al_xGa_{1-x}As$ layers in a heterostructure. (Experimental details are given in Part II; for general reviews of the 2DEG, see Refs. 2 and 3.) The electrons are confined in the GaAs by a potential well at the interface with the AlGaAs, which results from the repulsive barrier due to the conduction band offset between the two semiconductors (about 0.3 eV), and from the attractive electrostatic potential due to the positively charged ionized donors in the AlGaAs layer. The electrons thus are confined in a direction normal to the interface, but free to move along the interface. This implies that a two-dimensional subband is associated with each discrete confinement level in the well. Usually, the potential well is sufficiently narrow (about 10 nm) that only a single two-dimensional subband is occupied, and the density of states is strictly two-dimensional. At low temperatures, these states are occupied up to

the Fermi energy, $E_F \approx 10$ meV. Additional confinement occurs in a lateral direction if a narrow channel is defined electrostatically in the 2DEG. This leads to the formation of one-dimensional subbands, characterized by free motion in a single direction.

Throughout this chapter we will use a magnetic field perpendicular to the 2DEG as a tool to modify the nature of the quantum states. In a wide 2DEG, a perpendicular magnetic field eliminates the two degrees of freedom, and forms dispersionless Landau levels (which correspond classically to the motion of electrons in cyclotron orbits). One thus has a purely discrete density of states. Near the boundary of the 2DEG, the Landau levels transform into magnetic edge channels, which are free to move along the boundary, and correspond classically to skipping orbits. These edge channels have a one-dimensional dispersion (i.e., the energy depends on the momentum along the boundary). In this respect, they are similar to the one-dimensional subbands resulting from a purely electrostatic lateral confinement in a channel. Because of the one-dimensional dispersion law, both edge channels and one-dimensional subbands can be viewed as propagating modes in an electron waveguide. This similarity allows a unified description of the quantum Hall effect and of quantum-size effects in narrow conductors in the ballistic transport regime.

A really unequivocal and striking manifestation of a quantum-size effect on the conductance of a single narrow conductor came, paradoxically, with the experimental realization by the Delft–Philips collaboration[4] and by the Cambridge group[5] of the quantum point contact—a constriction that one would have expected to be too short for one-dimensional subbands to be well-developed. A major surprise was the nature of the quantum-size effect: The conductance of quantum point contacts is quantized in units of $2e^2/h$. This is reminiscent of the quantum Hall effect, but measured in the absence of a magnetic field. The basic reason for the *conductance quantization* (a fundamental cancellation of group velocity and density of states for quantum states with a one-dimensional dispersion law) already was appreciated in the original publications. More complete explanations came quickly thereafter, in which the mode-coupling with the wide 2DEG at the entrance and exit of the narrow constriction was treated explicitly. Rapid progress in the theoretical understanding of the conductance quantization, and of its subsequent ramifications, was facilitated by the availability of a formalism,[6,7] which turned out to be ideally suited for quasi-one-dimensional transport problems in the ballistic and adiabatic transport regimes. The Landauer–Büttiker formalism treats transport as a transmission problem for electrons at the Fermi level. The ohmic contacts are modeled as current injecting and collecting reservoirs, in which all inelastic scattering is thought to take place exclusively. As described by Büttiker in Chapter 4, the measured conductances then can be expressed as

rational functions of the transmission probabilities at the Fermi level between the reservoirs. The zero-field conductance quantization of an ideal one-dimensional conductor, and the smooth transition to the quantum Hall effect on applying a magnetic field, are seen to follow directly from the fact that a reservoir in equilibrium injects a current that is shared equally by all propagating modes (which can be one-dimensional subbands or magnetic edge channels).

Novel phenomena arise if a selective, non-equal distribution of current among the modes is realized instead. In the ballistic transport regime, directional selectivity can be effected by a quantum point contact, as a result of its horn-like shape and of the potential barrier present in the constriction.[8] The collimation of the electron beam injected by the point contact explains the strong non-additivity of the series resistance of two opposite point contacts observed in Ref. 9. A most striking manifestation of a non-equal distribution of current among the modes is realized in the adiabatic transport regime, where the selective population and detection of magnetic edge channels is the mechanism for the anomalous quantum Hall and Shubnikov–de Haas effects.[10–12]

Mode interference is another basic phenomenon. Its first unequivocal manifestation in quantum transport is formed by the large (nearly 100%) conductance oscillations found in the coherent electron focusing experiment.[13–15] They may be considered as the ballistic counterpart of the conductance fluctuations characteristic of the diffusive transport regime. In the adiabatic transport regime, mode interference is less important, because of the weakness in general of inter-edge channel coupling. Quantum interference phenomena still can be observed if a weak coupling exists between the edge channels at opposite edges of the conductor. Such a coupling can result naturally from the presence of an impurity in a narrow channel, or artificially at quantum point contacts. In this way, Aharonov–Bohm magnetoresistance oscillations can occur even in a singly connected geometry.[16,17]

In summary, transport phenomena in the ballistic and adiabatic regimes can be viewed as scattering or transmission experiments with modes in an electron waveguide. Quantization—i.e., the discreteness of the mode index—is essential for some phenomena (which necessarily require a description in terms of modes), but not for others (which could have been described semiclassically equally well in terms of the trajectories of electrons at the Fermi level). In this chapter, we consider the semiclassical limit along with a quantum mechanical formulation wherever this is appropriate. This serves to distinguish those aspects of the new phenomena that are intrinsically quantum mechanical from those that are not.

Most of the work described in the following has been published previously. The present chapter is the first comprehensive review. Many new details are

added, and a critical overview of experimental as well as theoretical aspects is provided. This is not intended to be a comprehensive review of the whole field of quasi-one-dimensional quantum transport. Because of the limited amount of space and time available, we have not included a detailed discussion of related work by other groups (some of which is described extensively in other chapters in this volume). For the same reason, we have excluded work by ourselves and others on the quasi-ballistic transport regime, and on ballistic transport in narrow-channel geometries. For a broader perspective, we refer readers to a review[18] and to recent conference proceedings.[19–22]

II. Split-Gate Quantum Point Contacts

The study of ballistic transport through point contacts in metals has a long history. Point contacts in metals act like small conducting orifices in a thin insulating layer, separating bulk metallic conductors (with a mean free path l much larger than the size of the orifice). Actual point contacts usually are fabricated by pressing a metal needle on a metallic single crystal, followed by spot-welding. Ballistic transport has been studied successfully in this way in a variety of metals.[23–26] Point contacts in bulk doped semiconductors have been fabricated by pressing two wedge-shaped specimens close together.[27] One limitation of these techniques is that the size of a point contact is not continuously variable.

Point contacts in a 2DEG cannot be fabricated by the same method, since the electron gas is confined at the GaAs-$Al_xGa_{1-x}As$ interface in the sample interior. The point contacts used in our studies are defined electrostatically[28,29] by means of a split gate on top of the heterostructure. (See Fig. 1a.) In this way, one can define short and narrow constrictions in the 2DEG, of variable width comparable to the Fermi wavelength (a quantum point contact). Other techniques can be used to define constrictions of fixed width, such as a deep[30] or shallow[31] mesa etch, or ion implantation using focused ion beams,[32] but a variable constriction width is crucial for our purpose. (An alternative technique for the fabrication of variable width constrictions employing a gate in the plane—rather than on top—of the 2DEG recently has been demonstrated.)[33] Starting point for the fabrication of our quantum point contact structures is a GaAs-$Al_xGa_{1-x}As$ heterostructure ($x = 0.3$) grown by molecular beam epitaxy. The layer structure is drawn schematically in Fig. 1b. The width of the opening in the gate is approximately 250 nm, its length being much shorter (50 nm). The 2DEG sheet carrier density n_s, obtained from the periodicity of the Shubnikov–de Haas oscillations in the magnetoresistance,[2] has a typical value of 3.5×10^{15} m^{-2}.

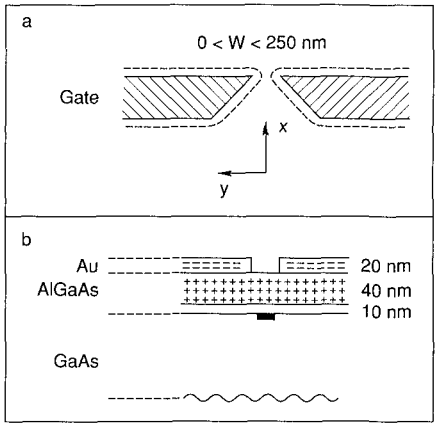

FIG 1 (*a*) Top view of a quantum point contact, defined using a split gate (shaded) on top of a GaAs-$Al_xGa_{1-x}As$ heterostructure The depletion boundary is indicated by the dashed curve The width W of the constriction can be reduced by increasing the negative voltage on the gate (*b*) Cross section of the quantum point contact The narrow quasi-one-dimensional electron gas channel in the constriction is indicated in black The positive ionized donors (+) in the AlGaAs layer are indicated, as well as the negative charge (−) on the gate

The electrons at the Fermi level then have a wave vector $k_F \equiv (2\pi n_s)^{1/2} \approx 0.15 \times 10^9\ m^{-1}$, a wavelength $\lambda_F \equiv 2\pi/k_F \approx 40$ nm, and a velocity $v_F \equiv \hbar k_F/m \approx 2.7 \times 10^5$ m/s. The transport mean free path $l \approx 10\ \mu$m follows from the zero-field resistivity $\rho \equiv h/e^2 k_F l \approx 16\ \Omega$. Note that $m = 0.067 m_e$ is the effective mass in GaAs. Most of the experimental work presented in this chapter has been done on samples made by the Philips–Delft collaboration. An exception is formed by the experiments described in Sections 4 and 9.a.ii, which were done on a sample fabricated in a collaboration between Philips and the Cavendish group.[34]

The fabrication procedure essentially is the same for all the samples. A standard mesa-etched Hall bar geometry is defined by wet etching. The split gate is fabricated using a combination of electron beam and optical lithography. The gate pattern of many of our samples has been designed specifically for the electron focusing experiments, which require two point contacts positioned next to each other on the 2DEG boundary. (See Fig. 2a.) Note that the actual 2DEG boundary between the two point contacts is a depletion potential wall below the gate (which extends laterally beyond the gate pattern, up to about 150 nm for large negative gate voltages). The effect of a negative gate voltage is to deplete gradually the electron gas under the gate structure. Beyond the depletion threshold (typically −0.6 V), no mobile carriers are present under the gate, and two conducting constrictions are formed with a width of about 250 nm. Two high-mobility 2DEG regions thus

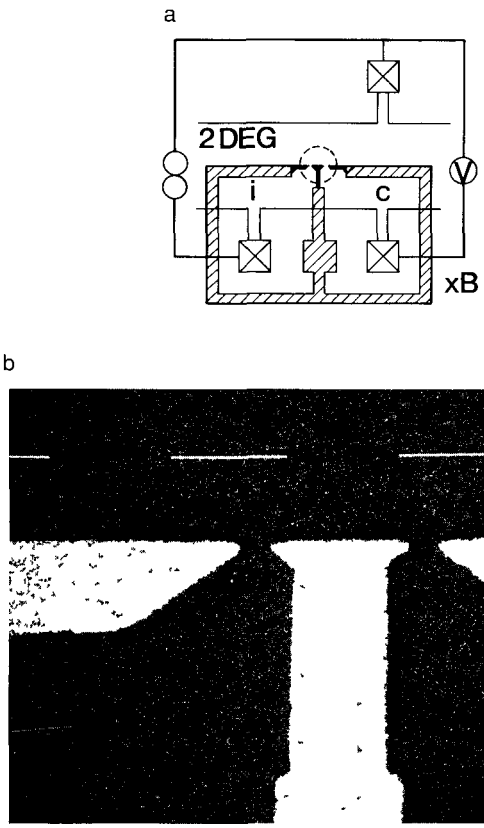

FIG 2 Schematic layout (*a*) of a double point contact device, in a three-terminal measurement configuration used in some of the electron focusing experiments The crossed squares are ohmic contacts to the 2DEG The split gate (shaded) separates injector (i) and collector (c) areas from the bulk 2DEG The fine details of the gate structure inside the dashed circle are shown in a scanning electron micrograph (*b*) The bar denotes a length of 1 μm (From Ref 15)

are isolated electrically from the rest of the 2DEG in the Hall bar, apart from the narrow constrictions, or point contacts, under the openings of the gate. A further increase of the negative gate voltage forces the constrictions to become progressively narrower, until they are fully pinched off. By this technique, it is possible to define point contacts of variable width W. To create constrictions with minimal length, the gates are tapered into a wedge. (See the scanning electron micrograph in Fig. 2b.) The precise functional dependence of width and carrier concentration on the gate voltage is dependent on the previous history of the sample. Thermal cycling and large positive gate voltages lead to a shift in the depletion threshold, although all transport measurements are

quite reproducible if the sample is kept cold and the gate voltage is not varied strongly.

A low-frequency ac lock-in technique is used to measure the resistances. Several ohmic contacts (alloyed Au-Ge-Ni) are positioned at the sides of the Hall bar (Fig. 2a), to serve as current and voltage terminals. The resistance $R_{ij,kl} \equiv (V_k - V_l)/I$ is defined as the voltage difference between terminal k and l divided by the current I, which flows from terminal i to j. One distinguishes between two- and four-terminal resistance measurements, depending on whether or not the voltage difference is measured between the current source and drain ($i, j \equiv k, l$), or between two separate ohmic contacts. Section 2 deals with two-terminal measurements of the point contact resistance in zero magnetic field. This resistance contains a spurious contribution of several $k\Omega$ from the rather large contact resistance of the current-carrying ohmic contacts. This correction can be estimated from a measurement of the two-terminal resistance at zero gate voltage, or can be eliminated entirely by performing a four-terminal measurement. Apart from the presence of this contact resistance, there is no significant difference between two- and four-terminal measurements in the absence of a magnetic field, provided the voltage probes do not introduce additional scattering in the vicinity of the point contact. In an external magnetic field, the behavior of two- and four-terminal resistances is quite different, however, as we will discuss in Sections 3 and 4.

In addition to the series resistance of the ohmic contacts, there are two additional small corrections to the quantized point contact resistance that are gate voltage-independent *beyond* the depletion threshold of the gate (-0.6 V), as we now discuss briefly. At the depletion threshold, the two-terminal resistance increases abruptly for three reasons:

1. The formation of the ballistic point contact, which is the quantity of interest.
2. The increase of the diffusive resistance of the wide 2DEG lead on one side of the constriction, because of a change in the lead geometry. (See the gate layout in Fig. 2a.) This term is $\rho \approx 16\ \Omega$ multiplied by the extra number of squares in the lead.
3. The appearance of the two-dimensional Maxwell spreading resistance,[35] associated with the spreading from a region of radius l (the mean free path) surrounding the point contact to one of radius W_{wide} (the width of the wide 2DEG leads). This term is approximately[1] $\pi^{-1}\rho \ln(W_{\text{wide}}/l)$.

The contributions 2 and 3 are independent of gate voltage beyond the depletion threshold to a very good approximation, but they are difficult to determine very accurately from the device geometry. For this reason, we

have opted to treat the total background resistance R_b in the two-terminal resistance R_{2t} of the point contact as a single adjustable parameter, chosen such that for one constant value of R_b, a uniform step height (between quantized plateaus) is obtained in the conductance $G = [R_{2t}(V_g) - R_b]^{-1}$ as a function of gate voltage V_g. This procedure always has yielded a uniform step height in G over the whole gate voltage range for a single value of R_b. Moreover, the resulting value of R_b is close to the value that one would have estimated from the preceding considerations.

III. Ballistic Quantum Transport

1. INTRODUCTION

In this section, we present a comprehensive review of the results of the study by the Philips–Delft collaboration of ballistic transport in geometries involving quantum point contacts in weak magnetic fields. To put this work in proper perspective, we first briefly discuss the two fields of research from which it has grown.

The first is that of point contacts in metals. Maxwell, in his *Treatise on Electricity and Magnetism*, investigated the spreading resistance of a small contact in the diffusive transport regime.[35] His results have been applied extensively in the technology of dirty metallic contacts.[36] The interest in point contacts gained new impetus with the pioneering work of Sharvin,[23] who proposed and subsequently realized[37] the injection and detection of a beam of electrons in a metal by means of point contacts much smaller than the mean free path. Sharvin's longitudinal electron focusing experiment was the analogue in the solid state of an experiment performed earlier in vacuum by Tricker[38] at the suggestion of Kapitza.[39] This technique since has been refined, especially with the introduction of the transverse electron focusing geometry by Tsoi.[24] (See Section 6.) Point contacts also can be used to inject electrons in a metal with an energy above the Fermi energy. This idea has been exploited in the field of point contact spectroscopy, and it has yielded a wealth of information on inelastic electron–phonon scattering.[25,26,40] Magnetotransport through ballistic point contacts and micro-bridges has been studied recently.[41,42] With the possible exception of the scanning tunneling microscope, which can be seen as a point contact on an atomic scale,[43–48] these studies in metals essentially are restricted to the *classical* ballistic transport regime, because of the extremely small Fermi wavelength ($\lambda_F \approx 0.5$ nm, of the same magnitude as the lattice spacing).

The second field is that of quasi-one-dimensional quantum transport in semiconductor microstructures, which started in the diffusive transport

regime in narrow silicon MOSFETs. That work has focused on the study of reproducible (universal) conductance fluctuations, as discussed in this volume. Clear manifestations of the quasi-one-dimensional density of states of a single narrow wire proved to be elusive, mainly because the irregular conductance fluctuations mask the structure due to the one-dimensional subbands in the wire. Devices containing many wires in parallel were required to average out these fluctuations and resolve the subband structure in a transport experiment.[49] This situation changed with the realization by various techniques of narrow channels in the two-dimensional electron gas (2DEG) of a GaAs-$Al_xGa_{1-x}As$ heterostructure.[28,29,31,50] This is an ideal model system because of the simple Fermi surface (a circle), the relatively long mean free path ($l \approx 10\ \mu m$ at low temperatures for material grown by molecular beam epitaxy), and the large Fermi wavelength ($\lambda_F \approx 40$ nm) resulting from the low electron density. Another essential advantage of this system is that its two-dimensionality allows the use of planar semiconductor technology to fabricate a rich variety of device structures. Finally, in contrast to metals, the low electron density in these semiconductor structures can be varied by means of a gate voltage. Thornton *et al.*[28] and Zheng *et al.*[29] have demonstrated that it is possible to realize structures of variable width and density by employing a split-gate lateral depletion technique. Other groups[51–54] have used the shallow mesa etch technique,[31] or other etch techniques,[55] to fabricate narrow channels of fixed width with many side probes for the study of quantum ballistic transport, as discussed by Timp in Chapter 3. An important result of these studies was the demonstration that in the ballistic transport regime, side probes are the dominant source of scattering.[56]

Our work on quantum ballistic transport builds on both fields summarized in the preceding. As discussed in Part I, the central vehicle for this investigation is the quantum point contact, a short and narrow constriction of variable width in the 2DEG, of dimensions comparable to λ_F and much smaller than l. This device yielded the first unequivocal demonstration of a quantum-size effect in a single narrow conductor,[4,5] in the form of the zero-field conductance quantization. We discuss the experiment and its theoretical explanation in Section 2. The quantization of the conductance provides us with an extremely straightforward way to determine the number of occupied subbands in the point contact. It is shown in Section 3 that a study of the magnetic depopulation of subbands directly yields the width and carrier density in the point contact.[57]

Two- and four-terminal resistance measurements are qualitatively different in the presence of an external magnetic field. The negative four-terminal magnetoresistance[34] arising from the suppression of backscattering at the point contact is the subject of Section 4. We then proceed to discuss in Section 5 the collimation of the beam injected by a point contact,[8] and its effect

on transport measurements in geometries involving two opposite point contacts.[9,58] The variety of magnetoresistance effects[59] in such geometries is even richer than for single-point contacts, as we will discuss. An important application of point contacts is as point-like electron sources and detectors with a large degree of spatial coherence. The first such application was the coherent electron focusing experiment[13] described in Section 6. This experiment exhibits the characteristic features of the quantum ballistic transport regime in a most extreme way. The results are interpreted in terms of mode interference of magnetic edge channels.[14] Ballistic transport far from equilibrium is the subject of Section 7, where we discuss the breakdown of the conductance quantization in the nonlinear transport regime,[60] and hot electron focusing.[61] In the latter experiment, the kinetic energy of the injected electrons in the 2DEG is measured in a similar way as in a β-spectrometer in vacuum.

2. CONDUCTANCE QUANTIZATION OF A QUANTUM POINT CONTACT

a. Experimental Observation of the Conductance Quantization

The first results on the resistance of a quantum point contact obtained by the Delft–Philips collaboration[4] are reproduced in Fig. 3. Equivalent results were obtained independently by the Cavendish group.[5] The resistance of the point contact is measured as a function of the voltage V_g on the split gate, at a temperature $T = 0.6$ K. The resistance measured at $V_g = 0$ V has been

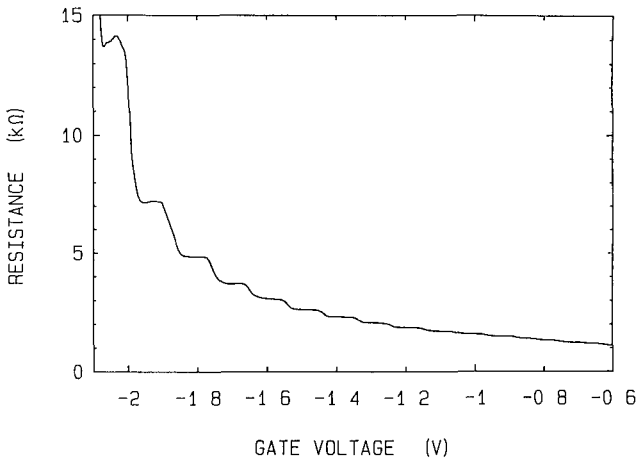

FIG. 3 Two-terminal point contact resistance measured as a function of gate voltage at 0.6 K. The resistance measured at zero gate voltage has been subtracted. (From Ref. 4.)

subtracted in Fig. 3, to eliminate the large ohmic contact resistance of about 4 kΩ in series with the point contact. At $V_g \approx -0.6$ V, the electron gas directly below the gate is depleted, and the constriction is formed. At this gate voltage, the constriction has its maximum width, roughly equal to the width of the opening in the gate (250 nm). On increasing the negative gate voltage, the width W gradually is reduced, and simultaneously the bottom of the conduction band is raised in the point contact region. The resulting bottleneck in real and energy space causes the point contact to have a nonzero resistance. This resistance increases without bound as the pinch-off voltage ($V_g \approx -2.2$ V) is approached. Classically, one expects this increase to be monotonic.

The unexpected characteristic of Fig. 3 is the sequence of plateaus and steps seen in the resistance versus gate voltage curve. The plateaus represent the conductance quantization of a quantum point contact in units of $2e^2/h$. This is seen most easily in Fig. 4, where the conductance is plotted (obtained by inverting the resistance of Fig. 3 after subtraction of an additional background resistance of 400 Ω, which accounts for the increase in lead resistance at the depletion threshold discussed in Part II). The conductance quantization is reminiscent of the quantum Hall effect,[62] but is observed in the absence of a magnetic field and thus can not have the same origin. The zero-field quantization is not as accurate as the quantum Hall effect. The deviations from exact quantization in the present experiments are estimated at 1%,[63] while in the quantum Hall effect, an accuracy of one part in 10^7 is obtained routinely.[64] It is very unlikely that in the case of the zero-field quantization, a similar accuracy can be achieved—if only because of the presence in series

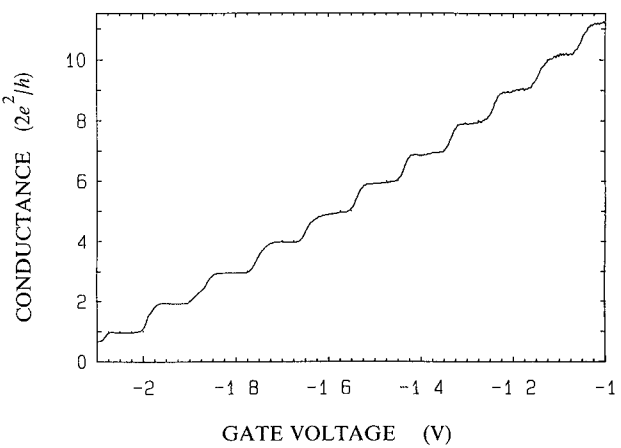

FIG. 4. Point contact conductance, obtained from the resistance in Fig. 3 after subtraction of an additional background resistance of 400 Ω. (From Ref. 4.)

with the point contact resistance of a background resistance whose magnitude can not be determined precisely. Both the degree of flatness of the plateaus and the sharpness of the transition between the plateaus vary among devices of identical design, indicating that the detailed shape of the electrostatic potential defining the constriction is important. There are many uncontrolled factors affecting this shape, such as small changes in the gate geometry, variations in the pinning of the Fermi level at the free GaAs surfaceor at the interface with the gate metal, not fully homogeneous doping of the heterostructure, and trapping of charges in the AlGaAs.

As will be discussed in Section 2.b, the sequence of plateaus is caused by the stepwise decrease of the number N of occupied one-dimensional subbands as the point contact gradually is pinched off, each subband contributing $2e^2/h$ to the conductance. In a simple approximation, the constriction is modeled as a straight channel of length L and width W, with a square-well lateral confining potential. The bottom of the well is at a height E_c above the conduction band bottom in the wide 2DEG. The density n_c in the constriction thus is reduced from the bulk density n_s by approximately a factor $(E_F - E_c)/E_F$. (This factor assumes a constant two-dimensional density of states in the constriction.) The stepwise reduction of N is due both to the decrease in W and the increase in E_c (or, equivalently, the reduction of n_c). If the latter effect is ignored, then the number of occupied subbands in the square well is $2W/\lambda_F$, with $\lambda_F = 40$ nm the Fermi wavelength in the wide 2DEG. The sequence of steps in Fig. 4 then would correspond to a gradual decrease in width from 320 nm (at $V_g \approx -1.0$ V) to 20 nm (at $V_g \approx -2.0$ V). This simple argument certainly overestimates the reduction in width, however, because of the unjustified neglect of the reduction in carrier density. By applying a perpendicular magnetic field, W and n_c can be determined independently, as discussed in Section 3.b. The length of the constriction is harder to assess, but the electrostatic depletion technique used is expected to create a constriction of length L, which increases with increasing negative gate voltage. Typically, $L \gtrsim W$. The actual two-dimensional shape of the confining potential certainly is smoother than a straight channel with hard walls. Nonetheless, for many applications, this simple model is adequate, and we will make use of it unless a more realistic potential is essential.

b. *Theory of the Conductance Quantization*

i. Classical point contact. It is instructive first to consider *classical* ballistic transport in a degenerate electron gas in some detail. The ballistic electron flow through a point contact is illustrated in Fig. 5a in real space, and in Fig. 5b in k-space, for a small excess electron density δn at one side of the point contact.[65,66] At low temperatures, this excess charge moves with

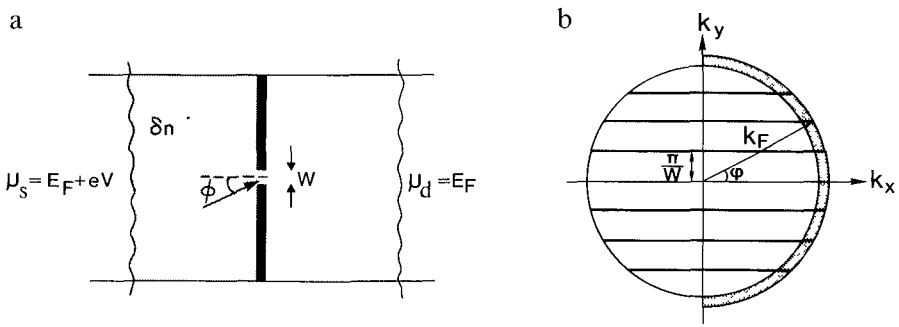

FIG. 5. (*a*) Classical ballistic transport through a point contact. The net concentration difference δn corresponds to a chemical potential difference eV between source (s) and drain (d). In reality, this concentration difference is eliminated by screening charges, but without changing the chemical potential difference or the current. (*b*) The net current through a quantum point contact is carried by the shaded region in k-space. In an ideal quasi-one-dimensional channel, the allowed states lie on the horizontal lines, which correspond to quantized values for $k_y = \pm n\pi/W$, and continuous values for k_x. The formation of these one-dimensional subbands gives rise to a quantized conductance. (From Refs. 65 and 66.)

the Fermi velocity v_F. The flux normally incident on the point contact is $\delta n\, v_F \langle \cos\phi\, \theta(\cos\phi)\rangle$, where $\theta(x)$ is the unit step function and the brackets denote an isotropic angular average. (The angle ϕ is defined in Fig. 5a.) In the ballistic limit $l \gg W$, the incident flux is fully transmitted, so that the total current I through the point contact is given by

$$I = eW\,\delta n\, v_F \int_{-\pi/2}^{\pi/2} \cos\phi \frac{d\phi}{2\pi} = \frac{e}{\pi} W v_F\, \delta n. \tag{1}$$

The transport problem in this formulation resembles the problem of Knudsen effusion in a non-degenerate gas. A fundamental difference is that in the former problem, I is linear in δn only for small $\delta n \ll n_s$, whereas this restriction is not necessary in the Knudsen problem. This distinction is a consequence of the interdependence of the velocity and the density in a degenerate gas described by Fermi–Dirac statistics.[66] We will return to this point in Section 7, when we discuss the nonlinear transport regime. To determine the conductance, we note that an electron density difference can not be maintained, because of the cost in electrostatic energy. Screening charges reduce δn without changing the chemical potential difference $\delta\mu \equiv eV$, which is assumed to be fixed by an external electron reservoir. The ratio $\delta n/\delta\mu$ is just the density of states at the Fermi level, $\delta n/\delta\mu = m/\pi\hbar^2$ for a two-dimensional electron gas with spin degeneracy. One thus finds from Eq. (1) for the con-

ductance $G \equiv I/V$, the result,[4]

$$G = \frac{2e^2}{h} \frac{k_F W}{\pi}, \text{ in 2D.} \tag{2}$$

Equation (2) is the two-dimensional analogue of Sharvin's well-known expression[23] for the point contact conductance in three dimensions,

$$G = \frac{2e^2}{h} \frac{k_F^2 S}{4\pi}, \text{ in 3D,} \tag{3}$$

where S now is the area of the point contact.

The experimental constriction geometry differs from the hole in a screen of Fig. 5a, in having a finite length with a smoothly varying width W, and an electron gas density that decreases on entering the constriction. The reduced density leads to a smaller value for k_F in the constriction than in the wide 2DEG. Equation (2) still can be applied to this situation, if the product $k_F W$ is evaluated at the *bottleneck* (such that all electrons that reach the bottleneck are transmitted through the constriction). This typically is halfway into the constriction, where k_F and W take on their minimal values.

ii. Conductance quantization of an ideal quasi-one-dimensional conductor. The basic mechanism for the quantization of the conductance given classically by Eq. (2) can be understood in quite simple terms.[4] The argument, which we present here in a somewhat modified form, refers to an *ideal* quasi-one-dimensional conductor that behaves as an electron waveguide connecting two reservoirs in thermal equilibrium at chemical potentials E_F and $E_F + \delta\mu$. All inelastic scattering is thought to take place in the reservoirs, not in the conductor itself. This is the viewpoint introduced by Landauer.[6] The Landauer formula relates the conductance to the transmission probability through the conductor from one reservoir to the other. The net current is injected into the conductor within a narrow range $\delta\mu$ above E_F into the N one-dimensional subbands or waveguide modes that can propagate at these energies. The dispersion relation $E_n(k)$ of the subbands is

$$E_n(k) = E_n + \frac{\hbar^2 k^2}{2m}, \tag{4}$$

where k is the wave number for propagation along the conductor and E_n is the energy associated with the lateral confinement of the nth subband.

(Equivalently, $E_n/\hbar$ is the cutoff frequency of the nth mode.) The number N of occupied subbands (or propagating modes) at the Fermi energy is the largest integer such that $E_N < E_F$. The current per unit energy interval injected into a subband is the product of the group velocity and the one-dimensional density of states. The group velocity is $v_n = dE_n(k)/\hbar\, dk$, and the density of states for one velocity direction and including spin degeneracy is $\rho_n = (\pi\, dE_n(k)/dk)^{-1}$. The product of v_n and ρ_n is seen to be independent of both energy and subband index n. The injected current thus is *equipartitioned* among the subbands, each subband carrying the same amount of current $e v_n \rho_n \delta\mu = (2e/h)\delta\mu$. The equipartitioning of current, which is the basic mechanism for the conductance quantization, is illustrated in Fig. 5b for a square-well lateral confining potential of width W. The one-dimensional subbands then correspond to the pairs of horizontal lines at $k_y = \pm n\pi/W$, with $n = 1, 2, \ldots N$ and $N = \mathrm{Int}[k_F W/\pi]$. The group velocity $v_n = \hbar k_x/m$ is proportional to $\cos\phi$, and thus decreases with increasing n. However, the decrease in v_n is compensated by an increase in the one-dimensional density of states. Since ρ_n is proportional to the length of the horizontal lines within the dashed area in Fig. 5b, ρ_n is proportional to $1/\cos\phi$ so that the product $v_n\rho_n$ does not depend on the subband index.

The total current $I = (2e/h)N\,\delta\mu$ yields a conductance $G = eI/\delta\mu$ given by

$$G = \frac{2e^2}{h} N. \tag{5}$$

This equation can be seen as a special limit of the Landauer formula for two-terminal conductances,[7,67–69]

$$G = \frac{2e^2}{h} \mathrm{Tr}\, \mathbf{t}\mathbf{t}^\dagger \equiv \frac{2e^2}{h} \sum_{n,m=1}^{N} |t_{nm}|^2, \tag{6}$$

where $\mathbf{t}$ is the matrix (with elements t_{nm}) of transmission probability amplitudes at the Fermi energy (from subband m at one reservoir to subband n at the other). The result, Eq. (5), follows from Eq. (6) if $\mathrm{Tr}\, \mathbf{t}\mathbf{t}^\dagger = N$. A sufficient condition for this is the absence of intersubband scattering, $|t_{nm}|^2 = \delta_{nm}$, a property that may be taken to define the ideal conductor. More generally, scattering among the subbands is allowed as long as it does not lead to backscattering (i.e., for zero reflection coefficients $r_{nm} = 0$ for all $n, m = 1, 2, \ldots N$). Equation (5) describes a stepwise increase in the conductance of an ideal quasi-one-dimensional conductor as the number of occupied subbands is increased. The conductance increases by $2e^2/h$ each time N increases by 1. For a square-well lateral confining potential, $N = \mathrm{Int}[k_F W/\pi]$, so that in the

classical limit, the result (2) for a two-dimensional point contact is recovered. Note that Eq. (5) also holds for three-dimensional point contacts, although in that case, no experimental system showing the conductance quantization as yet has been realized.

We emphasize that, although the classical formula, Eq. (2), holds only for a square-well lateral confining potential, the quantization, Eq. (5), is a general result for any shape of the confining potential. The reason simply is that the fundamental cancellation of the group velocity, $v_n = dE_n(k)/\hbar\, dk$, and the one-dimensional density of states, $\rho_n = (\pi\, dE_n(k)/dk)^{-1}$, holds *regardless* of the form of the dispersion relation $E_n(k)$. For the same reason, Eq. (5) is applicable equally in the presence of a magnetic field, when magnetic edge channels at the Fermi level take over the role of one-dimensional subbands. Equation (5) thus implies a continuous transition from the zero-field quantization to the quantum Hall effect, as we will discuss in Section 3.

The fact that the Landauer formula, Eq. (6), yields a *finite* conductance for a perfect (ballistic) conductor was a source of confusion in the early literature,[70–72] but now is understood as a consequence of the unavoidable contact resistances at the connection of the conductor to the reservoirs. The relation between ballistic point contacts and contact resistances of order h/e^2 for a one-dimensional subband first was pointed out by Imry.[68] This was believed to be only an *order of magnitude* estimate of the point contact resistance. One reason for this was that the Landauer formula follows from an idealized model of a resistance measurement; another one was that several multi-subband generalizations of the original Landauer formula[6] had been proposed,[67,73–76] which led to conflicting results. We refer to a paper by Stone and Szafer[69] for a discussion of this controversy, which now has been settled[7,77–79] in a way supported by the present experiments. This brief excursion into history may serve as a partial explanation of the fact that no prediction of the conductance quantization of a point contact was made, and why its experimental discovery came as a surprise.

iii. Conductance quantization of a quantum point contact. There are several reasons why the ideal quasi-one-dimensional conductor model given earlier, and related models,[80–82] are not fully satisfactory as an explanation of the experimentally observed conductance quantization. Firstly, to treat a point contact as a waveguide would seem to require a constriction much longer than wide, which is not the case in the experiments,[4,5] where $W \approx L$. (See Section 2.d. for a discussion of experiments in longer constrictions.) In a very short constriction, transmission through evanescent modes (with cutoff frequency above $E_F/\hbar$) becomes important. Secondly, the coupling of the modes in the wide 2DEG regions to those in the constriction has not been considered explicitly. In particular, diffraction and quantum mechanical

reflection at the entrance and exit of the constriction have been ignored. Finally, alloyed ohmic contacts, in combination with those parts of the wide 2DEG leads that are more than an inelastic scattering length away from the constriction, only are approximate realizations of the reservoirs in thermal equilibrium of the idealized problem.[83] As an example, electrons may be scattered back into the constriction by an impurity without intervening equilibration. This modifies the point contact conductance, as has been studied extensively in the classical case.[26]

To resolve these issues, it is necessary to solve the Schrödinger equation for the wave functions in the narrow point contact and the adjacent wide regions, and match the wave functions at the entrance and exit of the constriction. Following the experimental discovery of the quantized conductance, this mode coupling problem has been solved numerically for point contacts of a variety of shapes,[84–91] and analytically in special geometries.[92–95] As described in detail in Ref. 96, the problem has a direct and obvious analogue in the field of classical electromagnetism. Although it can be solved by standard methods (which we will not discuss here), the resulting transmission steps appear not to have been noted before in the optical or microwave literature. When considering the mode coupling at the entrance and exit of the constriction, it is important to distinguish between the cases of a gradual (*adiabatic*) and an *abrupt* transition from wide to narrow regions.

The case of an *adiabatic* constriction has been studied by Glazman *et al.*[97] and is the easiest case to solve analytically (cf. also a paper by Imry in Ref. 20). If the constriction width W changes gradually on the scale of a wavelength, the transport within the constriction is adiabatic; i.e., there is no intersubband scattering from entrance to exit. At the exit, where connection is made to the wide 2DEG regions, intersubband scattering becomes unavoidable and the adiabaticity breaks down. However, if the constriction width at the exit $W_{\max}$ is much larger than its minimal width $W_{\min}$, the probability for reflection back through the constriction becomes small. In the language of waveguide transmission, one has impedance-matched the constriction to the wide 2DEG regions.[98] Since each of the N propagating modes in the narrowest section of the constriction is transmitted without reflection, one has $\mathrm{Tr}\,\mathbf{t}\mathbf{t}^{\dagger} = N$, provided evanescent modes can be neglected. The conductance quantization, Eq. (5), then follows immediately from the Landauer formula, Eq. (6). Glazman *et al.*[97] have calculated that the contributions from evanescent modes through an adiabatic constriction are small even for rather short constriction length L (comparable to $W_{\min}$). The accuracy of the conductance quantization for an adiabatic constriction in principle can be made arbitrarily high, by widening the constriction more and more slowly, such that $(W_{\max} - W_{\min})/L \ll 1$. In practice, of course, the finite mean free path still poses a limitation. We note in this connection that the gradual widening of

the constriction has an interesting effect on the angular distribution of the electrons injected into the wide 2DEF.[8] This *horn collimation* effect is discussed in detail in Section 5.

An adiabatic constriction is not necessary to observe the quantization of the conductance. The calculations[84–95] show that well-defined conductance plateaus persist for *abrupt* constrictions—even if they are rather short compared to the width. In fact an optimum length for the observation of the plateaus is found to exist, given by[85] $L_{\rm opt} \approx 0.4\ (W\lambda_{\rm F})^{1/2}$. In shorter constrictions, the plateaus acquire a finite slope, although they do not disappear completely even at zero length. For $L > L_{\rm opt}$, the calculations exhibit regular oscillations that depress the conductance periodically below its quantized value. The oscillations are damped and usually have vanished before the next plateau is reached. A thermal average rapidly smears the oscillations and leads to smooth but non-flat plateaus. The plateaus disappear completely at elevated temperatures, when the thermal energy becomes comparable to the subband splitting. (See Section 2.c.) The plateaus also do not survive impurity scattering, either inside or near the constriction.[85,99,100]

Physical insight in these results can be obtained by treating the conduction through the constriction as a transmission problem, on the basis of the Landauer formula, Eq. (6). In the case of adiabatic transport discussed before, we had the simple situation that $|t_{nn'}|^2 = \delta_{n,n'}$ for $n, n' \leq N$, and zero otherwise. For an abrupt constriction, this is no longer true, and we have to consider the partial transmission of all the modes occupied in the wide regions. Semiclassically, the transverse momentum $\hbar n\pi/W$ of mode n is conserved at the abrupt transition from wide to narrow region. We thus can expect the coupling between modes n and n' in the narrow and wide regions (of width $W_{\rm min}$ and $W_{\rm wide}$), respectively, to be strongest if $n/n' \sim W_{\rm min}/W_{\rm wide}$. This leads to a large increase in mode index at the exit of an abrupt constriction. Szafer and Stone[87] have formulated a *mean-field* approximation that exploits such ideas by assuming that a particular propagating or evanescent mode n in the constriction couples exclusively and uniformly to all modes n' in the wide region for which the energy $E_{n'} = (\hbar k_{n'})^2/2m$ of transverse motion is within a level splitting of E_n. Figure 6 contrasts the mode coupling for the abrupt constriction with the case of fully adiabatic transport from $W_{\rm min}$ to $W_{\rm wide}$. Whereas in the adiabatic case, there is a one-to-one correspondence between the modes in the narrow and in the wide regions, in the abrupt case a mode in the constriction couples to a larger number (of order $W_{\rm wide}/W_{\rm min}$) of modes in the wide region.

Because of the abrupt widening of the constriction, there is a significant probability for backscattering at the exit of the constriction, in contrast to the adiabatic case considered previously. The conductance as a function of width, or Fermi energy, therefore is not a simple step function. On the nth

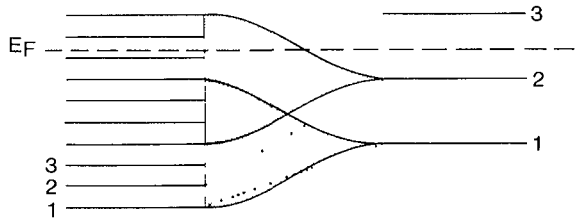

FIG 6 Mode coupling between a constriction and a wide 2DEG region The subband energies E_n are spaced closely in the wide region at the left For an abrupt constriction, off-diagonal mode coupling is important (indicated by the shaded areas in the mean-field approximation of Ref 87 The coupling is restricted between modes of the same parity.), while for an adiabatic constriction, this does not occur (dotted lines)

conductance plateau, backscattering occurs predominantly for the nth mode, since it has the largest longitudinal wavelength, $\lambda_n = h[2m\ (E_F - E_n)]^{-1/2}$. Resonant transmission of this mode occurs if the constriction length L is approximately an integer multiple of $\lambda_n/2$, and leads to the oscillations on the conductance plateaus found in the calculations referred to earlier. These transmission resonances are damped, because the probability for backscattering decreases with decreasing λ_n. The shortest value of λ_n on the nth conductance plateau is $h[2m\ (E_{n+1} - E_n)]^{-1/2} \approx (W\lambda_F)^{1/2}$ (for a square-well lateral confining potential). The transmission resonances thus are suppressed if $L \lesssim (W\lambda_F)^{1/2}$ (disregarding numerical coefficients of order unity). Transmission through evanescent modes, on the other hand, is predominant for the $(n + 1)$th mode, since it has the largest decay length $\Lambda_{n+1} = h[2m\ (E_{n+1} - E_F)]^{-1/2}$. The observation of a clear plateau requires that the constriction length exceed this decay length at the population threshold of the nth mode, or $L \gtrsim h[2m\,(E_{n+1} - E_n)]^{-1/2} \approx (W\lambda_F)^{1/2}$. The optimum length,[85] $L_{opt} \approx 0.4\ (W\lambda_F)^{1/2}$, thus separates a short constriction regime, in which transmission via evanescent modes cannot be ignored, from a long constriction regime, in which transmission resonances obscure the plateaus.

c. Temperature Dependence of the Conductance

i. Thermal averaging of the point contact conductance. In Fig. 7, we show[12] the conductance of a quantum point contact in zero magnetic field as a function of gate voltage, for various temperatures between 0.3 K and 4.2 K. On increasing the temperature, the plateaus acquire a finite slope until they no longer are resolved. This is a consequence of the thermal smearing of the Fermi–Dirac distribution,

$$f(E - E_F) = \left(1 + \exp\frac{E - E_F}{k_B T}\right)^{-1}.$$

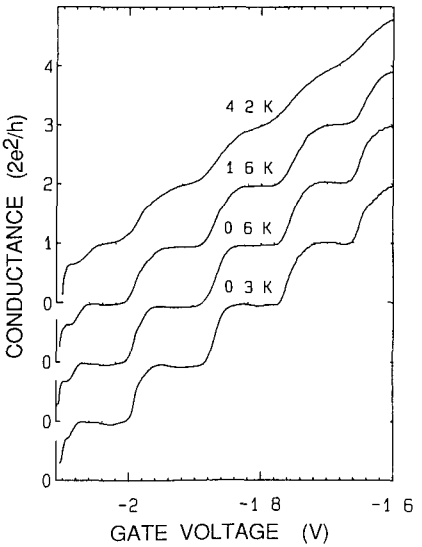

FIG. 7. Experimental temperature dependence of the conductance quantization in zero magnetic field. (From Ref. 12.)

If at $T = 0$ the conductance $G(E_F, T)$ has a step function dependence on the Fermi energy E_F, at finite temperatures it has the form,[80,101]

$$G(E_F, T) = -\int G(E, 0)\frac{df}{dE}dE = \frac{2e^2}{h}\sum_{n=1}^{\infty} f(E_n - E_F). \tag{7}$$

Here, as before, E_n denotes the energy of the bottom of the nth subband (*cf*, Eq. (4)). The width of the thermal smearing function df/dE is about $4k_BT$, so that the conductance steps should disappear above a characteristic temperature $T_{char} \approx \Delta E/4k_B$, with ΔE the subband splitting at the Fermi level. For the square-well confining potential, $\Delta E \approx 2(E_F - E_c)/N$. In Section 3.b, we estimate that ΔE increases from about 2 meV at $V_g = -1.0$ V (where $N = 11$) to 4 meV at $V_g = -1.8$ V (where $N = 3$). The increase in subband splitting thus qualitatively explains the experimental observation in Fig. 7 that the smearing of the plateaus is less pronounced for larger negative gate voltages. The temperature at which smearing becomes appreciable (≈ 4 K) implies $\Delta E \approx 2$ meV, which is of the correct order of magnitude.

It has been noted that a small but finite voltage drop across the constriction should have an effect that is qualitatively similar to that of a finite temperature.[101] This indeed is borne out by experiments.[12] Conduction at larger applied voltages in the nonlinear transport regime is discussed extensively in Section 7.

ii. Quantum interference effects at low temperatures. Interestingly, it was found experimentally[4,5] that, in general, a finite temperature yielded the best well-defined and flat plateaus as a function of gate voltage in the zero-field conductance. If the temperature is increased beyond this optimum (which is about 0.5 K), the plateaus disappear because of the thermal averaging discussed earlier. Below this temperature, oscillatory structure may be superimposed on the conductance plateaus, as demonstrated in Fig. 8, which shows[12] conductance traces at 40 mK (both in the absence and presence of a weak magnetic field). The strength and shape of the oscillations varies from device to device, probably due to the uncontrolled variations in the confining potential discussed in Part II. However, the data is quite reproducible if the sample is kept below 10 K. We believe that these oscillations are due at least in part to resonances in the transmission probability associated with reflections at the entrance and exit of the constriction. Indeed, similar oscillations were found in the numerical studies referred to earlier of the conductance of an abrupt constriction with $L \approx W$. Other groups[102,103] have measured comparable fine structure in the quantum point contact conductance.

In addition to these resonances, some of the structure may be a quantum interference effect associated with backscattering of electrons by impurities

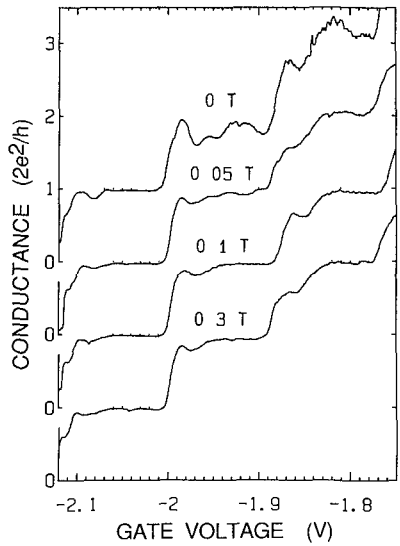

FIG. 8. Oscillatory structure observed in the zero-field conductance of a point contact at 40 mK (top curve). Some of the oscillatory structure is suppressed by a weak magnetic field (lower three curves). (From Ref. 12.)

near the opening of the constriction. The possibility that impurity scattering plays a role is supported by the fact that a very weak perpendicular magnetic field of 0.05 T leads to a suppression of some of the finest structure, leaving the more regular oscillations unchanged (Fig. 8). Increasing the magnetic field further has little effect on the flatness of the plateaus. The cyclotron radius for these fields is as long as 2 μm, so that the magnetic field hardly has any effect on the electron states *inside* the constriction. Such a field would be strong enough, however, to suppress the backscattering caused by one or a few impurities located within a few μm of the constriction. In contrast to the case of the conductance fluctuations in the diffusive transport regime,[104,105] the specific impurity configuration would be very important.

We thus believe that this data shows evidence of both impurity-related quantum interference oscillations and transmission resonances determined by the geometry. Only the latter survive in a weak perpendicular magnetic field. Provided this interpretation is correct, one in principle can estimate the length of the constriction from the periodicity of the relevant oscillations as a function of gate voltage. For a realistic modeling, one has to account for the complication that the gate voltage simultaneously affects the carrier density in the constriction, its width, and its length. Such calculations are not available, unfortunately. The effect of an increase in temperature on these quantum interference effects can be two-fold. Firstly, it leads to a suppression of the oscillations because of thermal averaging. Secondly, it reduces the phase coherence length as a result of inelastic scattering. The coherent electron focusing experiment discussed in Section 6 indicates that the latter effect is relatively unimportant for quantum ballistic transport at temperatures up to about 10 K. At higher temperatures, inelastic scattering induces a gradual transition to incoherent diffusive transport.

d. *Length Dependence of the Conductance*

Theoretically, one expects that the conductance quantization is preserved in longer channels than those used in the original publications[4,5] (in which, typically, $L \sim W \sim 100$ nm). Experiments on longer channels, however, did not show the quantization.[34,63,106] This is demonstrated in Fig. 9, where the resistance versus gate voltage is plotted[34] for a constriction with $L = 3.4$ μm. This is well below the transport mean free path in the bulk, which is about 10 μm in this material. The curve in Fig. 9 was taken at 50 mK, but the resistance is temperature-independent below 4 K. The sudden increase in the resistance at $V_g = -0.5$ V indicates the formation of the constriction. The lack of clear plateaus in Fig. 9 (compared with Fig. 3 for a short constriction) most likely is due to enhanced backscattering inside the constriction. Impurity

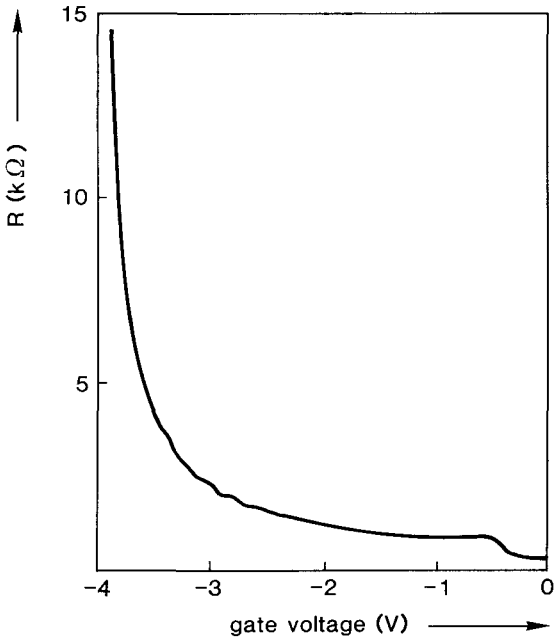

FIG. 9. Resistance of a long constriction ($L = 3.4$ μm) as a function of gate voltage, at T = 50 mK, showing the near absence of quantized plateaus. The shoulder at $V_g \approx -0.5$ V is a consequence of the formation of the constriction at the depletion threshold. (From Ref. 34.)

scattering may be one source of backscattering,[63,106] which is expected to be more severe in narrow channels due to the reduced screening in a quasi-one-dimensional electron gas.[107] Perhaps more importantly, backscattering can occur at channel wall irregularities. Thornton *et al.*[108] have found evidence of a small (5%) fraction of diffuse, rather than specular, reflections at boundaries defined electrostatically by a gate. In a 200 nm-wide constriction, this leads to an effective mean free path of about 200 nm/0.05 $\approx$ 4 μm, comparable to the constriction length in this device.

Long constrictions have been studied more extensively in the *quasi-ballistic* transport regime, where the mean free path is much larger than the channel width, but shorter than its length (*cf.* Refs. 50 and 109). Low-temperature transport in this regime is characterized by weak localization and electron–electron interaction effects, and by universal conductance fluctuations. It would be of interest to study the transition from the ballistic to the quasi-ballistic transport regime, by performing systematic studies on the length and width dependence of the quantum transport through smooth constrictions fabricated on material with different values for the mobility. Some recent work by Timp *et al.*[63,106] and Brown *et al.*[102] is in this direction.

3. MAGNETIC DEPOPULATION OF SUBBANDS

a. *Magneto-Electric Subbands*

If a magnetic field B is applied perpendicular to a wide 2DEG, the kinetic energy of the electrons is quantized[110] at energies $E_n = (n - \frac{1}{2})\hbar\omega_c$, with $\omega_c = eB/m$ the cyclotron frequency. The quantum number $n = 1, 2, \ldots$ labels the Landau levels. The number of Landau levels below the Fermi energy $N \sim E_F/\hbar\omega_c$ decreases as the magnetic field is increased. This *magnetic depopulation* of Landau levels is observed in the quantum Hall effect, where each occupied Landau level contributes e^2/h (per spin direction) to the Hall conductance. The Landau level quantization is the result of the periodicity of the circular motion in a magnetic field. In a narrow channel or constriction, the cyclotron orbit is perturbed by the electrostatic lateral confinement, and this modifies the energy spectrum. Instead of Landau levels, one now speaks of *magneto-electric subbands.* The effect of the lateral confinement on the number N of occupied subbands becomes important when the cyclotron orbit at the Fermi energy (of radius $l_{cycl} = \hbar k_F/eB$) no longer fits fully into the channel. If $l_{cycl} \gg W$, the effect of the magnetic field on the trajectories (and thus on the energy spectrum) can be neglected, and N becomes approximately B-independent. Simple analytic expressions for the B-dependence of N can be obtained for a parabolic confining potential,[111] or for a square-well potential.[15] For the square well, one finds in a semiclassical approximation (with an accuracy of ± 1), and neglecting the spin-splitting of the energy levels,

$$N \approx \mathrm{Int}\left[\frac{2}{\pi}\frac{E_F}{\hbar\omega_c}\left(\arcsin\frac{W}{2l_{cycl}} + \frac{W}{2l_{cycl}}\left[1 - \left(\frac{W}{2l_{cycl}}\right)^2\right]^{1/2}\right)\right], \quad \text{if } l_{cycl} > \frac{W}{2}, \tag{8a}$$

$$N \approx \mathrm{Int}\left[\frac{1}{2} + \frac{E_F}{\hbar\omega_c}\right], \text{ if } l_{cycl} < \frac{W}{2}. \tag{8b}$$

One easily verifies that for zero magnetic field, Eq. (8) yields $N = \mathrm{Int}[k_F W/\pi]$, as it should. If Eq. (8) is applied to a constriction containing a potential barrier of height E_c, then one should replace $E_F \to E_F - E_c$ and, consequently, $l_{cycl} \to l_{cycl}(1 - E_c/E_F)^{1/2}$. In Fig. 10, we show the depopulation of Landau levels with its characteristic $1/B$ dependence of N (dashed curve), and the much slower depopulation of magneto-electric subbands for $W/2l_{cycl} < 1$ (solid curve). These results are calculated from Eq. (8) for a square-well potential with $k_F W/\pi = 10$. Smoother confining potentials (e.g., parabolic) give similar results.[111]

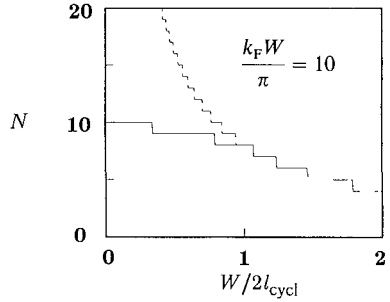

FIG. 10. Magnetic field dependence of the number of occupied subbands in a narrow channel, according to Eq. (8) (solid curve). The dashed curve gives the magnetic depopulation of Landau levels in a wide 2DEG, which has a $1/B$ dependence.

We note that in Fig. 10, a possible oscillatory B-dependence of E_F has been ignored, which would result from pinning of the Fermi level to the Landau levels—either in the narrow channel itself or in the adjacent wide 2DEG regions. To determine this B-dependence for a short constriction (where both pinning mechanisms compete) would require a self-consistent solution of the Schrödinger and Poisson equation, which has not been done yet in a quantizing magnetic field for such a geometry. In the application of Eq. (8) to the experiments in Section 3.b on a constriction containing a barrier, we similarly will neglect a possible oscillatory B-dependence of $E_F - E_c$.

In the Landau gauge for the vector potential $A = (0, Bx, 0)$ (for a channel along the y-axis), the translational invariance along the channel is not broken by the magnetic field, so that the propagating modes can still be described by a wave number k for propagation along the channel—just as in zero magnetic field (*cf.* Section 2.b). However, the dispersion relation $E_n(k)$ does not have the form of Eq. (4), and consequently, the group velocity $v_n \equiv dE_n/\hbar dk$ no longer is given by $\hbar k/m$ (as it is for $B = 0$). In a strong magnetic field ($l_{cycl} \lesssim W/2$), the propagating modes are extended along a boundary of the sample, and are referred to as magnetic *edge channels*. Classically, these states correspond to skipping orbits along a channel boundary (*cf.* Fig. 11a). In weaker fields ($l_{cycl} \gtrsim W/2$), the propagating modes extend throughout the bulk, and correspond to *traversing trajectories* that interact with both opposite channel boundaries (Fig. 11b). The wave functions and energy spectra for these various quantum states are very different, yet experimentally a gradual transition is observed from the zero-field conductance quantization of a quantum point contact to the strong-field quantum Hall effect. (See next section). The fundamental cancellation between group velocity and density of states for one-dimensional waveguide modes, which does not depend explicitly on the nature of the dispersion law $E_n(k)$, provides the theoretical explanation of the remarkable connection between these two quantum phenomena, which at first sight seem unrelated.

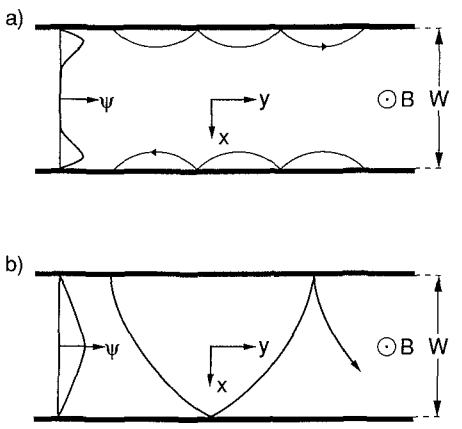

FIG. 11. Trajectories in a narrow channel in a perpendicular magnetic field (right) and the corresponding transverse profile of the wave function Ψ (left). Skipping orbits on both opposite edges and the corresponding edge states are shown in *a*, a traversing trajectory and the corresponding bulk state in *b*. Note that the wave functions shown correspond to the nodeless $n = 1$ mode.

b. Conductance Quantization in an External Magnetic Field

In Fig. 12, measurements[57] are shown of the conductance versus gate voltage for various values of the magnetic field (at $T = 0.6$ K). The point contact conductance has been obtained from the measured resistance after subtraction of a gate voltage-independent background resistance (*cf.* Part II). The measurements have been performed for values of the magnetic field where the 2DEG resistivity has a Shubnikov–de Haas minimum. The background resistance then is due mainly to the non-ideal ohmic contacts, and increases from about 4 kΩ to 8 kΩ between zero and 2.5 T.[57] Fig. 12 demonstrates that the conductance quantization is conserved in the presence of a magnetic field, and shows a smooth transition from zero-field quantization to quantum Hall effect. The main effect of the magnetic field is to reduce the number of plateaus in a given gate voltage interval. This provides a direct demonstration of depopulation of 1D subbands, as analyzed later. In addition, one observes that the flatness of the plateaus improves in the presence of the field. This is due to the spatial separation at opposite edges of the constriction of the left- and right-moving electrons (illustrated in Fig. 11a), which reduces the probability for backscattering in a magnetic field.[34,77] We return to the magnetic suppression of backscattering in Section 4. Finally, in strong magnetic fields, the spin degeneracy of the energy levels is removed, and additional plateaus appear at *odd* multiples of e^2/h. They are much less well-resolved than the even-numbered plateaus, presumably because the Zeeman

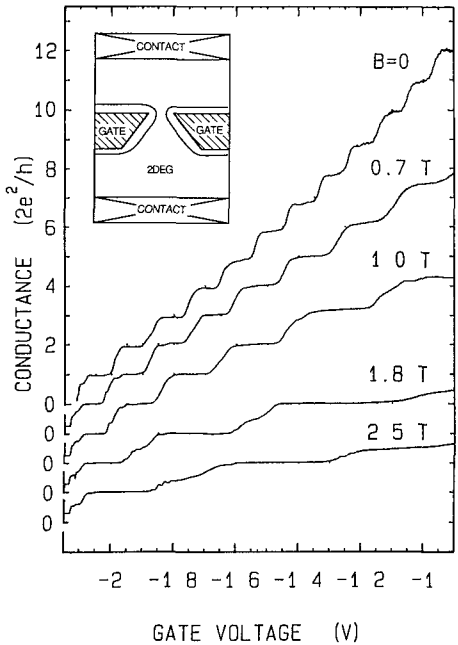

FIG. 12. Point contact conductance (corrected for a series lead resistance) as a function of gate voltage for several magnetic field values, illustrating the transition from zero-field quantization to quantum Hall effect. The curves have been offset for clarity. The inset shows the device geometry. (From Ref. 57.)

spin-splitting energy $|g\mu_B B|$ is considerably smaller than the subband splitting ΔE. (If one uses the low-field value $g = -0.44$ for the Landé g-factor in GaAs, and the definition $\mu_B = e\hbar/2m_e$ for the Bohr magneton, one finds a splitting as small as 0.025 meV per T, while ΔE in general is more than 1 meV, as discussed later.) We note that the spin degeneracy of the quantized plateaus also can be removed by a strong parallel (rather than perpendicular) magnetic field, as shown by Wharam *et al.*[5]

Because the arguments leading to Eq. (5) are valid regardless of the nature of the subbands involved, we can conclude that in the presence of a magnetic field, the conductance remains quantized according to $G = (2e^2/h)N$ (ignoring spin-splitting, for simplicity). Calculations[95,112–114] done for specific point contact geometries confirm this general conclusion. The number of occupied (spin degenerate) subbands N is given approximately by Eq. (8), for a square-well confining potential. In the high-magnetic field regime $W \gtrsim 2l_{cycl}$, the quantization of G with N given by Eq. (8b) is just the quantum Hall effect in a two-terminal configuration (which has been shown[115–117] to be equivalent to the quantization of the Hall resistance in the more usual four-

terminal configuration, *cf.* Section 9.a; the ohmic contact resistance is exceptionally large in our sample, but usually is much smaller so that accurate quantization becomes possible in a two-terminal measurement in high magnetic fields). At lower magnetic fields, the quantization of the point contact conductance provides a direct and extremely straightforward method to measure, via $N = G(2e^2/h)^{-1}$, the depopulation of magneto-electric subbands in the constriction. Previously, this effect in a narrow channel had been studied indirectly by measuring the deviations from the $1/B$ periodicity of the Shubnikov–de Haas oscillations[118–120] (the observation of which is made difficult by the irregular conductance fluctuations that result from quantum interference in a disordered system).

Figure 13 shows the number N of occupied subbands obtained from the measured G (Fig. 12), as a function of reciprocal magnetic field for various gate voltages.[57] Also shown are the theoretical curves according to Eq. (8), with the potential barrier in the constriction taken into account. The barrier

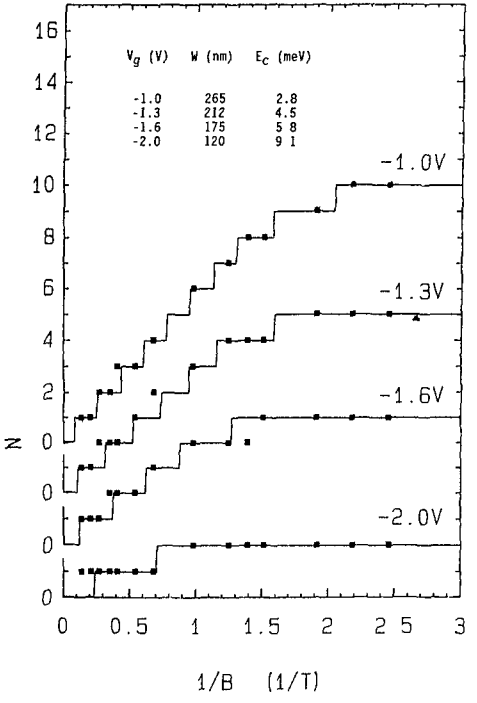

FIG. 13. Number of occupied subbands as a function of reciprocal magnetic field for several values of the gate voltage. Data points have been obtained directly from the quantized conductance (Fig. 12); solid curves are calculated from Eq. (8), with the parameters tabulated in the inset. (From Ref. 57.)

height E_c is obtained from the high-field conductance plateaus (where $N \approx (E_F - E_c)/\hbar\omega_c$), and the constriction width W then follows from the zero-field conductance (where $N \approx [2m(E_F - E_c)/\hbar^2]^{1/2}W/\pi$) The good agreement found over the entire field range confirms our expectation that the quantized conductance is determined exclusively by the number of occupied subbands—irrespective of their electric or magnetic origin The present analysis is for a square-well confining potential For the narrowest constrictions, a parabolic potential should be more appropriate, it has been used to analyze the data of Fig 12 in Refs 12 and 121 The most realistic potential shape is a parabola with a flat section inserted in the middle,[122 123] but this potential contains an additional undetermined parameter (the width of the flat section) Wharam *et al* [124] have analyzed their depopulation data using such a model (*cf* also Ref 121) Because of the uncertainties in the actual shape of the potential, the parameter values tabulated in Fig 13 only are rough estimates, but we believe that the observed trends in the dependence of W and E_c on V_g are significant

In Fig 14, we have plotted this trend (assuming a square-well confining potential) for the point contact discussed before (curves labeled 2) and for another (nominally identical) point contact (curves 3) For comparison, we also show the results obtained in Section 4 for a longer and wider constriction [34] (curves 1) The electron density n_c in the constriction has been calculated approximately by $n_c \approx (E_F - E_c)m/\pi\hbar^2$ (i e , using the two-dimensional density of states, with neglect of the subband quantization) The dependence of the width and electron density on the gate voltage is qualitatively similar for the three devices The quantitative differences between the two nominally identical quantum point contacts (curves 2 and 3) serve to emphasize the

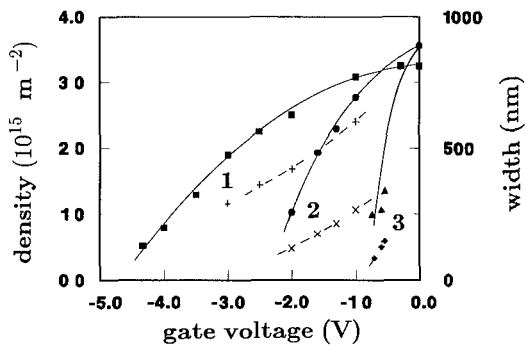

FIG 14 Electron gas density (solid curves) and width (dashed curves) of the constrictions defined by three different split gate devices The curves labeled 1 are for the wide split gate of Ref 34 (discussed in Section 4) Curves 2 are for the point contact of Ref 57 (*cf* Fig 13), and curves 3 for another point contact of identical design

importance of the uncontrolled variations in the device electrostatics discussed in Part II. (It should be noted, though, that curve 2 is representative for several other samples studied.) The larger constriction (curve 1) needs a much higher gate voltage for pinch-off simply because of its different dimensions. It would be of interest to compare these results with a self-consistent solution of the three-dimensional Poisson and Schrödinger equation, which now are starting to become available.[122,123]

A significant reduction of the electron density n_c in the constriction with increasing negative gate voltage occurs in all the samples (*cf.* Fig. 14). The potential barrier in the constriction thus cannot be neglected (except at low gate voltages). As an example, one finds for a typical quantum point contact (Fig. 13 or curve 2 in Fig. 14) that E_c/E_F varies from 0 to 0.7 (with $E_F = 12.7$ meV) as the gate voltage is varied from 0 to -2.0 V. This corresponds to a reduction of n_c by a factor of 3.5. Because of the relatively large potential barrier, the N-dependence of the zero-field subband splitting at the Fermi energy $\Delta E \approx 2(E_F - E_c)/N$ for a small number of occupied subbands in the square well is found to be substantially reduced from the $1/N$ dependence that would follow on ignoring the barrier. For the typical sample mentioned previously, one finds at $V_g = -1.8$ V, where $N = 3$, a subband splitting $\Delta E \approx 3.5$ meV. This is only a factor of 2 larger than the splitting $\Delta E \approx 1.8$ meV that one finds at $V_g = -1.0$ V, although $N = 11$ has increased by almost a factor of 4.

4. MAGNETIC SUPPRESSION OF BACKSCATTERING AT A POINT CONTACT

Only a small fraction of the electrons injected by the current source into the 2DEG is transmitted through the point contact. The remaining electrons are scattered back into the source contact. This is the origin of the nonzero resistance of a ballistic point contact. In this section, we shall discuss how a relatively weak magnetic field leads to a suppression of the *geometrical backscattering* caused by the finite width of the point contact, while the amount of backscattering caused by the potential barrier in the point contact remains essentially unaffected.

The reduction of backscattering by a magnetic field is observed as a *negative* magnetoresistance (i.e., $R(B) - R(0) < 0$) in a *four-terminal* measurement of the point contact resistance.[34] The distinction between two- and four-terminal resistance measurements already has been mentioned in Part II. In Sections 2 and 3, we considered the two-terminal resistance R_{2t} of a point contact. This resistance is the total voltage drop between source and drain divided by the current, and has a particular significance as the quantity that

determines the dissipated power I^2R_{2t}. Two-terminal resistance measurements, however, do not address the issue of the *distribution* of the voltage drop along the sample. In the ballistic (or adiabatic) transport regime, the measurement and analysis of the voltage distribution are non-trivial, because the concept of a local resistivity tensor (associated with that of local equilibrium) breaks down. (We will discuss non-local transport measurements in ballistic and adiabatic transport in Section 6 and Part IV, respectively.) In this section, we are concerned with the four-terminal *longitudinal* resistance R_L, measured with two adjacent (not opposite) voltage probes, one at each side of the constriction (*cf.* the inset in Fig. 15). We speak of a (generalized) longitudinal resistance, by analogy with the longitudinal resistance measured in a Hall bar, because the line connecting the two voltage probes does not intersect the line connecting the current source and drain (located at the far left and right of the conductor shown in Fig. 15). The voltage probes are positioned on wide 2DEG regions, well away from the constriction. This allows the establishment of local equilibrium near the voltage probes, at least in weak magnetic fields (*cf.* Part IV), so that the measured four-terminal resistance does not depend on the properties of the probes.

The experimental results[34] for R_L in this geometry are plotted in Fig. 15. This quantity shows a negative magnetoresistance, which is temperature-independent (between 50 mK and 4 K), and is observed in weak magnetic fields once the narrow constriction is defined (for $V_g \lesssim 0.3$ V). (The very small

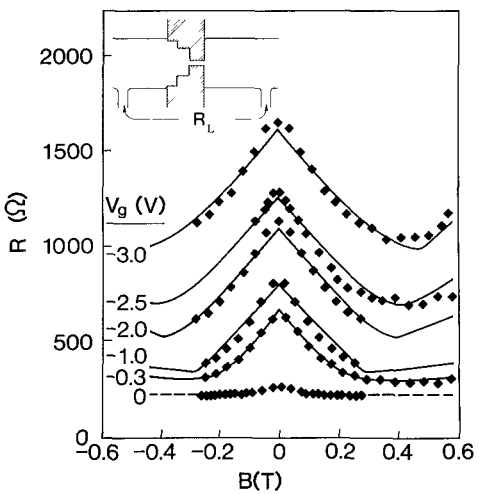

FIG. 15. Four-terminal longitudinal magnetoresistance R_L of a constriction for a series of gate voltages from 0 V (lowest curve) to -3 V. Solid lines are according to Eqs. (8) and (10), with the constriction width as adjustable parameter. The inset shows schematically the device geometry, with the two voltage proves used to measure R_L. (From Ref. 34.)

effect seen in the trace for $V_g = 0$ V probably is due to a density reduction by the Schottky barrier under the gate.) At stronger magnetic fields ($B > 0.4$ T), a crossover is observed to a positive magnetoresistance. The zero-field resistance, the magnitude of the negative magnetoresistance, the slope of the positive magnetoresistance, as well as the crossover field, all increase with increasing negative gate voltage.

The magnetic field dependence of the four-terminal resistance shown in Fig. 15 is qualitatively different from that of the two-terminal resistance R_{2t} considered in Section 3. In fact, R_{2t} is approximately B-independent in weak magnetic fields (below the crossover fields of Fig. 15). We recall that R_{2t} is given by (*cf.* Eq. (5))

$$R_{2t} = \frac{h}{2e^2} \frac{1}{N_{\min}}, \tag{9}$$

with $N_{\min}$ the number of occupied subbands at the bottleneck of the constriction (where it has its minimum width and electron gas density). In weak magnetic fields such that $2l_{cycl} > W$, the number of occupied subbands remains approximately constant (*cf.* Fig. 10 or Eq. (8)), which is the reason for the weak dependence on B of the two-terminal resistance in this field regime. For stronger fields, Eq. (9) describes a *positive* magnetoresistance, because $N_{\min}$ decreases due to the magnetic depopulation of subbands discussed in Section 3. Why then do we find a *negative* magnetoresistance in the four-terminal measurements of Fig. 15? Qualitatively, the answer is shown in Fig. 16 for a constriction without a potential barrier. In a magnetic field the left- and right-moving electrons are separated spatially by the

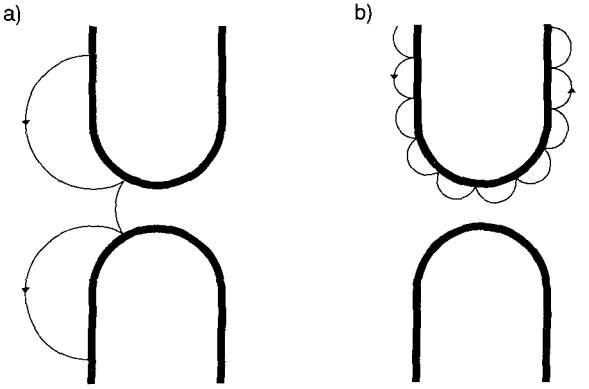

FIG. 16. Illustration of the reduction of backscattering by a magnetic field, which is responsible for the negative magnetoresistance of Fig. 15. Shown are trajectories approaching a constriction without a potential barrier, in a weak (*a*) and strong (*b*) magnetic field.

Lorentz force at opposite sides of the constriction. Quantum mechanically, the skipping orbits in Fig. 16 correspond to magnetic edge states (*cf.* Fig. 11). Backscattering thus requires scattering across the width of the constriction, which becomes increasingly improbable as l_{cycl} becomes smaller and smaller compared to the width. (Compare Figs. 16a, b.) For this reason, in a magnetic field the role of the constriction as the dominant bottleneck limiting the current is taken over increasingly by the contact resistance at the connection of the current contacts with the 2DEG. Quantitatively, this can be treated as follows.[34]

Consider the four-terminal geometry of Fig. 17. A current I flows through a constriction due to a chemical potential difference between the source (at chemical potential $\mu_s = E_F + \delta\mu$) and the drain (at chemical potential $\mu_d = E_F$). Unless the magnetic field is very weak, we can assume that the left- and right-moving electrons are separated spatially at the lower and upper boundary of the wide 2DEG, and that backscattering can occur only at the constriction. The four-terminal longitudinal resistance is defined as $R_L \equiv (\mu_l - \mu_r)/eI$, where μ_l and μ_r are the chemical potentials measured by the two voltage probes shown in Fig. 17, at the upper boundary to the left and right of the constriction. The left voltage probe, which is in equilibrium with the electrons coming from the source, has $\mu_l = E_F + \delta\mu$. A fraction N_{min}/N_{wide} of these electrons is transmitted through the constriction, the remainder returning to the source contact via the opposite edge. Here, N_{min} and N_{wide} are the number of propagating modes in the narrow and wide regions, respectively. If we assume a local equilibrium near the voltage probes, the excess chemical potential, $\mu_r - E_F$, is reduced by the same factor, $\mu_r - E_F = (N_{min}/N_{wide})(\mu_l - E_F)$. (In the absence of local equilibrium, the measured chemical potential depends on how the voltage probe couples to the 2DEG.) The transmitted current itself is determined by the two-terminal resistance, from Eq. (9), which gives $I = (2e/h)N_{min}\,\delta\mu$. Collecting results, we find the

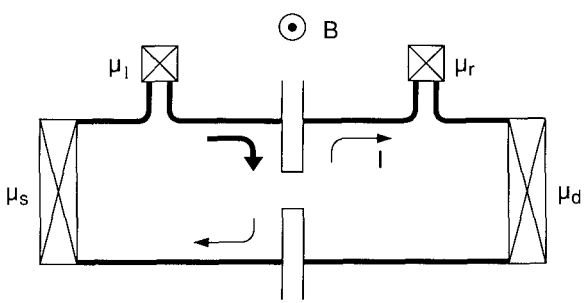

FIG 17 Schematic arrangement of a four-terminal conductor containing a constriction, used in the text to derive Eq (10) The spatial separation of the left- and right-moving electrons in the wide regions is indicated (*cf* Fig 11a)

simple formula,[34]

$$R_L = \frac{h}{2e^2}\left(\frac{1}{N_{min}} - \frac{1}{N_{wide}}\right), \tag{10}$$

obtained independently by Büttiker.[77] For a more formal derivation of this result using the Landauer–Büttiker formalism, see Section 9.

At small magnetic fields, N_{min} is approximately constant, while N_{wide} decreases linearly with B (*cf.* Eq. (8)). Equation (10) thus predicts a *negative* magnetoresistance. Physically, the resistance reduction is due to the fact that, as B is increased, a larger and larger fraction of the edge states is transmitted through the constriction (as is illustrated in Fig. 16). If the electron density in the wide and narrow regions is equal (i.e., the barrier height $E_c = 0$), then the resistance R_L *vanishes* for fields $B > B_{crit} \equiv 2\hbar k_F/eW$. This follows from Eq. (10), because in this case, N_{min} and N_{wide} are identical. If, on the other hand, the electron density in the constriction is less than its value in the wide region, then Eq. (10) predicts a crossover at B_{crit} to a strong-field regime of *positive* magnetoresistance described by

$$R_L \approx \frac{h}{2e^2}\left(\frac{\hbar\omega_c}{E_F - E_c} - \frac{\hbar\omega_c}{E_F}\right), \text{ if } B > B_{crit}. \tag{11}$$

The solid curves in Fig. 15 have been obtained from Eqs. (8) and (10) (after addition of the background resistance found at gate voltage zero), with the constriction width W_{min} as the single freeparameter. The barrier height E_c has been determined independently from the two-terminal resistance in high magnetic fields (*cf.* Section 3). The agreement found is quite good, confirming the validity of Eq. (10) in the weak-field regime, and providing a means to determine the constriction width (which is found to vary from 0.8 to 0.3 μm as V_g varies from -0.3 to -3.0 V; see the curves labeled 1 in Fig. 14). The constriction in the present experiment is relatively long ($L \approx 3.4$ μm), so that it does not exhibit clear quantized plateaus in the zero-field two-terminal conductance (*cf.* Section 2.d and Fig. 9, measured on this same sample). For this reason, the discreteness of N was ignored in the theoretical curves in Fig. 15. We emphasize, however, that the preceding analysis is equally applicable to the quantum case (as will be discussed in Section 9). For example, Eq. (10) describes the quantization in *fractions* of $2e^2/h$ of the longitudinal conductance R_L^{-1} of a point contact observed experimentally.[125,126] (See Section 9.a).

In high magnetic fields in the quantum Hall effect regime, the validity of the result, Eq. (10), is not restricted to point contacts, but holds also for a wide Hall bar (having N_{wide} occupied Landau levels), of which a segment has

a reduced electron density (so that only $N_{\min}$ Landau levels are occupied in that region). Many such experiments have been performed.[127–132] In these papers, the simplicity of an analysis in terms of transmitted and reflected edge states had not been appreciated yet, in contrast to more recent experimental work[133–135] in which a narrow gate across the Hall bar induces a potential barrier in the 2DEG. Deviations from Eq. (10) can result from inter-edge channel scattering; *cf.* Refs. 136, 137, and 138.

The preceding argument predicts a Hall resistance, $R_H = R_{2t} - R_L$, in the wide regions given by

$$R_H = \frac{h}{2e^2}\frac{1}{N_{\text{wide}}}, \tag{12}$$

unaffected by the presence of the constriction. This is a direct consequence of our assumption of local equilibrium near the voltage probes. In the weak-field regime of Fig. 15, this result, Eq. (12), has been confirmed experimentally, but deviations were found for higher magnetic fields.[34] These are discussed further in Section 9. Anticipating that discussion, we note that in a strong magnetic field, the assumption of local equilibrium near the voltage probes is a sufficient but not necessary condition for Eqs. (10) and (12) to hold. Even in the absence of local equilibrium, these equations remain valid if the voltage probes are much wider than l_{cycl}, so that all edge states on one edge of the wide 2DEG region are fully absorbed by the voltage contact. Such an *ideal* contact induces itself a local equilibrium among the edge states[77] (*cf.* Section 9.a).

For completeness, we mention that one also can measure the two four-terminal *diagonal* resistances R_{D^+} and R_{D^-} across the constriction, in such a way that the two voltage probes are on opposite (not adjacent) edges of the 2DEG, on either side of the constriction. (See Fig. 18). Additivity of voltages on contacts tells us that $R_{D^\pm} = R_L \pm R_H$ (for the magnetic field direction of Fig. 18), so that

$$R_{D^+} = \frac{h}{2e^2}\frac{1}{N_{\min}}; \quad R_{D^-} = \frac{h}{2e^2}\left(\frac{2}{N_{\text{wide}}} - \frac{1}{N_{\min}}\right). \tag{13}$$

On field reversal, R_{D^+} and R_{D^-} are interchanged. Thus, a four-terminal resistance (R_{D^+} in Eq. (13)) can be equal in principle to the two-terminal resistance (R_{2t} in Eq. (9)). The main difference between these two quantities is that the additive contribution of the ohmic contact resistance, and of a part of the diffusive background resistance, is eliminated in the four-terminal resistance measurement (*cf.* Part II).

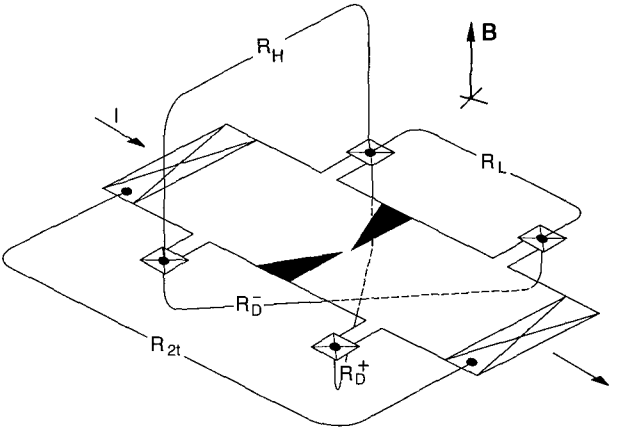

FIG 18 Perspective view of a six-terminal Hall bar containing a point contact, showing the various two- and four-terminal resistances mentioned in the text.

In conclusion, we have demonstrated that a magnetic field suppresses the backscattering at the point contact that results from its finite width. In Ref. 34, it was suggested that a magnetic field also may suppress the backscattering caused by impurities in a narrow channel, thereby explaining a negative magnetoresistance effect in the quasi-ballistic transport regime first observed by Choi *et al.*[50] (See also Ref. 109.) A theory of the quantum Hall effect based on the suppression of backscattering by a magnetic field has been developed by Büttiker;[77] see Part IV.

5. ELECTRON BEAM COLLIMATION AND POINT CONTACTS IN SERIES

a. *Introduction*

The first experimental study of ballistic transport through two opposite point contacts was carried out by Wharam *et al.*,[9] who discovered that the series resistance is considerably less than the sum of the two individual resistances. Subsequent experiments confirmed this result.[139,151] To explain this observation theoretically, two of us proposed[8,66] that *collimation* of the electron beam injected by a point contact enhances the direct transmission probability from one point contact to the other, thereby significantly reducing the series resistance below its ohmic value. An alternative measurement configuration was suggested, in which the deflection of the beam by a magnetic field can be sensitively detected, to provide direct experimental proof of

a reduced electron density (so that only $N_{\min}$ Landau levels are occupied in that region). Many such experiments have been performed.[127–132] In these papers, the simplicity of an analysis in terms of transmitted and reflected edge states had not been appreciated yet, in contrast to more recent experimental work[133–135] in which a narrow gate across the Hall bar induces a potential barrier in the 2DEG. Deviations from Eq. (10) can result from inter-edge channel scattering; *cf.* Refs. 136, 137, and 138.

The preceding argument predicts a Hall resistance, $R_H = R_{2t} - R_L$, in the wide regions given by

$$R_H = \frac{h}{2e^2}\frac{1}{N_{\text{wide}}}, \tag{12}$$

unaffected by the presence of the constriction. This is a direct consequence of our assumption of local equilibrium near the voltage probes. In the weak-field regime of Fig. 15, this result, Eq. (12), has been confirmed experimentally, but deviations were found for higher magnetic fields.[34] These are discussed further in Section 9. Anticipating that discussion, we note that in a strong magnetic field, the assumption of local equilibrium near the voltage probes is a sufficient but not necessary condition for Eqs. (10) and (12) to hold. Even in the absence of local equilibrium, these equations remain valid if the voltage probes are much wider than l_{cycl}, so that all edge states on one edge of the wide 2DEG region are fully absorbed by the voltage contact. Such an *ideal* contact induces itself a local equilibrium among the edge states[77] (*cf.* Section 9.a).

For completeness, we mention that one also can measure the two four-terminal *diagonal* resistances R_{D^+} and R_{D^-} across the constriction, in such a way that the two voltage probes are on opposite (not adjacent) edges of the 2DEG, on either side of the constriction. (See Fig. 18). Additivity of voltages on contacts tells us that $R_{D^\pm} = R_L \pm R_H$ (for the magnetic field direction of Fig. 18), so that

$$R_{D^+} = \frac{h}{2e^2}\frac{1}{N_{\min}}; \quad R_{D^-} = \frac{h}{2e^2}\left(\frac{2}{N_{\text{wide}}} - \frac{1}{N_{\min}}\right). \tag{13}$$

On field reversal, R_{D^+} and R_{D^-} are interchanged. Thus, a four-terminal resistance (R_{D^+} in Eq. (13)) can be equal in principle to the two-terminal resistance (R_{2t} in Eq. (9)). The main difference between these two quantities is that the additive contribution of the ohmic contact resistance, and of a part of the diffusive background resistance, is eliminated in the four-terminal resistance measurement (*cf.* Part II).

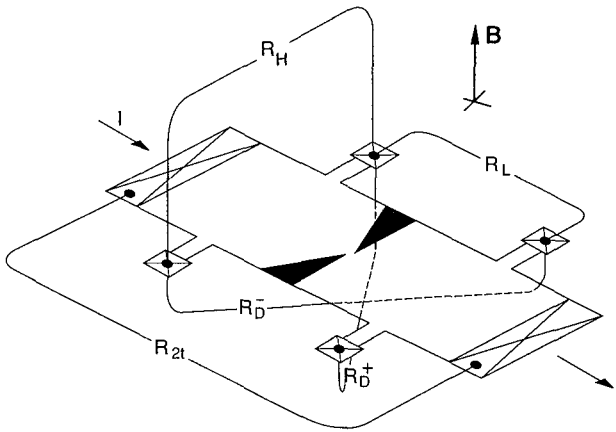

FIG. 18. Perspective view of a six-terminal Hall bar containing a point contact, showing the various two- and four-terminal resistances mentioned in the text.

In conclusion, we have demonstrated that a magnetic field suppresses the backscattering at the point contact that results from its finite width. In Ref. 34, it was suggested that a magnetic field also may suppress the backscattering caused by impurities in a narrow channel, thereby explaining a negative magnetoresistance effect in the quasi-ballistic transport regime first observed by Choi *et al.*[50] (See also Ref. 109.) A theory of the quantum Hall effect based on the suppression of backscattering by a magnetic field has been developed by Büttiker;[77] see Part IV.

5. ELECTRON BEAM COLLIMATION AND POINT CONTACTS IN SERIES

a. *Introduction*

The first experimental study of ballistic transport through two opposite point contacts was carried out by Wharam *et al.*,[9] who discovered that the series resistance is considerably less than the sum of the two individual resistances. Subsequent experiments confirmed this result.[139,151] To explain this observation theoretically, two of us proposed[8,66] that *collimation* of the electron beam injected by a point contact enhances the direct transmission probability from one point contact to the other, thereby significantly reducing the series resistance below its ohmic value. An alternative measurement configuration was suggested, in which the deflection of the beam by a magnetic field can be sensitively detected, to provide direct experimental proof of

collimation. Such an experiment now has been performed,[58] and will be discussed shortly in some detail. We will not consider here the obvious alternative geometry of two adjacent point contacts in parallel (studied in Refs. 140, 141, and 142). In that geometry, the collimation effect can not enhance the coupling of the two point contacts, so only small deviations from Ohm's law are to be expected.

The collimation effect has an importance in ballistic transport that goes beyond the point contact geometry for which it originally was proposed. Recent theoretical[143,144] and experimental[145,146] work has shown that collimation is at the origin of the phenomenon of the quenching of the Hall effect (a suppression of the Hall resistance at low magnetic fields).[52] In fact, collimation is one of a set of semiclassical mechanisms[144] that together can explain a whole variety of magnetoresistance anomalies found experimentally in ballistic narrow-channel geometries, including quenched and negative Hall resistances, the last Hall plateau, bend resistances, and geometrical resonances. These recent developments emphasize the general importance of the collimation effect, but will not be discussed here any further. In this chapter, we restrict ourselves to point contact geometry, which happens to be an ideal geometry for the study of collimated electron beams.

b. Collimation

Collimation follows from the constraints on the electron momentum imposed by the potential energy barrier in the point contact (barrier collimation), and by the gradual flaring of the confining potential at the entrance and exit of the point contact (horn collimation).[8,66] Semiclassically, collimation results from the adiabatic invariance of the product of channel width W and absolute value of the transverse momentum $\hbar k_y$. (This product is proportional to the action for motion transverse to the channel.)[147] Therefore, if the electrostatic potential in the point contact region is sufficiently smooth, the quantity $S = |k_y|W$ is approximately constant from point contact entrance to exit. Note that S/π corresponds to the quantum mechanical one-dimensional subband index n. The quantum mechanical criterion for adiabatic transport thus is that the potential in the point contact region does not cause intersubband transitions. To this end, it should be smooth on the scale of a wavelength, and the width should change gradually on the same length scale. As we discussed in Section 2.b.iii, adiabatic transport breaks down at the exit of the point contact, where it widens abruptly into a 2DEG of essentially infinite width. Barrier- and horn-collimation reduce the *injection/acceptance cone* of the point contact from its original value of π to a value of $2\alpha_{\max}$. This effect is illustrated in Fig. 19. Electrons incident at an angle $|\alpha| > \alpha_{\max}$ from normal incidence are reflected. (The geometry

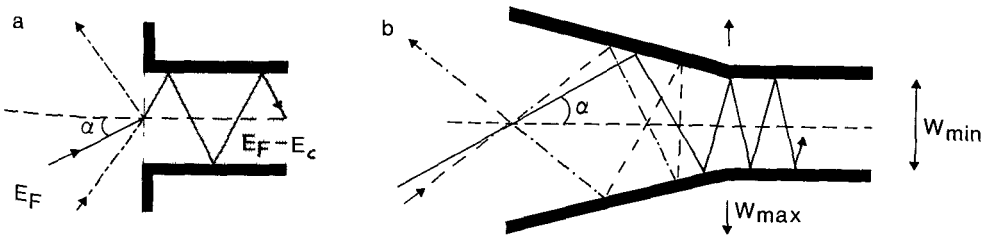

FIG. 19. Illustration of the collimation effect for an abrupt constriction (*a*) containing a potential barrier of height E_c, and for a horn-shaped constriction (*b*) that is flared from a width W_{min} to W_{max}. The dash-dotted trajectories approaching at an angle α outside the injection/acceptance cone are reflected. (From Ref. 66.)

of Fig. 19b is known in optics as a *conical reflector*.)[148] On the other hand, all electrons leave the constriction at an angle $|\alpha| < \alpha_{max}$; i.e., the injected electrons form a collimated beam of angular opening $2\alpha_{max}$.

To obtain an analytic expression for the collimation effect, we describe the shape of the potential in the point contact region by three parameters: W_{min}, W_{max}, and E_c. (See Fig. 19.) We consider the case that the point contact has its minimal width W_{min} at the point where the barrier has its maximal height E_c above the bottom of the conduction band in the broad regions. At that point, the largest possible value of S is

$$S_1 \equiv (2m/\hbar^2)^{1/2}(E_F - E_c)^{1/2} W_{min}.$$

We assume that adiabatic transport (i.e., S = constant) holds up to a point of zero barrier height and maximal width W_{max}. The abrupt separation of adiabatic and non-adiabatic regions is a simplification that can be—and has been—tested by numerical calculations. (See the following.) At the point contact exit, the largest possible value of S is

$$S_2 \equiv (2m/\hbar^2)^{1/2}(E_F)^{1/2} \sin\alpha_{max} W_{max}.$$

The invariance of S implies that $S_1 = S_2$, so that

$$\alpha_{max} = \arcsin\left(\frac{1}{f}\right); \quad f \equiv \left(\frac{E_F}{E_F - E_c}\right)^{1/2} \frac{W_{max}}{W_{min}}. \tag{14}$$

The *collimation factor* $f \geq 1$ is the product of a term describing the collimating effect of a barrier of height E_c, and a term describing collimation due to a gradual widening of the point contact width from W_{min} to W_{max}. In the adiabatic approximation, the angular injection distribution $P(\alpha)$ is proportional to $\cos\alpha$ with an abrupt truncation at $\pm\alpha_{max}$. The cosine angular

dependence follows from the cosine distribution of the incident flux in combination with time reversal symmetry, and thus is not affected by the reduction of the *injection/acceptance* cone. (This also can be seen from the quantum mechanical correspondence, by noting that in the absence of inter subband transitions, the relative magnitude of the contributions of the transmitted one-dimensional subbands to the injected current cannot change.) We conclude, therefore, that in the adiabatic approximation, $P(\alpha)$ (normalized to unity) is given by

$$P(\alpha) = \frac{1}{2} f \cos\alpha, \text{ if } |\alpha| < \alpha_{\max} \equiv \arcsin(1/f),$$
$$P(\alpha) = 0, \text{ otherwise.} \tag{15}$$

We defer to Section 5.d a comparison of the analytical result, Eq. (15), with a numerical calculation.

The injection distribution, Eq. (15), can be used to obtain (in the semiclassical limit) the direct transmission probability T_d through two identical opposite point contacts separated by a large distance L. To this end, first note that T_d/N is the fraction of the current injected through the first point contact, which is transmitted through the second point contact (since the transmission probability through the first point contact is N, for N occupied subbands in the point contact). Electrons injected within a cone of opening angle $W_{\max}/L$ centered at $\alpha = 0$ reach the opposite point contact, and are transmitted. If this opening angle is much smaller than the total opening angle $2\alpha_{\max}$ of the beam, then the distribution function $P(\alpha)$ can be approximated by $P(0)$ within this cone. This approximation requires $W_{\max}/L \ll 1/f$, which is satisfied experimentally in devices with a sufficiently large point contact separation. We thus obtain $T_d/N = P(0)W_{\max}/L$, which, using Eq. (15), can be written as[8]

$$T_d = f \frac{W_{\max}}{2L} N. \tag{16}$$

This simple analytical formula can be used to describe the experiments on transport through identical opposite point contacts in terms of one empirical parameter f, as discussed in the following two subsections.

c. *Series Resistance*

The expression for the series resistance of two identical opposite point contacts in terms of the preceding transmission probability can be obtained directly from the Landauer–Büttiker formalism,[7] as was done in Ref. 8.

We give here an equivalent, somewhat more intuitive derivation. Consider the geometry shown in Fig. 20a. A fraction T_d/N of the current GV injected through the first point contact by the current source is transmitted directly through the second point contact (and then drained to ground). Here, $G = (2e^2/h)N$ is the conductance of the individual point contacts, and V is the source-drain voltage. The remaining fraction, $1 - T_d/N$, equilibrates in the region between the point contacts, as a result of inelastic scattering. (Elastic scattering is sufficient if phase coherence does not play a role.) Since that region cannot drain charge (as contacts 2 and 4 are not connected to ground in Fig. 20a), these electrons eventually will leave via one of the two point contacts. For a symmetric structure, we may assume that the fraction $\frac{1}{2}(1 - T_d/N)$ of the injected current GV is transmitted through the second point contact after equilibration. The total source-drain current I is the sum of the direct and indirect contributions,

$$I = \frac{1}{2}\left(1 + \frac{T_d}{N}\right)GV.$$

The series conductance $G_{series} = I/V$ becomes

$$G_{series} = \frac{1}{2}G\left(1 + \frac{T_d}{N}\right). \tag{17}$$

In the absence of direct transmission ($T_d = 0$), one recovers the ohmic addition law for the resistance, as expected for the case of complete intervening

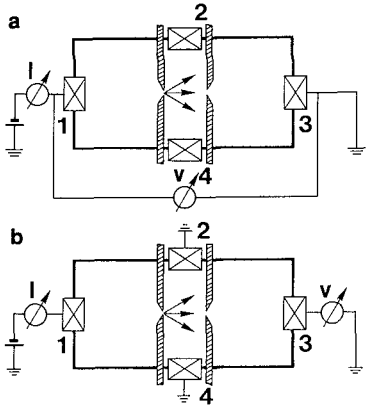

FIG. 20. Configuration for a series resistance measurement (*a*) and for a measurement of the collimation peak (*b*). The gates defining the two opposite point contacts are shaded; the squares indicate ohmic contacts to the 2DEG. (From Ref. 66.)

equilibration (*cf.* the related analysis by Büttiker of tunneling in series barriers).[149,150] At the opposite extreme, if all transmission is direct ($T_d = N$), the series conductance is identical to that of the single-point contact. Substituting the expression, Eq. (16), into Eq. (17), we obtain the result[8] for small but nonzero direct transmission,

$$G_{\text{series}} = \frac{1}{2} G\left(1 + f\frac{W_{\max}}{2L}\right). \tag{18}$$

The plateaus in the series resistance as a function of gate voltage, observed experimentally,[9] of course are not obtained in the semiclassical calculation leading to Eq. (18). However, since the non-additivity essentially is a semiclassical collimation effect, the present analysis should give a reasonably reliable estimate of deviations from additivity for not too narrow point contacts. Once the point contact width becomes less than a wavelength, diffraction inhibits collimation of the electron beam. In the limit $k_F W \ll 1$, the injection distribution becomes proportional to $\cos^2\alpha$ for all α, independent of the shape of the potential in the point contact region.[15] Here, we will compare Eq. (18) only with experiments on rather wide point contacts. A fully quantum mechanical calculation of the series conductance, with which the experiments and the semiclassical result could be compared, unfortunately is not available. (The calculation of Ref. 89 can not be used for this purpose because equilibration in the region between the point contacts due to inelastic scattering is ignored.)

Wharam *et al.*[9] find for relatively large point contact widths ($G = 5 \times (2e^2/h)$ for both point contacts) a ratio $G/G_{\text{series}} = 1.4$, considerably less than the ratio of 2 expected from ohmic addition. From Eq. (17), we infer that a fraction $T_d/N \approx 0.4$ of the injected current is transmitted directly. As discussed in Ref. 8, this number is consistent with Eq. (16) using an estimate of $f \approx 2.4$ and $W_{\max}/L \approx 0.4$ in their geometry.

In a recent study, Beton *et al.*[151] have applied the preceding results to their experiment on transport through point contacts in series. For the two widest identical point contacts considered ($G = 4 \times (2e^2/h)$), they infer a fraction $T_d/N = 0.6$ of directly transmitted current. From their scale diagram, we estimate $L \approx 440$ nm; From the relation,

$$G = \left(\frac{2e^2}{h}\right)\left(\frac{2m}{\hbar^2}\right)^{1/2}(E_F - E_c)^{1/2}\frac{W_{\min}}{\pi}$$

(with $E_c \approx 0$ and the experimentally given value of E_F), we estimate $W_{\min} \approx 90$ nm. The maximal width for adiabatic transport $W_{\max}$ is difficult to estimate reliably, and the lithographic opening of 240 nm in the gates defining

the point contact presumably is an overestimate. Using $E_c \approx 0$ and $W_{max} \lesssim 240$ nm, we find from Eq. (14) a collimation factor, $f \lesssim 2.7$, which from Eq. (16) gives the theoretical value, $T_d/N \lesssim 0.7$—consistent with the experimental value[151] of 0.6. (In Ref. 151, a much larger theoretical value is stated without derivation.) At smaller point contact widths, the agreement between experiment and theory becomes worse, possibly as a result of the diffraction effects mentioned earlier.

In both these experiments (as well as in a similar experiment of Hirayama and Saku)[139] L is not much larger than W_{max}, so that the requirement for the validity of Eqs. (16) and (18) of small fW_{max}/L (or, equivalently, small T_d/N) is not well satisfied. A more significant comparison between the present analytical theory and experiment would require a larger point contact separation. Unfortunately, the non-additivity of the resistance then is only a small correction to the series resistance (since $T_d \ll N$). For a more sensitive study of the collimated electron beam, one needs to eliminate the background signal from electrons transmitted after equilibration, which obscures the direct transmission in a series resistance measurement at large point contact separation. As proposed in Refs. 8 and 66 and discussed in Sec. 5.d, this uninteresting background can be largely eliminated by maintaining the region between the point contacts at ground potential and operating one of the point contacts as a voltage probe drawing no net current. (See Fig. 20b.)

So far, we have considered only the case of zero magnetic field. In a weak magnetic field ($2l_{cycl} > L$), the situation is rather complicated. As discussed in detail in Ref. 8, there are two competing effects in weak fields: On the one hand, the deflection of the electron beam by the Lorentz force reduces the direct transmission probability, with the effect of decreasing the series conductance; on the other hand, the magnetic field enhances the indirect transmission, with the opposite effect. The result is an initial *decrease* in the series conductance for small magnetic fields in the case of strong collimation, and an *increase* in the case of weak collimation. This is expected to be a relatively small effect compared to the effects at stronger fields discussed next (*cf.* Ref. 8).

In stronger fields ($2l_{cycl} < L$), the direct transmission probability vanishes, which greatly simplifies the situation. If we assume that all transmission between the opposite point contacts is with intervening equilibration, then the result is[8]

$$G_{series} = \frac{2e^2}{h}\left(\frac{2}{N} - \frac{1}{N_{wide}}\right)^{-1/2}. \tag{19}$$

Here, N is the (B-independent) number of occupied subbands in the point contacts, and N_{wide} is the number of occupied Landau levels in the 2DEG

equilibration (*cf.* the related analysis by Büttiker of tunneling in series barriers).[149,150] At the opposite extreme, if all transmission is direct ($T_d = N$), the series conductance is identical to that of the single-point contact. Substituting the expression, Eq. (16), into Eq. (17), we obtain the result[8] for small but nonzero direct transmission,

$$G_{\text{series}} = \frac{1}{2} G\left(1 + f\frac{W_{\max}}{2L}\right). \tag{18}$$

The plateaus in the series resistance as a function of gate voltage, observed experimentally,[9] of course are not obtained in the semiclassical calculation leading to Eq. (18). However, since the non-additivity essentially is a semiclassical collimation effect, the present analysis should give a reasonably reliable estimate of deviations from additivity for not too narrow point contacts. Once the point contact width becomes less than a wavelength, diffraction inhibits collimation of the electron beam. In the limit $k_F W \ll 1$, the injection distribution becomes proportional to $\cos^2\alpha$ for all α, independent of the shape of the potential in the point contact region.[15] Here, we will compare Eq. (18) only with experiments on rather wide point contacts. A fully quantum mechanical calculation of the series conductance, with which the experiments and the semiclassical result could be compared, unfortunately is not available. (The calculation of Ref. 89 can not be used for this purpose because equilibration in the region between the point contacts due to inelastic scattering is ignored.)

Wharam *et al.*[9] find for relatively large point contact widths ($G = 5 \times (2e^2/h)$ for both point contacts) a ratio $G/G_{\text{series}} = 1.4$, considerably less than the ratio of 2 expected from ohmic addition. From Eq. (17), we infer that a fraction $T_d/N \approx 0.4$ of the injected current is transmitted directly. As discussed in Ref. 8, this number is consistent with Eq. (16) using an estimate of $f \approx 2.4$ and $W_{\max}/L \approx 0.4$ in their geometry.

In a recent study, Beton *et al.*[151] have applied the preceding results to their experiment on transport through point contacts in series. For the two widest identical point contacts considered ($G = 4 \times (2e^2/h)$), they infer a fraction $T_d/N = 0.6$ of directly transmitted current. From their scale diagram, we estimate $L \approx 440$ nm; From the relation,

$$G = \left(\frac{2e^2}{h}\right)\left(\frac{2m}{\hbar^2}\right)^{1/2}(E_F - E_c)^{1/2}\frac{W_{\min}}{\pi}$$

(with $E_c \approx 0$ and the experimentally given value of E_F), we estimate $W_{\min} \approx 90$ nm. The maximal width for adiabatic transport $W_{\max}$ is difficult to estimate reliably, and the lithographic opening of 240 nm in the gates defining

the point contact presumably is an overestimate. Using $E_c \approx 0$ and $W_{max} \lesssim 240$ nm, we find from Eq. (14) a collimation factor, $f \lesssim 2.7$, which from Eq. (16) gives the theoretical value, $T_d/N \lesssim 0.7$—consistent with the experimental value[151] of 0.6. (In Ref. 151, a much larger theoretical value is stated without derivation.) At smaller point contact widths, the agreement between experiment and theory becomes worse, possibly as a result of the diffraction effects mentioned earlier.

In both these experiments (as well as in a similar experiment of Hirayama and Saku)[139] L is not much larger than W_{max}, so that the requirement for the validity of Eqs. (16) and (18) of small fW_{max}/L (or, equivalently, small T_d/N) is not well satisfied. A more significant comparison between the present analytical theory and experiment would require a larger point contact separation. Unfortunately, the non-additivity of the resistance then is only a small correction to the series resistance (since $T_d \ll N$). For a more sensitive study of the collimated electron beam, one needs to eliminate the background signal from electrons transmitted after equilibration, which obscures the direct transmission in a series resistance measurement at large point contact separation. As proposed in Refs. 8 and 66 and discussed in Sec. 5.d, this uninteresting background can be largely eliminated by maintaining the region between the point contacts at ground potential and operating one of the point contacts as a voltage probe drawing no net current. (See Fig. 20b.)

So far, we have considered only the case of zero magnetic field. In a weak magnetic field ($2l_{cycl} > L$), the situation is rather complicated. As discussed in detail in Ref. 8, there are two competing effects in weak fields: On the one hand, the deflection of the electron beam by the Lorentz force reduces the direct transmission probability, with the effect of decreasing the series conductance; on the other hand, the magnetic field enhances the indirect transmission, with the opposite effect. The result is an initial *decrease* in the series conductance for small magnetic fields in the case of strong collimation, and an *increase* in the case of weak collimation. This is expected to be a relatively small effect compared to the effects at stronger fields discussed next (*cf.* Ref. 8).

In stronger fields ($2l_{cycl} < L$), the direct transmission probability vanishes, which greatly simplifies the situation. If we assume that all transmission between the opposite point contacts is with intervening equilibration, then the result is[8]

$$G_{series} = \frac{2e^2}{h}\left(\frac{2}{N} - \frac{1}{N_{wide}}\right)^{-1/2}. \tag{19}$$

Here, N is the (B-independent) number of occupied subbands in the point contacts, and N_{wide} is the number of occupied Landau levels in the 2DEG

between the point contacts. The physical origin of the simple addition rule, Eq. (19), is additivity of the four-terminal longitudinal resistance as in Eq. (10). From this additivity, it follows that for n different point contacts in series, Eq. (19) generalizes to

$$\frac{1}{G_{\text{series}}} - \frac{h}{2e^2}\frac{1}{N_{\text{wide}}} = \sum_{i=1}^{n} R_{\text{L}}(i), \tag{20}$$

where

$$R_{\text{L}}(i) = \left(\frac{h}{2e^2}\right)\left(\frac{1}{N_i} - \frac{1}{N_{\text{wide}}}\right)$$

is the four-terminal longitudinal resistance of point contact i. Equation (19) predicts a non-monotonic B-dependence for G_{series}. This can be seen most easily by disregarding the discreteness of N and N_{wide}. We then have $N_{\text{L}} \approx E_{\text{F}}/\hbar\omega_{\text{c}}$, while the magnetic field dependence of N (for a square-well confining potential in the point contacts) is given by Eq. (8). The resulting B-dependence of G_{series} is shown in Fig. 21 (dotted curves). The non-monotonic behavior is due to the delayed depopulation of subbands in the point contacts, compared to the broad 2DEG. While the number of occupied Landau levels N_{wide} in the region between the point contacts decreases steadily with B for $2l_{\text{cycl}} < L$, the number N of occupied subbands in the point contacts remains approximately constant until $2l_{\text{c,min}} \approx W_{\text{min}}$ (with $l_{\text{c,min}} \equiv l_{\text{cycl}}\,(1 - E_{\text{c}}/E_{\text{F}})^{1/2}$ denot-

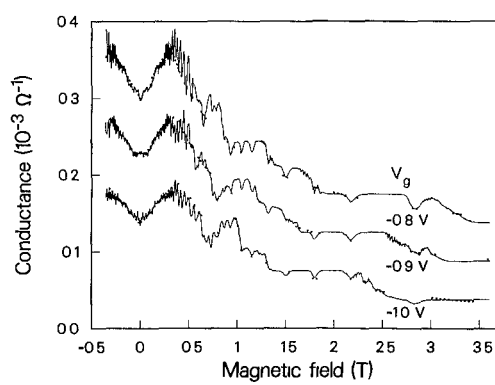

FIG 21 Magnetic field dependence of the series conductance of two opposite point contacts ($L = 1.0$ μm) for three different values of the gate voltage (solid curves) at $T = 100$ mK. For clarity, subsequent curves from bottom to top are offset by $0.5 \times 10^{-4}\,\Omega^{-1}$, with the lowest curve shown at its actual value The dotted curves are calculated from Eqs (8) and (19), with the point contact width as adjustable parameter (From Ref. 59)

ing the cyclotron radius in the point contact region). In this field interval, G_{series} increases with B, according to Eq. (19). For stronger fields, depopulation in the point contacts begins to dominate G_{series}, leading finally to a decreasing conductance (as is the rule for single point contacts; see Section 3). The peak in G_{series} thus occurs at $2l_{c,\min} \approx W_{\min}$.

The remarkable camelback shape of G_{series} versus B predicted by Eq. (19) now has been observed experimentally.[59] The data is shown in Fig. 21 (solid curves) for three values of the gate voltage V_g, at $T = 100$ mK. The measurement configuration is as shown in Fig. 20a, with a point contact separation $L = 1.0$ μm. The voltage is the same on all the gates defining the point contacts, so that we expect the two opposite point contacts to be similar. Moreover, since $|V_g|$ is rather small, we can neglect the barrier in the point contact and assume that the electron density in the point contacts is the same as in the 2DEG channel that separates the point contacts. (The latter density can be obtained independently from the Hall resistance of the channel, and is found to decrease from 1.1 to 0.8×10^{15} m^{-2} as V_g varies from -0.8 to -1.0 V.) The point contact width $W_{\min}$ then remains the only free parameter, which is determined by a fit using Eqs. (8) and (19). (A square-well confining potential is assumed in the point contacts; $W_{\min}$ is found to decrease from 320 to 200 nm over the gate voltage range of Fig. 21). The dotted curves in Fig. 21 show the result of such a fit[59] (after correction for a constant background resistance of 2.0 kΩ, estimated from the two-terminal measurement of the quantized resistance of the individual point contacts). It is seen that Eq. (19) provides a good description of the overall magnetoresistance behavior from low magnetic fields up to the quantum Hall effect regime. The additional structure in the experimental curves has several different origins, for which we refer to Ref. 59. Similar structure in the two-terminal resistance of a single point contact will be discussed in detail in Section 11.

We emphasize that Eq. (19) is based on the assumption of complete equilibration of the current-carrying edge states in the region between the point contacts. In a quantizing magnetic field, local equilibrium is reached by inter-Landau-level scattering. If the potential landscape (both in the point contacts themselves and in the 2DEG region in between) varies by less than the Landau-level separation $\hbar\omega_c$ on the length scale of the magnetic length $(\hbar/eB)^{1/2}$, then inter-Landau-level scattering is suppressed in the absence of other scattering mechanisms. (See Part IV.) This means that the transport from one point contact to the other is adiabatic. The series conductance then simply is $G_{\text{series}} = (2e^2/h)N$ for two identical point contacts, where $N \equiv \min(N_1, N_2)$ for two different point contacts in series. This expression differs from Eq. (19) if a barrier is present in the point contacts, since that causes the number N of occupied Landau levels in the point contact to be less than the number N_{wide} of occupied levels in the wide 2DEG. (In a strong magnetic

field, $N \approx (E_F - E_c)/\hbar\omega_c$, while $N_{wide} \approx E_F/\hbar\omega_c$.) Adiabatic transport in a magnetic field through two point contacts in series has been studied experimentally in Ref. 152.

d. Magnetic Deflection of a Collimated Electron Beam

Consider the geometry of Fig. 20b. The current I_i through the injecting point contact is drained to ground at the two ends of the 2DEG channel separating the point contacts. The opposite point contact, the collector, serves as a voltage probe (with the voltage V_c being measured relative to ground). Since the terminals 2 and 4 are grounded, the indirect transmission probability from injector to collector is suppressed.[8,66] The collector voltage divided by the injected current is given by

$$\frac{V_c}{I_i} = \frac{1}{G}\frac{T_d}{N}, \quad T_d \ll N, \tag{21}$$

with $G = (2e^2/h)N$ the two-terminal conductance of the individual point contacts (which are assumed to be identical), and T_d the direct transmission probability through the two point contacts, as calculated in Section 5b. Equation (21) can be obtained either from the Landauer–Büttiker formalism (as done in Ref. 66), or simply by noting that the incoming current $I_i T_d/N$ through the collector has to be counterbalanced by an equal outgoing current GV_c (since the collector draws no net current). Using Eq. (16), we find

$$\frac{V_c}{I_i} = \frac{h}{2e^2} f^2 \frac{\pi}{2k_F L}, \tag{22}$$

where we have used the relation

$$G = \left(\frac{2e^2}{h}\right)\left(\frac{k_F W_{max}}{\pi}\right)\left(\frac{1}{f}\right)$$

($k_F \equiv (2mE_F/\hbar^2)^{1/2}$ being the Fermi wave vector in the region between the point contacts). In an experimental situation, L and k_F are known, so that the collimation factor f can be determined directly from the collector voltage by means of Eq. (22).

The result Eq. (22) holds in the absence of a magnetic field. A small magnetic field B will deflect the collimated electron beam past the collector. Simple geometry leads to the criterion, $L/2l_{cycl} = \alpha_{max}$, for the cyclotron radius at which T_d is reduced to zero by the Lorentz force. One thus would expect to see in V_c/I_i a peak around zero field, of height given by Eq. (22)

and of width,

$$\Delta B = \frac{4\hbar k_F}{eL} \arcsin \frac{1}{f} \tag{23}$$

according to Eq. (14).

In Fig. 22, this collimation peak is shown[58] (solid curve), at $T = 1.8$ K in a device with an $L = 4.0$ μm separation between injector and collector. The actual measurement configuration differed from Fig. 20b in an inessential way, in that only one end of the region between the point contacts was grounded. The current I_i thus flows from contact 1 to 2, and the voltage V_c is measured between contacts 3 and 4. In the notation of Part II, $V_c/I_i \equiv R_{12,34}$. This four-terminal resistance is referred to in narrow Hall bar geometries as a *bend resistance* measurement.[54,56] One can show,[58] using the Landauer–Büttiker formalism,[7] that the height of the collimation peak still is given by Eq. (22) if one replaces f^2 by $f^2 - 1/2$. The expression (23) for the width is not modified. The experimental result in Fig. 22 shows a peak height of 150 Ω (measured relative to the background resistance at large magnetic fields). Using $L = 4.0$ μm and the value $k_F = 1.1 \times 10^8$ m^{-1} obtained from Hall resistance measurements in the channel between the point contacts, one deduces a collimation factor, $f \approx 1.85$. The corresponding opening angle of the injection/acceptance cone is $2\alpha_{max} \approx 65°$. The calculated value of f would imply a width, $\Delta B \approx 0.04$ T, which is not far from the measured full width at half maximum of 0.03 T.

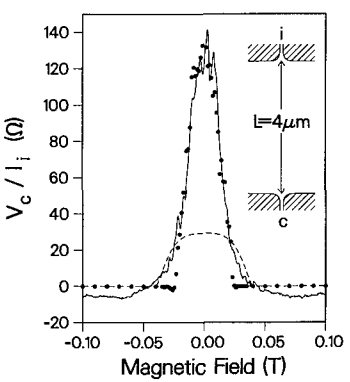

FIG. 22. Detection of a collimated electron beam over a distance of 4 μm. In this four-terminal measurement, two ohmic contacts to the 2DEG region between the point contacts are used: One of these acts as a drain for the current I_i through the injector, and the other is used as a zero-reference for the voltage V_c on the collector. The drawn curve is the experimental data, at $T = 1.8$ K. The black dots are the result of a semiclassical simulation, using a hard-wall potential with contours as shown in the inset. The dashed curve results from a simulation without collimation (corresponding to rectangular corners in the potential contour). (From Ref. 58.)

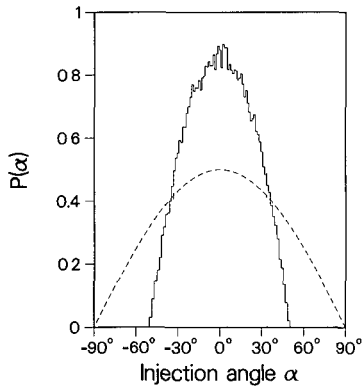

FIG. 23. Calculated angular injection distributions in zero magnetic field. The solid histogram is the result of a simulation of the classical trajectories at the Fermi energy in the geometry shown in the inset of Fig. 22; The dotted curve follows from the adiabatic approximation, Eq. (15), with the experimental collimation factor $f = 1.85$. The dashed curve is the cosine distribution in the absence of any collimation. (From Ref. 58.)

The experimental data in Fig. 22 are compared with the result[58] from a numerical simulation of classical trajectories of the electrons at the Fermi level (following the method of Ref. 144). This semiclassical calculation was performed to relax the assumption of adiabatic transport in the point contact region, and of small T_d/N, on which Eqs. (16) and (21) are based. The dashed curve is for point contacts defined by hard-wall contours with straight corners (no collimation); the dots are for the smooth hard-wall contours shown in the inset, which lead to collimation via the horn effect (*cf.* Fig. 19b; the barrier collimation of Fig. 19a presumably is unimportant at the small gate voltage used in the experiment, and is not taken into account in the numerical simulation). The angular injection distributions $P(\alpha)$ that follow from these numerical simulations are compared in Fig. 23 (solid histogram) with the result, Eq. (15), from the adiabatic approximation for $f = 1.85$ (dotted curve). The uncollimated distribution $P(\alpha) = (\cos\alpha)/2$ also is shown for comparison (dashed curve). Taken together, Figs. 22 and 23 unequivocally demonstrate the importance of collimation for the transport properties, as well as the adequateness of the adiabatic approximation as an estimator of the collimation cone.

6. COHERENT ELECTRON FOCUSING

a. Introduction

Electron focusing in metals was pioneered by Sharvin[23] and Tsoi[24] as a powerful tool to investigate the shape of the Fermi surface, surface scattering, and the electron–phonon interaction.[40] The experiment is the analogue in

the solid state of magnetic focusing in vacuum. Required is a large mean free path for the carriers at the Fermi surface, to ensure *ballistic* motion as in vacuum. The mean free path (which can be as large as 1 cm in pure metallic single crystals) should be much larger than the length L on which the focusing takes place. Experimentally, $L = 10^{-2} - 10^{-1}$ cm is the separation of two metallic needles or point contacts (of typical width $W \sim 1$ μm) pressed on the crystal surface, which serve to inject a divergent electron beam and detect its focusing by the magnetic field. In metals, electron focusing essentially is a *classical* phenomenon because of the small Fermi wavelength λ_F (typically 0.5 nm, on the order of the inter-atomic separation). Both the ratios λ_F/L and λ_F/W are much larger in a 2DEG than in a metal, typically by factors of 10^4 and 10^2, respectively. *Coherent* electron focusing[13–15] is possible in a 2DEG because of this relatively large value of the Fermi wavelength, and turns out to be strikingly different from classical electron focusing in metals.

The geometry of the experiment[13] in a 2DEG is the transverse focusing geometry of Tsoi,[24] and consists of two point contacts on the same boundary in a perpendicular magnetic field. (In metals, one also can use the geometry of Sharvin[23] with opposite point contacts in a longitudinal field. This is not possible in two dimensions.) Two split-gate quantum point contacts and the intermediate 2DEG boundary are created electrostatically by means of split gates, as described in Part II. On applying a negative voltage to the split-gate electrode shown in Fig. 2a, the electron gas underneath the gate structure is depleted, creating two 2DEG regions (i and c) electrically isolated from the rest of the 2DEG—apart from the two quantum point contacts under the 250 nm wide openings in the split gate. The devices studied had point contact separations L of 1.5 and 3.0 μm, both values being below the mean free path of 9 μm estimated from the mobility.

Electron focusing can be seen as a transmission experiment in electron optics; *cf.* Ref. 96 for a discussion from this point of view. An alternative point of view (emphasized in Ref. 15) is that coherent electron focusing is a prototype of a non-local resistance[153] measurement in the quantum ballistic transport regime, such as that studied extensively in narrow-channel geometries (cf. Chapter 3 by Timp). Longitudinal resistances that are negative, not $\pm B$ symmetric, and dependent on the properties of the current and voltage contacts as well as on their separation; periodic and aperiodic magnetoresistance oscillations; absence of local equilibrium—these all are characteristic features of this transport regime that appear in a most extreme and bare form in the electron focusing geometry. One reason for the simplification offered by this geometry is that the current and voltage contacts, being point contacts, are not nearly as invasive as the wide leads in a Hall bar geometry. Another reason is that the electrons interact with only one boundary (instead of two in a narrow channel). Apart from the intrinsic interest of electron focusing in a 2DEG, the experiment also can be seen as a method to study electron scattering—as in metals. For two such applications, see Refs. 154

and 155 A search for inelastic scattering far from equilibrium by means of hot electron focusing[61] is the subject of Section 7c

The outline of this section is as follows In Section 6 b, the experimental results on electron focusing[13 15] are presented A theoretical description[14 15] is given in Section 6 c, in terms of mode interference in the waveguide formed by the magnetic field at the 2DEG boundary A discussion of the anomalous quantum Hall effect in the electron focusing geometry[10 12 15] is deferred to Section 9, where we consider adiabatic quantum transport The present section is based on our earlier review[156] of this subject

b Experiment

Figure 24 illustrates electron focusing in two dimensions as it follows from the classical mechanics of electrons at the Fermi level The injector (i) injects electrons ballistically into the 2DEG The injected electrons all have the same Fermi velocity, but in different directions Electrons are detected if they reach the adjacent collector (c), after one or more specular reflections at the boundary connecting i and c These *skipping orbits* are composed of translated circular arcs of cylotron radius, $l_{\mathrm{cycl}} \equiv \hbar k_{\mathrm{F}}/eB$ The focusing action of the magnetic field is evident in Fig 24 (top) from the black lines of high density of trajectories These lines are known in optics as *caustics*, and are plotted separately in Fig 24 (bottom) The caustics intersect the 2DEG boundary at multiples of the cyclotron diameter from the injector As the magnetic field is

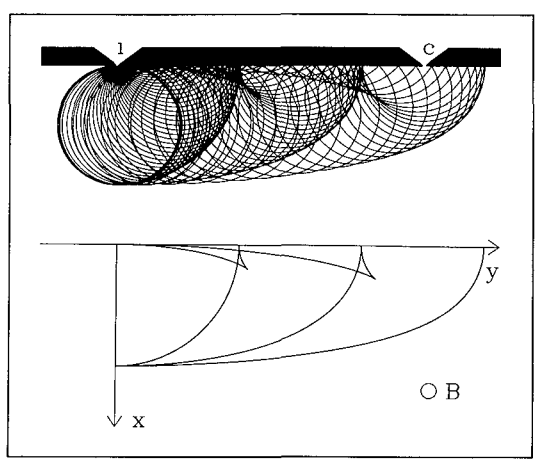

FIG 24 (Top) Skipping orbits along the 2DEG boundary The trajectories are drawn up to the third specular reflection (Bottom) Plot of the caustics, which are the collection of focal points of the trajectories (From Ref 15)

increased, a series of these focal points shifts past the collector. The electron flux incident on the collector thus reaches a maximum whenever its separation L from the injector is an integer multiple of $2l_{cycl}$. This occurs when

$$B = pB_{focus},\ p = 1, 2, \ldots, \text{with}$$
$$B_{focus} = \frac{2\hbar k_F}{eL}. \tag{24}$$

For a given injected current I_i, the voltage V_c on the collector is proportional to the incident flux. The classical picture thus predicts a series of equidistant peaks in the collector voltage as a function of magnetic field.

In Fig. 25 (top), we show such a classical focusing spectrum, calculated for parameters corresponding to the experiment discussed in this section ($L = 3.0\ \mu m, k_F = 1.5 \times 10^8\ m^{-1}$). The spectrum consists of equidistant focusing peaks of approximately equal magnitude superimposed on the Hall resistance (dashed line). The pth peak is due to electrons injected perpendicularly to the boundary that have made $p - 1$ specular reflections between injector and collector. Such a classical focusing spectrum commonly is observed in metals,[157] albeit with a decreasing height of subsequent peaks

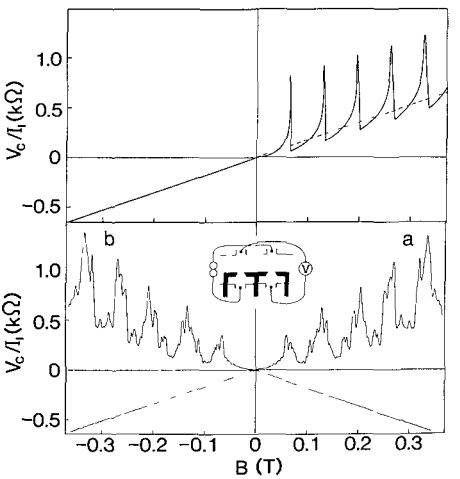

FIG. 25. (Bottom) Experimental electron focusing spectrum ($T = 50$ mK, $L = 3.0\ \mu m$) in the generalized Hall resistance configuration depicted in the inset. The two traces a and b are measured with interchanged current and voltage leads, and demonstrate the injector–collector reciprocity as well as the reproducibility of the fine structure. (Top) Calculated classical focusing spectrum corresponding to the experimental trace a. (50 nm-wide point contacts were assumed.) The dashed line is the extrapolation of the classical Hall resistance seen in reverse fields. (From Ref. 15.)

because of partially diffuse scattering at the metal surface Note that the peaks occur in one field direction only, in reverse fields, the focal points are at the wrong side of the injector for detection, and the normal Hall resistance is obtained The experimental result for a 2DEG is shown in the bottom half of Fig 25 (trace *a*, trace *b* is discussed later) A series of five focusing peaks is evident at the expected positions Note that the observation of multiple focusing peaks immediately implies that the electrostatically defined 2DEG boundary scatters predominantly *specularly* This conclusion is supported by magnetotransport experiments in a narrow channel defined by a split gate [108] In contrast, it has been found that a 2DEG boundary defined by ion beam exposure induces a large amount of *diffuse* scattering [108 155]

Fig 25 is obtained in a measuring configuration (inset) in which an imaginary line connecting the voltage probes crosses that between the current source and drain This is the configuration for a generalized Hall resistance measurement Alternatively, one can measure a generalized longitudinal resistance (*cf* Section 4) in the configuration shown in the inset of Fig 26 One then measures the focusing peaks without a superimposed Hall slope Note that the experimental longitudinal resistance (Fig 26, bottom) becomes *negative* This is a classical result of magnetic focusing, as demonstrated by the calculation shown in the top half of Fig 26 Buttiker[158 159] has studied negative longitudinal resistances in a different (Hall bar) geometry

On the experimental focusing peaks, a fine structure is evident The fine structure is quite reproducible (as is evident when comparing Figs 25 and 26), but sample-dependent It is resolved only at low temperatures (below 1 K) and small injection voltages (The measurements shown are taken at 50 mK and a few μV ac voltage over the injector) A nice demonstration of the reproducibility of the fine structure is obtained upon interchanging current and voltage leads, so that the injector becomes the collector and vice versa The resulting focusing spectrum shown in Fig 25 (trace *b*) is almost

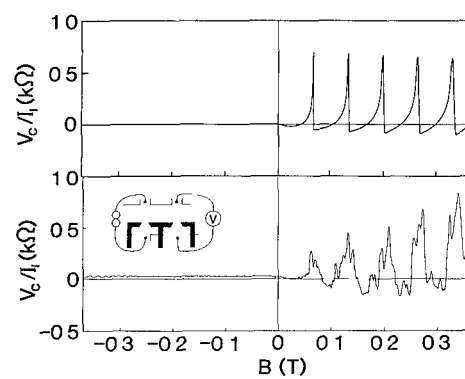

FIG 26 Same as Fig 25, but in the longitudinal resistance configuration (From Ref 15)

the precise mirror image of the original one (trace *a*)—although this particular device had a strong asymmetry in the widths of injector and collector. The symmetry in the focusing spectra is a consequence of the fundamental reciprocity relation derived by Büttiker,[7,158] which generalizes the familiar Onsager–Casimir symmetry relation for the resistivity tensor to resistances.

The fine structure on the focusing peaks in Figs. 25 and 26 is the first indication that electron focusing in a 2DEG is qualitatively different from the corresponding experiment in metals. At higher magnetic fields, the resemblance to the classical focusing spectrum is lost. (See Fig. 27.) A Fourier transform of the spectrum for $B \geq 0.8$ T (inset in Fig. 27) shows that the large-amplitude high-field oscillations have a dominant periodicity of 0.1 T, which is approximately the same as the periodicity B_{focus} of the much smaller focusing peaks at low magnetic fields. (B_{focus} in Fig. 27 differs from Fig. 25 because of a smaller $L = 1.5$ μm.) This dominant periodicity is the result of quantum interference between the different trajectories in Fig. 24 that take an electron from injector to collector. In Section 6.c, we demonstrate this in a mode picture, which in the WKB approximation is equivalent to calculating the interferences of the (complex) probability amplitude along classical trajectories. The latter ray picture is treated extensively in Ref. 15. The theoretical analysis implies for the experiment that the injector acts as a *coherent* point source with the coherence maintained over a distance of several microns to the collector.

c. *Edge Channels and Mode Interference*

To explain the characteristics features of coherent electron focusing mentioned previously, it is necessary to go beyond the classical description.[14,15]

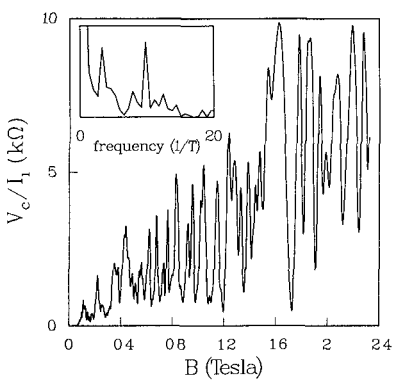

FIG. 27. Experimental electron focusing spectrum over a larger field range and for very narrow point contacts (estimated width 20–40 nm; $T = 50$ mK, $L = 1.5$ μm). The inset gives the Fourier transform for $B \geq 0.8$ T. The high-field oscillations have the same dominant periodicity as low-field focusing peaks—but with a much larger amplitude. (From Ref. 15.)

As discussed briefly in Section 3 (*cf* Fig 11a), quantum ballistic transport along the 2DEG boundary in a magnetic field takes place via magnetic edge states, which are the propagating modes of this problem [160] The modes at the Fermi level are labeled by a quantum number $n = 1, 2, \quad N$ Since the injector has a width below λ_F, it excites these modes coherently For $k_F L \gg 1$, the interference of modes at the collector is dominated by their rapidly varying phase factors $\exp(ik_n L)$ The wave number k_n in the y direction (along the 2DEG boundary, see Fig 24 for the choice of axes) corresponds classically to the x coordinate of the center of the cyclotron orbit, which is a conserved quantity upon specular reflection at the boundary [161] In the Landau gauge $A = (0, Bx, 0)$, this correspondence may be written as $k_n = k_F \sin\alpha_n$, where α is the angle with the x axis under which the cyclotron orbit is reflected from the boundary ($|\alpha| < \pi/2$) The quantized values α_n follow in this semiclassical description from the Bohr–Sommerfeld quantization rule[160 161] that the flux enclosed by the cyclotron orbit and the boundary equals $(n - \frac{1}{4})$ h/e (for an infinite barrier potential) Simple geometry shows that this requires that

$$\frac{\pi}{2} - \alpha_n - \frac{1}{2}\sin 2\alpha_n = \frac{2\pi}{k_F l_{cycl}}\left(n - \frac{1}{4}\right), n = 1, 2, \quad N, \tag{25}$$

with N the largest integer smaller than $\frac{1}{2}k_F l_{cycl} + \frac{1}{4}$ As plotted in Fig 28, the dependence on n of the phase $k_n L$ is close to linear in a broad interval This also follows from expansion of Eq (25) around $\alpha_n = 0$, which gives

$$k_n L = \text{constant} - 2\pi n\frac{B}{B_{focus}} + k_F L \times \text{order}\left(\frac{N - 2n}{N}\right)^3 \tag{26}$$

If B/B_{focus} is an integer, a fraction of order $(1/k_F L)^{1/3}$ of the N edge states interferes constructively at the collector (The edge states outside the domain of linear n-dependence of the phase give rise to additional interference struc-

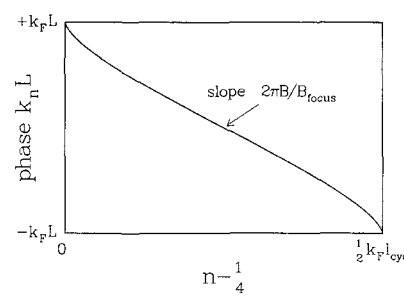

FIG 28 Phase $k_n L$ of the edge channels at the collector, calculated from Eq (25) Note the domain of approximately linear n-dependence of the phase, responsible for the oscillations with B_{focus}-periodicity (From Ref 15)

ture which, however, does not have a simple periodicity.) Because of the 1/3 power, this is a substantial fraction even for the large $k_F L \sim 10^2$ of the experiment. The resulting mode interference oscillations with B_{focus}-periodicity can become much larger than the classical focusing peaks. This has been demonstrated in Refs. 14 and 15, where the collector voltage has been determined in WKB approximation with neglect of the finite width of the injector and detector. The result obtained there can be written in the form,

$$\frac{V_c}{I_i} = \frac{h}{2e^2}\left|\frac{1}{N}\sum_{n=1}^{N} e^{ik_n L}\right|^2. \tag{27}$$

Note that this equation implies that in the absence of interference among the modes, the normal quantum Hall resistance $h/2Ne^2$ is obtained. This is *not* a general result, but depends specifically on the properties of the injector and collector point contacts—as we will discuss in Section 9.

Figure 29 gives the focusing spectrum from Eq. (27), with parameter values corresponding to the experimental Fig. 27. The inset shows the Fourier transform for $B \geq 0.8$ T. There is no detailed one-to-one correspondence between the experimental and theoretical spectra. No such correspondence was to be expected in view of the sensitivity of the experimental spectrum to small variations in the voltage on the gate defining the point contacts and the 2DEG boundary.[13,15] Those features of the experimental spectrum that are insensitive to the precise measurement conditions are well-produced however, by the calculation: We recognize in Fig. 29 the low-field focusing peaks and the large-amplitude high-field oscillations with the same periodicity. (The reason that the periodicity B_{focus} in Fig. 29 is somewhat larger than in Fig. 27 most likely is the experimental uncertainty in the effective point contact separation of the order of the split-gate opening of 250 nm.) The high-field oscillations

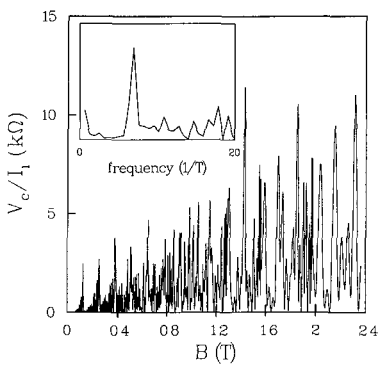

FIG. 29. Focusing spectrum calculated from Eq. (27) for parameters corresponding to the experimental Fig. 27. The inset shows the Fourier transform for $B \geq 0.8$ T. Infinitesimally small point contact widths are assumed in the calculation. (From Ref. 156.)

range from about 0 to 10 kΩ in both theory and experiment. This maximum amplitude is not far below the theoretical upper bound of $h/2e^2 \approx 13$ kΩ, which follows from Eq. (27) if we assume that *all* the modes interfere constructively. This indicates that a *maximal phase coherence* is realized in the experiment, and implies that:

1. The experimental injector and collector point contacts resemble the idealized point source/detector in the calculation.
2. Scattering events other than specular scattering on the boundary can be largely ignored (since any other inelastic as well as elastic scattering events would scramble the phases and reduce the oscillations with B_{focus}-periodicity).

The theory can be improved in several ways. This will affect the detailed form of the spectra, but probably not the fundamental periodicity. Since the exact wave functions of the edge states are known (Weber functions), one could go beyond the WKB approximation. This will become important at large magnetic fields, when the relevant edge states have small quantum numbers. In this regime, one also would have to take into account a possible B-dependence of E_F relative to the conduction band bottom (due to pinning of the Fermi energy at the Landau levels; *cf.* the related discussion in Section 3.a). It would be interesting to find out to what extent this bulk effect is reduced at the 2DEG boundary by the presence of edge states to fill the gap between the Landau levels. Another direction of improvement is towards a more realistic modeling of the injector and collector point contacts. Since the maximum amplitude of the theoretical and experimental oscillations is about the same (as is evident when comparing Figs. 27 and 29), the loss of spatial coherence due to the finite point contact size does not seem to be particularly important in this experiment. (Infinitesimal point contact width was assumed in the calculation.) On the other hand, the experimental focusing spectrum does not contain as many rapid oscillations as the calculation would predict. This may be due to the collimating properties of the point contacts. (See Ref. 156).

7. Breakdown of the Conductance Quantization and Hot Electron Focusing

a. Mechanisms for Nonlinear Ballistic Transport

Nonlinear transport in semiconductor devices is the rule, rather than an exception, but its analysis can be quite complicated.[162–164] In this section, we are concerned with several experiments that extend the study of ballistic

transport to the regime of a nonlinear dependence of current on voltage. Although our theoretical understanding is far from being complete, we shall demonstrate that the overall behavior is readily understood on the basis of simple considerations. This simplicity is due to the ballistic nature of the transport on the length scales probed by the experiments. This allows us to ignore the energy dependence of the transport mean free path, which underlies much of the complexity of nonlinear transport in the diffusive transport regime.

Nonlinear transport through metallic point contacts has been investigated widely because of the possibility to observe phonon-related structure in the second derivative of the current-voltage characteristics. This is known as *point contact spectroscopy.*[25,26] Since typical optical phonon energies are of the order of 30 meV, while the Fermi energy in a metal is several eV, classical nonlinearities governed by the parameter eV/E_F do not play a significant role in metals. In a 2DEG, where E_F typically is only 10 meV, the situation is reversed, and the latter effects so far have obscured possible structure due to inelastic scattering processes in single quantum point contacts. The typical energy scale for nonlinearities in the quantum ballistic transport regime is even smaller than E_F. In experiments on the conductance quantization of a quantum point contact, the maximum breakdown voltage is found to be the energy separation between consecutive subbands.[60] This is discussed in Section 7.b. A complication is formed by the presence of a potential barrier in the point contact. The barrier height E_c depends on the applied bias voltage, thereby forming an additional mechanism for nonlinear transport. This complication may be turned into an advantage because it allows the injection of hot electrons over the barrier for sufficiently large bias voltage. Hot electron transport has been studied widely in devices inspired by the hot electron transistors pioneered by Shannon,[165] and this field of research has matured since the advent of vertically layered structures. A high sensitivity to inelastic scattering processes may be achieved in geometries involving two barriers, if one is used as a hot electron injector and the other as an energy-selective collector. This technique, known as *hot-electron spectroscopy,*[166–168] was adapted recently to transport in the plane of a 2DEG.[169–171] Among the results, we quote the demonstration by Palevski *et al.*[169] of emission of single longitudinal optical phonons (which in GaAs have an energy of 36 meV) and the discovery by Sivan *et al.*[171] of a long inelastic mean free path (exceeding 2 μm) for hot electrons with an excess energy up to the phonon energy. This finding contradicts theoretical predictions that the inelastic mean free path for electron–electron interactions should be one or two orders of magnitude smaller.[172,173] An entirely new way of detecting ballistic hot electrons is the *electron focusing* technique, which is the solid state analog of a β-spectrometer. As discussed in Section 7.c, the experimental results[61] corroborate the finding of Sivan *et al.* of ballistic hot-electron transport on long-length scales.

The hot-electron focusing technique is special for another reason: It can be used to determine the magnitude of the voltage drop in the immediate vicinity of a quantum point contact.[61] This may be the first realization of a really non-invasive voltage probe.

Two additional sources of nonlinearity are characteristic for quantum transport, but are not discussed beyond this introduction. The first is (resonant) tunneling through potential barriers, which has been investigated widely in vertical layered semiconductor structures following original work by Chang, Esaki, and Tsu.[174–176] Only recently has tunneling been studied in double-barrier structures defined in a 2DEG.[169,177] Clear signatures of tunneling currents have not been found yet in the current-voltage characteristics of single quantum point contacts, but the physics should be very similar to that of Ref. 169. An interesting difference between point contacts and wide tunnel barriers is the spatial resolution, which in a geometry with two opposite point contacts may be exploited to impose constraints on the lateral momentum of the tunneling electrons.[171]

A second source of quantum mechanical nonlinearity[178] arises in experiments that, like coherent electron focusing, probe the coherence of the injected electrons. Energy averaging due to a small nonzero bias voltage V is very similar to that due to a finite temperature $T \approx eV/k_B$, as discussed in Section 2.c (Ref. 101). This similarity is supported by experimental observations. (For the coherent electron focusing experiment, one finds,[15] for example, that energy averaging becomes important if the temperature is raised above 1 K, or for injection voltages beyond 100 μV.)

b. Breakdown of the Conductance Quantization

i. Experiments. The breakdown of the conductance quantization has been studied[60] by measuring the dc current-voltage (I–V) characteristics of a quantum point contact device at a temperature of 0.6 K. The I–V traces were obtained for a set of gate voltages V_g in an interval from -2.0 to -2.1 V, corresponding to the $N = 1$ conductance plateau at small bias voltages. (See inset of Fig. 30a.) These measurements are representative for the case that the point contact at zero bias is not yet pinched off, which implies that in equilibrium the Fermi energy E_F exceeds the energy E_1 of the bottom of the lowest subband in the constriction. We will refer to E_1 as the height of the effective barrier in the constriction. The case $E_F < E_1$ of a pinched-off point contact is discussed in the next paragraph. The experimental I–V traces reproduced in Figs. 30a,b have been obtained from a two-terminal resistance measurement, after a correction for the background resistance originating in the ohmic contacts and the wide 2DEG regions. (See Part II and Ref. 60.) The dotted lines in Figs. 30a,b represent the quantized conduc-

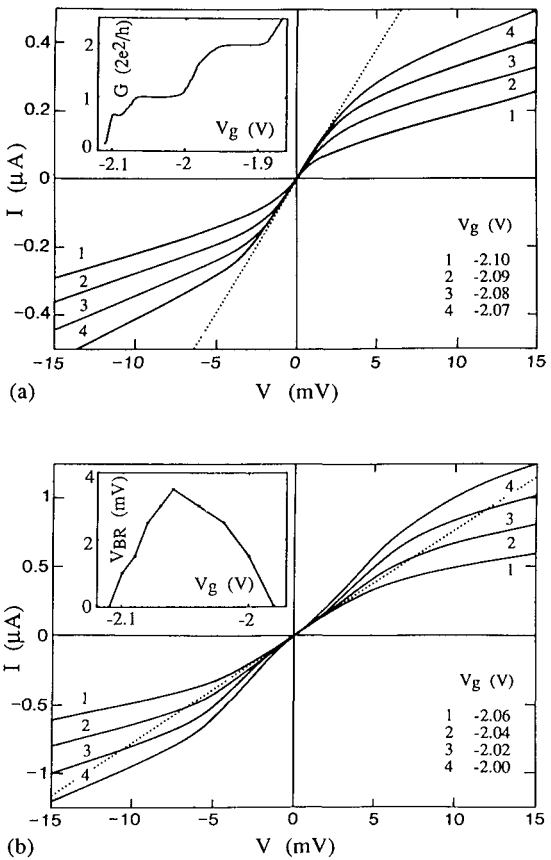

FIG. 30. Current-voltage characteristics of a quantum point contact at 0.6 K for different values of the gate voltage V_g at which the conductance G for small bias voltages is quantized at $2e^2/h$. The insets show the gate voltage dependence of the small-bias conductance G, and of the voltage V_{BR} beyond which the quantization $I = (2e^2/h)V$ (indicated by the dotted line) breaks down. The $I-V$ characteristics shown have been corrected for a background resistance. (From Ref. 60.)

tance $I/V = 2e^2/h$. The conductance quantization breaks down beyond a critical voltage V_{BR}, which depends sensitively on the gate voltage (inset of Fig. 30b). The dependence of V_{BR} on V_g has a characteristic triangular shape, with a maximum of about 3 mV. Note that the maximum of V_{BR} is comparable to the subband separation at the Fermi level (calculated in Sec. 3b). On increasing the bias voltage V beyond V_{BR}, the differential conductance $\partial I/\partial V$ is seen initially to be either smaller (Fig. 30a) or larger (Fig. 30b) than the quantized value. As discussed shortly, this can be understood qualitatively as a consequence of the unequal number of populated subbands for the two

opposite velocity directions in the constriction (whereby the bias voltage-induced lowering of the effective barrier height E_1 in the constriction plays a central role) Eventually, $\partial I/\partial V$ is found to drop below $2e^2/h$, regardless of the value of the gate voltage (in the interval considered in Fig 30) We note that no evidence for a saturation of the current has been found for voltages up to 200 mV

On increasing the negative gate voltage beyond -2 1 V, the point contact is pinched off, and hardly any current flows for small voltages The residual current in the regime in which the point contact is just pinched off may be caused by a combination of tunneling[179] and thermionic emission over the barrier[180] (depending on the temperature and the value of the gate voltage) Measurements in this regime will be meaningful only if the leakage current through the gate is much less than this current We have not investigated this regime systematically yet (See in this connection Ref 169) When the point contact is pinched off, the effective barrier height in the constriction exceeds the Fermi level in equilibrium ($E_F < E_1$) An appreciable current starts to flow only if the bias voltage is sufficiently large to tilt the barrier such that electrons injected by one reservoir can pass over it freely The experimental results are shown in Fig 31 It is seen that a clear conduction threshold exists, which depends sensitively on the gate voltage Beyond this threshold, the differential conductance $\partial I/\partial V$ following from these $I-V$ curves is roughly independent of the gate voltage (with a value of $(80\ \mathrm{k\Omega})^{-1}$), while it also is seen to be of comparable magnitude as the *trans*-conductance $\partial I/\partial V_g$ The latter result is significant in view of the symmetry relation discussed next

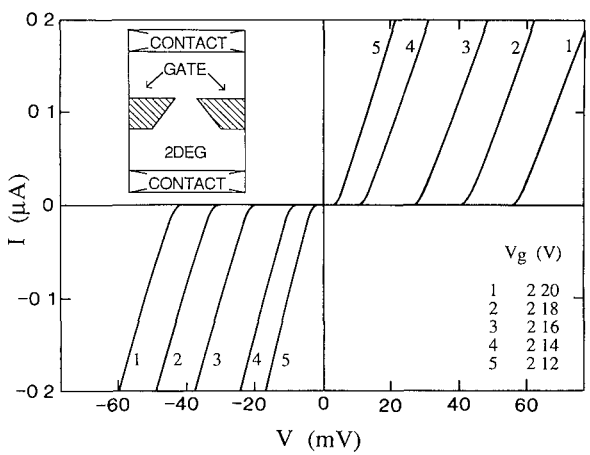

FIG 31 *I V* characteristics at different gate voltages V_g for which the constriction is pinched off at small bias voltage The inset shows the sample layout (From Ref 60)

ii. Symmetry of the current-voltage characteristics. Before we turn to an account of some models that explain the basic features of the experimental results, it is of interest to examine the symmetry of the $I-V$ curves under a polarity change of the voltage V between source and drain. None of the experimental $I-V$ curves shown in Figs. 30 and 31 is anti-symmetric in V. This does not necessarily imply an intrinsic asymmetry in the device geometry. The reason is that point contacts are three- rather than two-terminal devices, and one of the current-carrying contacts also serves as the zero-reference of the gate voltage V_g, for either polarity of V. The presented current-voltage characteristics have been obtained without changing the choice of zero reference or the value of the gate voltage. The proper symmetry relation for a three-terminal device with mirror symmetry between source and drain is

$$I(V, V_g) = -I(-V, V_g - V), \tag{28}$$

rather than simply $I(V) = -I(-V)$. Here, we have assumed that V as well as V_g are defined relative to the same current contact, as is the case in Figs. 30 and 31. The difference between the two relations is significant, even though V is smaller than V_g by up to two orders of magnitude, because of the comparable magnitude of the differential conductance and the differential trans-conductance noted earlier. Equation (28) can be applied to Figs. 30 and 31 by comparing selected data points on curves measured for different gate voltages. It becomes readily apparent from such a procedure that gross deviations from Eq. (28) are found for the point contact under investigation, indicating that it did not have the mirror symmetry between source and drain. Deviations from Eq. (28) in other devices were found to be much smaller, however.

It will be clear from the preceding considerations that a precise modeling of the observed results would require a knowledge of the complex interdependence of the shape and height of the effective barrier in the constriction on the bias voltage and the gate voltage. In the following, we employ a highly simplified model, which suffices nevertheless for a qualitative description of the experiments.

iii. $I-V$ characteristic of a semiclassical point contact. Consider semiclassical transport at large voltages through a point contact in a 2DEG defined by a hard-wall confining potential of width W. In contrast to our earlier treatment in the linear transport regime (Section 2.b.i), the net current now can be considered no longer to be carried exclusively by electrons at the Fermi energy. Instead, electrons in a finite energy interval contribute to the current. The fundamental origin of the resulting nonlinear $I-V$ characteristic is the interdependence of the Fermi velocity v_F and the electron gas density n_s in a degenerate electron gas, given by $v_F = (2\pi n_s)^{1/2}\hbar/m$ in two dimensions. We

illustrate this by a simple model,[66] related to those commonly used in the literature on vertical transport in hot-electron spectroscopy devices,[166–168] and also to models for tunneling through a potential barrier under finite bias.[179,181,182]

The reservoirs on either side of the constriction are assumed to maintain a Fermi–Dirac distribution in the 2DEG for the electrons that move *towards* the constriction, with a Fermi energy difference eV. The electric field thus is assumed to be nonzero only in the constriction itself. The distribution of electrons moving *away* from the constriction can be far from equilibrium, however. According to these assumptions, the current is carried by conduction electrons in an energy interval from max $(E_F - e|V|, 0)$ to E_F, and can be written in the form (compared with Section 2.b.i),

$$I = eW \int_{\max(E_F - e|V|,0)}^{E_F} \rho(E) \left(\frac{2E}{m}\right)^{1/2} dE \int_{-\pi/2}^{\pi/2} \cos\phi \frac{d\phi}{2\pi}, \tag{29}$$

with $\rho(E) = m/\pi\hbar^2$ the density of states, which is energy-independent in two dimensions. The resulting current no longer is linear in the applied voltage,

$$I = I_{\max}\left[1 - \left(1 - \frac{e|V|}{E_F}\right)^{3/2}\right], \text{ for } e|V| < E_F, \tag{30a}$$

$$I = I_{\max}, \text{ for } e|V| > E_F, \tag{30b}$$

with

$$I_{\max} = e\frac{m}{\pi\hbar^2}\frac{2W}{3\pi}\left(\frac{2}{m}\right)^{1/2} E_F^{3/2} = en_s v_F \frac{2W}{3\pi}. \tag{31}$$

For $e|V| \ll E_F$, one recovers Eqs. (1) and (2). For $e|V| > E_F$, we find that the current is limited by a saturation value $I_{\max}$. This saturation primarily is a consequence of our assumption that the entire voltage drop is localized at the point contact, with neglect of any accelerating fields outside the point contact region. We return to this point shortly.

The effect of a potential barrier of height E_c in the constriction can be taken into account by replacing E_F with $E_F - E_c$ in Eqs. (30) and (31). A complete description also has to determine the dependence of the barrier height on the applied bias voltage. (See next section).

iv. Breakdown of the conductance quantization. A model for the nonlinear I–V characteristics in the quantum ballistic case has been given in Ref. 60, on which our present discussion is based. This problem also has been considered in Refs. 183 and 101. To be specific, we consider an idealized model of

a trapezoidal effective potential barrier in the constriction,[125] as illustrated in Fig. 32. The applied source-drain voltage V is assumed to result in a constant electric field between entrance and exit of the point contact. Due to this electric field, the potential barrier in the point contact is tilted, and thereby lowered, as illustrated in Figs. 32b,c. Due to the lateral confinement, one-dimensional subbands are formed in the constriction. On entering the constriction, the bottom of the nth subband rises relative to the bulk 2DEG, as a combined result of the increased lateral confinement and the electrostatic barrier (responsible for the reduced density in the constriction). The nth subband bottom has a maximal energy E_n, constituting a bottleneck for the current. We calculate the net current I_n through the constriction carried by the nth subband by considering the occupation of the right- and left-moving states at the bottleneck. The right-moving states are filled from E_n up to μ_s, the electrochemical potential of the source at the left of the constriction (provided $\mu_s > E_n$). Analogously, provided that $\mu_d > E_n$, the left-moving states

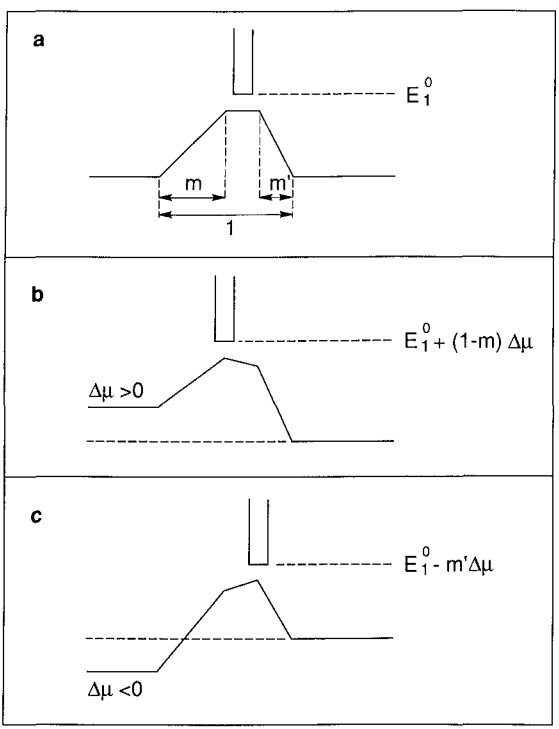

FIG. 32. Illustration of the voltage-induced tilting of the barrier in the constriction, for the case of an asymmetric trapezoidal shape of the barrier, with aspect ratios m and m'. The lowest one-dimensional subband is indicated. The energy E_1^0 of the subband bottom at the maximum of the trapezium is the effective barrier height at zero bias. (From Ref. 125.)

are filled from E_n up to μ_d, the electrochemical potential of the drain to the right of the constriction. We assume that the electrons in the nth subband with energy $\mu > E_n$ are transmitted fully through the constriction, and neglect intersubband scattering. In the following, we take $\mu_s > \mu_d$ (as in Fig. 32b), and define $\Delta\mu \equiv \mu_s - \mu_d \equiv eV$. The difference in occupation of right- and left-moving states for nonzero $\Delta\mu$ gives rise to a net current in the nth subband, which, according to the cancellation of group velocity and one-dimensional density of states (discussed in Section 2.b.ii), is given by

$$I_n = \frac{2e}{h}[\mu_s - \max(\mu_d, E_n)]\theta(\mu_s - E_n), \tag{32}$$

with $\theta(x)$ the unit step function. The bottom of the subband E_n at the bottleneck differs from its equilibrium position (which we denote by E_n^o in this section) because of the tilting of the potential barrier, and is given by $E_n = E_n^o + (1 - m)\Delta\mu$. (See Fig. 32b.) Here, m is a phenomenological parameter between 0 and 1, which gives the fraction of $\Delta\mu$ that drops to the left of the bottleneck. (The value of m is determined by the shape of the effective potential barrier, as shown in Fig. 32a.) At a characteristic voltage V_c or V_c', either μ_s or μ_d crosses the subband bottom E_n, thereby changing the contribution $\partial I_n/\partial V$ from the nth subband to the differential conductance, $g \equiv \partial I/\partial V$. From Eq. (32), one finds[60] for an initially unpopulated subband,

$$\frac{\partial I_n}{\partial V} = \frac{2e^2}{h} m\theta(V - V_c), \text{if } E_F < E_n^o, \tag{33}$$

while for an initially populated subband, one has

$$\frac{\partial I_n}{\partial V} = \frac{2e^2}{h}[1 - (1 - m)\theta(V - V_c')], \text{if } E_F > E_n^o. \tag{34}$$

The breakdown voltages are given by

$$V_c = \frac{(E_n^o - E_F)}{me}, \tag{35}$$

$$V_c' = \frac{(E_F - E_n^o)}{(1 - m)e}, \tag{35b}$$

and they depend on the subband index n. (Note that m itself depends on n as well, since the effective potential barrier is due in part to the lateral confinement.)

Eq. (35b) applies to a subband that is occupied in equilibrium ($E_F > E_n^o$). Beyond a critical voltage V_c', the differential conductance due to this subband *decreases* from its normal quantized value of $2e^2/h$. Equation (35a) applies to a subband that in equilibrium is not occupied at the bottleneck of the constriction ($E_F < E_n^o$). The differential conductance of this suband *increases* beyond a critical voltage V_c to a value that is smaller than the quantized value. Although the expressions for the critical voltage depend on the parameter m, these general conclusions presumably are model-independent. The smallest breakdown voltage (which corresponds either to the highest occupied subband or the lowest unoccupied one) determines the breakdown voltage observed experimentally. If V is smaller than the breakdown voltages V_c, V_c' for all n, then we recover the conductance quantization $G \equiv I/V = (2e^2/h)N$ (with N the number of subbands occupied in equilibrium). This result exemplifies an interesting difference between quantum ballistic and semiclassical ballistic transport, in that it predicts linear transport *exactly* in a finite voltage interval, in contrast to Eq. (30).

v. Discussion. To illustrate the consequences[60] of Eqs. (33) and (34) for the differential conductance $g \equiv \partial I/\partial V$, we have shown schematically in Fig. 33

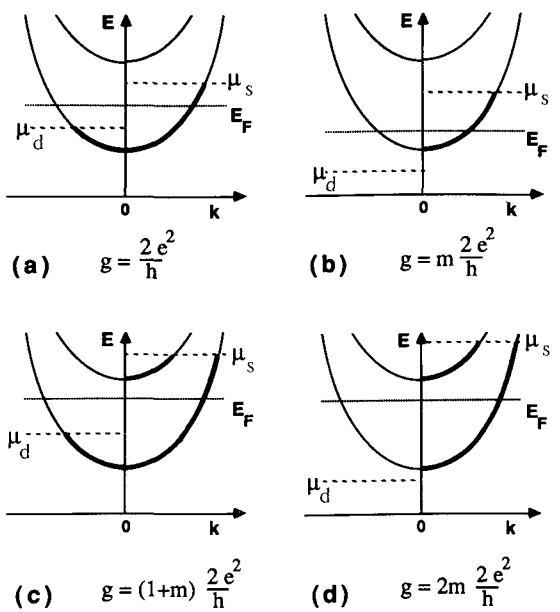

FIG. 33. Occupation of two one-dimensional subbands (represented by the parabolic dispersion curves $E_n(k)$) at the bottleneck in the constriction. Four situations are illustrated for different voltages $V = (\mu_s - \mu_d)/e$ across the point contact, and for different values of E_F (relative to the potential barrier height). (From Ref. 60.)

the energy of the two lowest one-dimensional subbands at the bottleneck as a function of longitudinal wave vector k. In equilibrium ($V = 0$), the subbands are occupied up to the Fermi energy E_F. A voltage V across the constriction gives a difference $\mu_s - \mu_d = eV$ in occupation between the two velocity directions (Fig. 33a), resulting in a net current. As long as the number of occupied subbands is the same for the two velocity directions, the conductance is quantized. However, at larger applied voltages, μ_d can fall below the bottom of a subband. Then g is reduced to a fraction m of $2e^2/h$ as shown in Fig. 33b, where E_F is near the bottom of the lowest subband. This is observed experimentally in Fig. 30a. The subband occupation of Fig. 33b also can be reached from the situation $E_F < E_1$, where there are no occupied states in equilibrium. For low voltages $g = 0$, but at a voltage V_c, the chemical potential μ_s crosses E_1, so that g increases to $m(2e^2/h)$ (*cf.* Eq. (33)). This is the case studied experimentally in Fig. 31.

Figures 33c,d correspond to the situation where E_F is close to the bottom of the second subband, as in the experimental Fig. 30b. On increasing V, first the second subband starts to be populated (Fig. 30c), leading to an increase in g to $(1 + m)2e^2/h$. A further increase of V causes μ_d to fall below the bottom of the first subband (Fig. 30d), which then reduces g to $2m(2e^2/h)$. This explains qualitatively the increasing and then decreasing slope in Fig. 30b. We note that the situation of Fig. 33d also can be reached directly from that in Fig. 33a, which actually is happening at $V_g = -2.06$ V in Fig. 30b.

So far, we have considered the case $\Delta\mu > 0$. If $\Delta\mu < 0$, then the parameter m is replaced by m', which will be different from m in the case of an assymmetric barrier in the constriction. (See Fig. 32c.) This may explain part of the asymmetry in the $I-V$ characteristics of Figs. 30 and 31, discussed earlier.

Both the quantum mechanical result, Eqs. (33)–(34), and its semiclassical counterpart Eq. (30), predict a saturation current of order I_{max} given by Eq. (31). The magnitude of the expected saturation current for $E_F = 10$ meV is $I_{max} \sim 1$ μA for a constriction with a few occupied one-dimensional subbands. Such a saturation has not been observed experimentally. The likely reason is that for large applied voltages, the electric field no longer is localized at the point contact. The electric field outside the point contact region will give rise to an acceleration of the electrons before they enter the point contact (i.e., in a wider region), thereby enhancing I_{max}. This point is discussed further in Section 7.c.

More recently, Brown *et al.*[102,184,185] have studied the differential resistance of a quantum point contact. In particular, they predict that the current saturation of Eq. (31) should cross over to a negative differential resistance regime at sufficiently large bias voltages as a consequence of quantum mechanical reflection. Further experiments will be required to test this prediction.

Here, we have considered only $I-V$ characteristics at fixed V_g. Glazman and Khaetskii[183] have predicted that the differential conductance as a func-

tion of gate voltage at finite V should exhibit additional plateaus between those at multiples of $2e^2/h$. Some evidence has been found for such plateaus (which also follow from Eqs. (33) and (34)), but these were not well resolved.[60]

The results Eqs. (33) and (34) predict a maximum breakdown voltage given by the one-dimensional subband separation at the Fermi level. This criterium is equivalent to that obtained by Jain and Kivelson[186] for the breakdown of the quantum Hall effect, where the Landau levels take over the role of the one-dimensional subbands. The triangular dependence of the breakdown voltage on the gate voltage shown in Fig. 30b also is reminiscent of experiments on the breakdown of the quantum Hall effect, where the breakdown voltage has a triangular dependence on the magnetic field. (Recent discussions of this topic are Refs. 30, 136, 187–191.) These similarities emphasize the correspondence between the zero-field quantization and the quantum Hall effect, discussed in Section 3.

c. *Hot Electron Focusing*

The voltage drop in a current-carrying sample containing a point contact in general is highly localized at or near the constriction. As mentioned in Section 4, two-terminal resistance measurements do not yield information on the voltage distribution. Four-terminal measurements could do so in principle, but such measurements in practice are hampered by the fact that conventional voltage probes located within a mean free path of the point contact are *invasive*. Indeed, in narrow multi-probe conductors, the voltage probes are the dominant source of scattering, as discussed by Timp in Chapter 3. In addition, conventional probes measure an electrochemical potential rather than an electrostatic voltage. In the linear response regime, a knowledge of the actual electric field distribution is not required to know the dissipation in the system (*cf.* Section 4). For many applications beyond linear transport, however, the electric field distribution does matter, as emphasized repeatedly by Landauer.[6,83,178,192–194] We thus are faced with a challenge to overcome the limitations of conventional resistance measurements in the ballistic regime. In this section, we discuss how an extension of the electron focusing technique of Section 6 to finite applied voltages meets this challenge at least in part.[61] (A promising alternative technique is to use a scanning tunneling microscope.)[194,195]

In Section 6, we emphasized the great difference in *length scales* between electron focusing in metals and in a 2DEG. The experiments discussed next demonstrate another qualitative difference, that of *energy scales*. In metals,[196] electrons are injected at energies above E_F, which generally are much less than $E_F \approx 5$ eV. In contrast, the Fermi energy in a 2DEG is only about 10 meV, so that dc-biasing the small ac injection voltage used in the electron

the energy of the two lowest one-dimensional subbands at the bottleneck as a function of longitudinal wave vector k. In equilibrium ($V = 0$), the subbands are occupied up to the Fermi energy E_F. A voltage V across the constriction gives a difference $\mu_s - \mu_d = eV$ in occupation between the two velocity directions (Fig. 33a), resulting in a net current. As long as the number of occupied subbands is the same for the two velocity directions, the conductance is quantized. However, at larger applied voltages, μ_d can fall below the bottom of a subband. Then g is reduced to a fraction m of $2e^2/h$ as shown in Fig. 33b, where E_F is near the bottom of the lowest subband. This is observed experimentally in Fig. 30a. The subband occupation of Fig. 33b also can be reached from the situation $E_F < E_1$, where there are no occupied states in equilibrium. For low voltages $g = 0$, but at a voltage V_c, the chemical potential μ_s crosses E_1, so that g increases to $m(2e^2/h)$ (*cf.* Eq. (33)). This is the case studied experimentally in Fig. 31.

Figures 33c,d correspond to the situation where E_F is close to the bottom of the second subband, as in the experimental Fig. 30b. On increasing V, first the second subband starts to be populated (Fig. 30c), leading to an increase in g to $(1 + m)2e^2/h$. A further increase of V causes μ_d to fall below the bottom of the first subband (Fig. 30d), which then reduces g to $2m(2e^2/h)$. This explains qualitatively the increasing and then decreasing slope in Fig. 30b. We note that the situation of Fig. 33d also can be reached directly from that in Fig. 33a, which actually is happening at $V_g = -2.06$ V in Fig. 30b.

So far, we have considered the case $\Delta\mu > 0$. If $\Delta\mu < 0$, then the parameter m is replaced by m', which will be different from m in the case of an assymmetric barrier in the constriction. (See Fig. 32c.) This may explain part of the asymmetry in the $I-V$ characteristics of Figs. 30 and 31, discussed earlier.

Both the quantum mechanical result, Eqs. (33)–(34), and its semiclassical counterpart Eq. (30), predict a saturation current of order I_{max} given by Eq. (31). The magnitude of the expected saturation current for $E_F = 10$ meV is $I_{max} \sim 1$ μA for a constriction with a few occupied one-dimensional subbands. Such a saturation has not been observed experimentally. The likely reason is that for large applied voltages, the electric field no longer is localized at the point contact. The electric field outside the point contact region will give rise to an acceleration of the electrons before they enter the point contact (i.e., in a wider region), thereby enhancing I_{max}. This point is discussed further in Section 7.c.

More recently, Brown *et al.*[102,184,185] have studied the differential resistance of a quantum point contact. In particular, they predict that the current saturation of Eq. (31) should cross over to a negative differential resistance regime at sufficiently large bias voltages as a consequence of quantum mechanical reflection. Further experiments will be required to test this prediction.

Here, we have considered only $I-V$ characteristics at fixed V_g. Glazman and Khaetskii[183] have predicted that the differential conductance as a func-

tion of gate voltage at finite V should exhibit additional plateaus between those at multiples of $2e^2/h$. Some evidence has been found for such plateaus (which also follow from Eqs. (33) and (34)), but these were not well resolved.[60]

The results Eqs. (33) and (34) predict a maximum breakdown voltage given by the one-dimensional subband separation at the Fermi level. This criterium is equivalent to that obtained by Jain and Kivelson[186] for the breakdown of the quantum Hall effect, where the Landau levels take over the role of the one-dimensional subbands. The triangular dependence of the breakdown voltage on the gate voltage shown in Fig. 30b also is reminiscent of experiments on the breakdown of the quantum Hall effect, where the breakdown voltage has a triangular dependence on the magnetic field. (Recent discussions of this topic are Refs. 30, 136, 187–191.) These similarities emphasize the correspondence between the zero-field quantization and the quantum Hall effect, discussed in Section 3.

c. *Hot Electron Focusing*

The voltage drop in a current-carrying sample containing a point contact in general is highly localized at or near the constriction. As mentioned in Section 4, two-terminal resistance measurements do not yield information on the voltage distribution. Four-terminal measurements could do so in principle, but such measurements in practice are hampered by the fact that conventional voltage probes located within a mean free path of the point contact are *invasive*. Indeed, in narrow multi-probe conductors, the voltage probes are the dominant source of scattering, as discussed by Timp in Chapter 3. In addition, conventional probes measure an electrochemical potential rather than an electrostatic voltage. In the linear response regime, a knowledge of the actual electric field distribution is not required to know the dissipation in the system (*cf.* Section 4). For many applications beyond linear transport, however, the electric field distribution does matter, as emphasized repeatedly by Landauer.[6,83,178,192–194] We thus are faced with a challenge to overcome the limitations of conventional resistance measurements in the ballistic regime. In this section, we discuss how an extension of the electron focusing technique of Section 6 to finite applied voltages meets this challenge at least in part.[61] (A promising alternative technique is to use a scanning tunneling microscope.)[194,195]

In Section 6, we emphasized the great difference in *length scales* between electron focusing in metals and in a 2DEG. The experiments discussed next demonstrate another qualitative difference, that of *energy scales*. In metals,[196] electrons are injected at energies above E_F, which generally are much less than $E_F \approx 5$ eV. In contrast, the Fermi energy in a 2DEG is only about 10 meV, so that dc-biasing the small ac injection voltage used in the electron

focusing experiment leads to a noticeable shift in the focusing peaks The magnitude of the shift allows a direct determination of the kinetic energy of the injected electrons, in direct analogy with a β-spectrometer In metals, shifts in the peak position in electron focusing experiments also have been observed,[37] but are attributed to the magnetic field induced by the current at large dc bias voltages [196] This field is totally negligible at the current levels used in a 2DEG

1 *Experiment* Hot electron focusing[61] spectra have been measured in a geometry identical to that used for the coherent electron focusing experiments of Section 6 (See inset of Fig 36) The electrons are injected from a 2DEG region 1 through the injector point contact, and are collected by a second point contact in region c The 2DEG region where the focusing takes place is denoted by s The point contact spacing was approximately 1 5 μm A four-terminal generalized longitudinal resistance configuration was employed, with a small ac modulation voltage of 100 μV superimposed on a dc bias voltage V_{DC} on the order of a few mV between terminals 1 and 2 The differential focusing signal $\partial V_c/\partial I_1$ is obtained by measuring the ac collector volt-

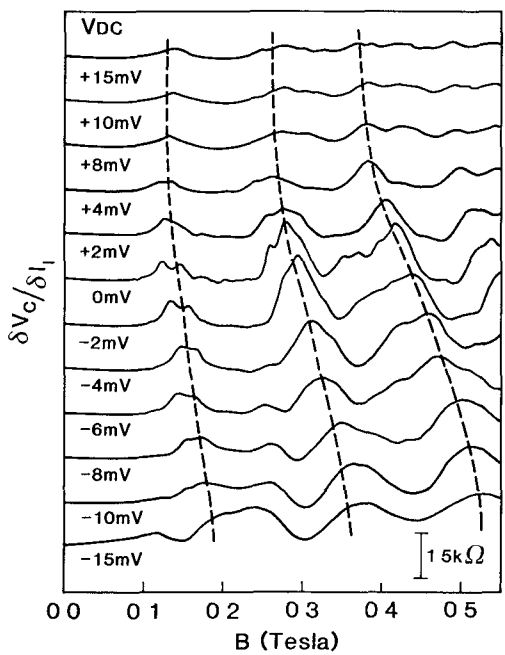

FIG 34 Differential electron focusing spectra for various applied dc bias voltages, demonstrating the effects of electron acceleration and deceleration over the point contact region The dashed lines connect peaks of the same index (From Ref 61)

age between terminals 3 and 4, and dividing by the ac current from terminal 1 to 2. In Fig. 34, the evolution of the focusing spectra with bias voltage V_{DC} is shown in the range $+15$ to -15 meV. Because of the rather large ac voltage, the fine structure due to quantum interference (studied in detail in Section 6) is smeared out, and only the classical focusing peaks remain—which are of primary interest in this experiment. A clear shift of these peaks is observed as the dc bias is increased, in particular for negative bias voltages. The peak positions are directly related to the energy of the injected electrons, as we now will demonstrate with a simple model,[61] which is based on the one used to explain the nonlinear conductance of a single-point contact,[60] discussed in Section 7.b.

ii. Model. Figure 35 illustrates the injection process at large bias voltages for four different positions of the chemical potentials μ_i and μ_s in regions i and s, relative to the subband bottom E_1 at the bottleneck of the injector

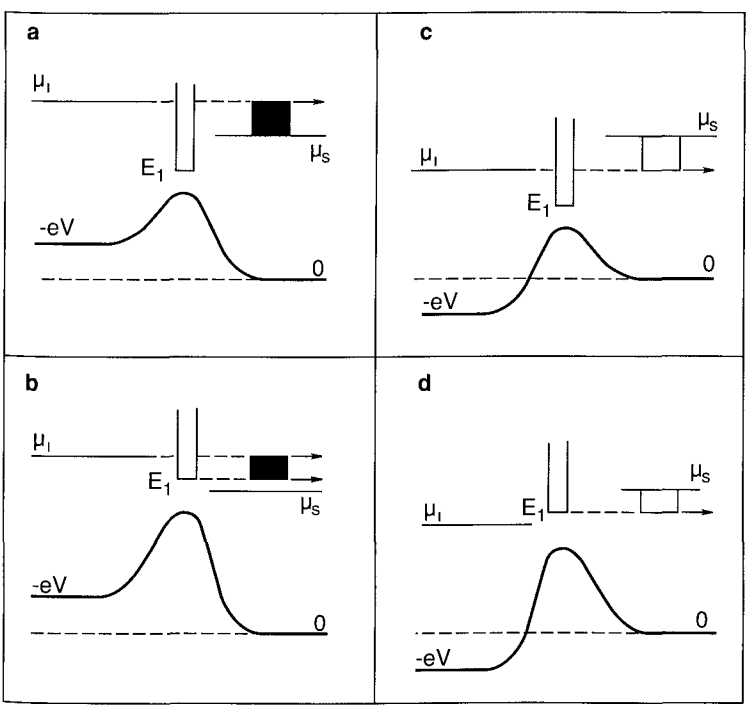

FIG. 35. Schematic representation of the injection of hot electrons (in an energy range indicated in black) or cold holes (indicated in white) from the injector at the left to the wide 2DEG region at the right. The lowest one-dimensional subband is indicated by the column. Depending on the positions of the Fermi levels μ_i and μ_s, the subband bottom E_1 can act as a low-energy cutoff for the injected carriers. The arrows denote the extremal energy of carriers detected in a differential focusing experiment. (From Ref. 61.)

point contact. The electric field caused by the applied voltage is assumed to be small outside the immediate vicinity of the point contact. In the point contact, adiabatic transport is assumed, and for simplicity only the lowest subband in the point contact is considered. (In fact, in the experiment of Fig. 34, only a single subband is occupied.) A negative voltage ($-eV > 0$) implies injection from i to s of electrons with energy in excess of E_F, as shown in panels *a* and *b* of Fig. 35. The electron flow at positive voltages is from region *s* into the injector, or equivalently, *unoccupied* electron states in region s move away from the injector (panels *c* and *d*). These states will be referred to as *holes* in the conduction band, below the local Fermi level μ_s. Peaks in the focusing spectrum also are observed for positive bias voltages (Fig. 34) resulting from focusing of ballistic holes rather than electrons.

The distribution of injected electrons or holes extends over a wide range of energies, corresponding to the height of the black and white boxes in Fig. 35. The differential measurement technique selects primarily only those electrons with maximal or minimal injection energy, as indicated by arrows in Fig. 35. One thus can study the transport of the injected *hot* electrons for $V < 0$ and of *cool* holes for $V > 0$, with an energy resolution determined by the magnitude of the ac voltage. The kinetic energy E_{focus} of the injected carriers is directly related to the position B_{focus} of the nth focusing peak if E_1 is below both μ_i and μ_s (the cases shown in Figs. 35a,c). By requiring that the point contact separation L is a multiple n of the cyclotron diameter at energy E_{focus} and magnetic field B_{focus}, one obtains the relation,

$$E_{\text{focus}} = \frac{(eLB_{\text{focus}})^2}{8mn^2}. \tag{36}$$

The two cases shown in Figs. 35c,d (where E_1 is below either μ_i or μ_s) are somewhat more complicated, because one has to take into account that the effective barrier height E_1 depends on the bias voltage. (See Sec. 7.b.) As analyzed in Ref. 61, for $V < 0$, Eq. (36) remains valid also when $E_1 < \mu_i$, but in that case, an additional small dip in the focusing spectrum is expected due to the appearance of a second extremal energy. (Note the two arrows in Fig. 35b.) This feature is not resolved clearly in Fig. 34. For $V > 0$, Eq. (36) no longer holds when $E_1 < \mu_s$ (the case of Fig. 35d). In addition, one expects a suppression of the focusing peaks for large positive bias voltages due to the cutoff imposed by the barrier in the collector. This does not play a role for negative bias voltages.

Figure 36 shows a plot of E_{focus} obtained via Eq. (36) from the position of the $n = 3$ focusing peak, as a function of the dc bias voltage.[61] For V_{DC} between -8 and $+3$ mV, E_{focus} varies linearly with V_{DC}. A linear least-squares

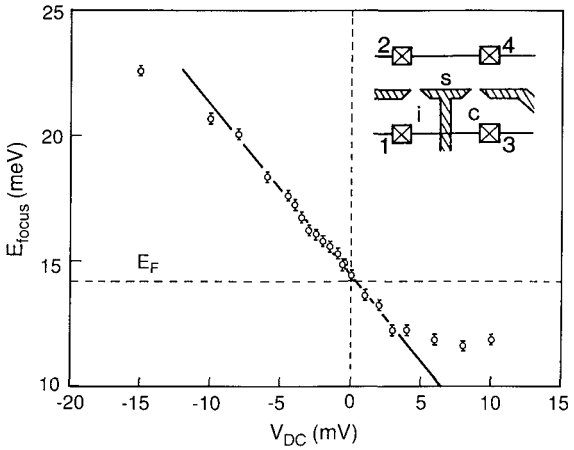

FIG. 36. Energy E_{focus} extracted from the focusing peak spacing as a function of applied dc bias voltage. The error bars shown reflect the estimated uncertainty in the measurement of the peak position, and the full straight line represents a least-squares fit. The Fermi energy E_F obtained from the Shubnikov–de Haas oscillations in the wide 2DEG region is indicated by the horizontal dashed line. The inset shows a schematic device diagram. The shaded parts indicate the gate used to define the point contacts and the 2DEG boundary, and the squares denote the ohmic contacts. (From Ref. 61.)

fit in this region yields

$$E_{focus} = -0.68eV_{DC} + 14.4 \text{ meV}. \tag{37}$$

For $V_{DC} = 0$, the electron kinetic energy E_{focus} agrees very well with $E_F = 14.2$ meV obtained from the Shubnikov–de Haas oscillations, as it should. The deviations from linearity below -8 mV and above $+3$ mV are the subject of Section 7.c.iv. We first discuss the implication of Eq. (37) for the electric field at the injector point contact.

iii. The dipole field at the injector point contact. The slope of the straight line in Fig. 36 tells us what fraction of the total dc voltage drop V_{DC} across the sample is localized in the immediate vicinity of the injector point contact. The experimental results represented by Eq. (37) indicate that this fraction is 0.68.

We now will argue that the part of the voltage drop across the sample associated with the (quantized) contact resistance nearly is completely localized at the injector point contact. (In this subsection, the term *voltage drop* refers to that part of the electrochemical potential difference that is associated with the electric field.) The total sample resistance in the experiment[61] was 19.4 ± 0.3 kΩ. Applying the voltage division rule for series resistors, one

finds that the voltage drop localized at the point contact corresponds to a fraction 0.68 of this resistance, i.e., to 13.2 ± 0.3 kΩ. This value agrees within the experimental uncertainty with the quantized resistance $h/2e^2 = 12.9$ kΩ of a quantum point contact with a single occupied one-dimensional subband. The remainder of the total sample resistance corresponds to the background discussed in Part II, which is due primarily to the ohmic contacts. It is important to realize that the observation of focusing peaks requires that the acceleration through the injector is completed on a length scale that is short compared to the cyclotron radius, which in this experiment is about 0.25 μm (for a point contact separation of 1.5 μm and a peak index $n = 3$). The fact that a correct value for the quantized point contact resistance is found using no information but the focusing peak spacing and the total sample resistance thus implies that the point contact resistance is associated with an electric field localized within this length scale. The significance of such dipole fields for transport measurements was first stressed by Landauer in the context of the residual resistivity of metals.[6] For a more recent discussion of the same topic, see Ref. 194. We stress that the spatial scale (of 0.25 μm or less) determined here for the dipole field at the point contact is about two orders of magnitude below the elastic or inelastic mean free path.

iv. Further observations. For hot electron injection at large negative bias voltages $V_{DC} < -8$ mV, the kinetic energy E_{focus} inferred from the focusing peaks (via Eq. (36)) increases more weakly with V_{DC} than at smaller biases. In addition, there is some evidence of new peaks in the focusing spectra, with positions corresponding roughly to injection of electrons with the Fermi energy. (See Fig. 34.) These two features may be indicative of a rapid energy relaxation process close to the injector point contact. Another possibility is that in this large bias regime, the voltage drop no longer is well localized at the point contact, in contrast to the case for smaller biases. We recall in this connection the absence of saturation in the experimental $I-V$ characteristics discussed in Section 7.b. It would seem that the observation of well-defined peaks in a focusing experiment precludes energy relaxation on length scales longer than the cyclotron radius as a possible explanation.

The deviation from linearity in Fig. 36 for positive injection voltages sets in for relatively small biases, $V_{DC} > +3$ mV. According to the model of Ref. 61, this deviation may arise when the subband bottom E_1 in the injector exceeds the Fermi energy μ_i in region i. In addition, the barrier in the collector point contact may impose an additional energy selection on the electron focusing signal. This will be important especially at high positive V_{DC} (but not for negative biases). Further experiments done in the regime where the injector point contact is close to pinch-off are discussed in Ref. 61.

In this device, hot electrons travel $\pi L/2 = 2.3$ μm between injector and collector. From theoretical work,[172,173] we estimate that the mean free path

for electrons with an excess energy 50% of the Fermi energy of 14 meV (which still is considerably smaller than the optical phonon energy of 36 meV) should be limited to about 400 nm as a result of electron–electron interaction effects. Such a short mean free path seems irreconcilable with the present data,[61] since it would imply a two order of magnitude reduction of the focusing peak height. An even larger discrepancy was found by Sivan *et al.*[171] in a different experiment involving two opposite point contacts. It remains a theoretical challenge to explain these unanticipated long scattering lengths observed by two independent experiments.

IV. Adiabatic Transport in the Quantum Hall Effect Regime

8. INTRODUCTION

Both the quantum Hall effect (QHE) and the quantized conductance of a ballistic point contact are described by one and the same relation, $G = (2e^2/h)N$, between the conductance G and the number N of propagating modes at the Fermi level. The smooth transition from zero-field quantization to QHE that follows from this relation was the subject of Section 3. The resemblance between ballistic quantum transport and transport in the QHE regime becomes superficial, however, if one considers the entirely different role of scattering processes in weak and strong magnetic fields. First of all, the zero-field conductance quantization is destroyed by a small amount of elastic scattering (due to impurities or roughness of the channel boundaries; *cf.* Section 2.d), while the QHE is not. This difference is a manifestation of the suppression of backscattering by a magnetic field, discussed in Section 4. Absence of backscattering by itself does not imply adiabatic transport, which requires a total suppression of scattering among the modes. In weak magnetic fields, adiabaticity is of importance within a point contact, but not on longer length scales (*cf.* Sections 2.b.iii and 5). In the wide 2DEG region, scattering among the modes in weak fields establishes local equilibrium on a length scale given by the inelastic scattering length (which, in a high-mobility material, presumably is not much longer than the elastic scattering length $l \sim 10$ μm). The situation is strikingly different in a strong magnetic field, where the *selective* population and detection of the modes at the Fermi level demonstrated in Ref. 10 (to be discussed in Section 9.b) is made possible by the persistence of adiabaticity outside the point contact. In the words of Ref. 197, application of a magnetic field induces a transition from a *local* to a *global* adiabatic regime. Over some longer distance (which is not yet known precisely), adiabaticity breaks down, but surprisingly enough, local equilibrium remains absent even on macroscopic length scales (exceeding 0.25 mm).[11,12,34,198–200] Since local equilibrium is a prerequisite for the use

of a local resistivity tensor, these findings imply a non-locality of the transport that had not been anticipated in theories of the QHE (which commonly are expressed in terms of a local resistivity; *cf.* Ref. 64 for a review). An important exception, to which we will return, is Büttiker's theory[77] on the role of contacts in the establishment of the quantized Hall resistance.

In the QHE regime, the propagating modes at the Fermi level are located at the edges of the sample,[201,202] under circumstances such that in the bulk, the Fermi level lies in a band of localized states.[62,203] These edge states originate from Landau levels that in the bulk lie below the Fermi level, but rise in energy on approaching the sample boundary. The point of intersection of the nth Landau level ($n = 1, 2, \ldots$) with the Fermi level forms the site of edge states belonging to the nth edge *channel*. The energy of each state can be separated into a part $(n - \frac{1}{2})\hbar\omega_c$ due to the quantized cyclotron motion, a part E_G due to the potential energy in an electrostatic potential $V(x, y)$, and the Zeeman energy $\pm\frac{1}{2}g\mu_B B$ depending on the spin direction. For a given total Fermi energy E_F, one has

$$E_G = E_F - (n - \tfrac{1}{2})\hbar\omega_c \pm \tfrac{1}{2}g\mu_B B. \tag{38}$$

Adiabatic transport is motion with constant n, implying that the cyclotron orbit center is guided along contours of constant $V(x, y) = E_G$. The energy E_G of this equipotential is referred to as the *guiding center energy*.

The simplicity of this guiding center drift along equipotentials has been exploited in the percolation theory[204,205] of the QHE, soon after its experimental discovery.[62] The physical requirements for the absence of inter-Landau-level scattering have received considerable attention[206,207] in that context, and more recently[197,208,209] in the context of adiabatic transport in edge channels. One requirement is that strong potential variations occur on a spatial scale that is large compared to the magnetic length $l_m \equiv (\hbar/eB)^{1/2}$,which corresponds to the cyclotron radius in the QHE ($l_{cycl} \equiv l_m(2n - 1)^{1/2} \approx l_m$ if the Landau level index $n \approx 1$). More rapid potential fluctuations may be present provided their amplitude is much less than $\hbar\omega_c$ (the energy separation of Landau levels).

Because edge channels at opposite edges of the sample move in opposite directions (as the drift velocity along an equipotential is given by $\mathbf{E} \times \mathbf{B}/B^2$, with $\mathbf{E}$ the electric field), backscattering requires scattering from one edge to the other. *Selective backscattering* for all edge channels with $n \geq n_0$ is imposed by a potential barrier[133–135] across the sample if its height is between the guiding center energies of edge channel n_0 and $n_0 - 1$. (Note that the edge channel with a larger index n has a smaller value of E_G.) The anomalous Shubnikov–de Haas effect,[11] which we will discuss in Section 10.a, has demonstrated that selective backscattering also can occur *naturally* in the

absence of an imposed potential barrier. The edge channel with the highest index $n = N$ is selectively backscattered when the Fermi level approaches the energy $(N - \frac{1}{2})\hbar\omega_c$ of the Nth bulk Landau level (disregarding the Zeeman energy for simplicity of notation). The guiding center energy of the Nth edge channel then approaches zero, and backscattering either by tunneling or by thermally activated proccesses becomes effective—but only for that edge channel, which remains almost completely decoupled from the other $N - 1$ edge channels over distances as large as 250 μm. It was believed initially[11] that the transport might be fully adiabatic over this macroscopic length scale. However, Alphenaar *et al.*[200] now have demonstrated experimentally that on this length scale, the edge channels wth $n \leq N - 1$ have equilibrated to a large extent. The absence of scattering was found to persist only between this group of edge channels and the Nth edge channel, and then only if the Fermi level lies in (or near) the Nth bulk Landau level. As a qualitative explanation of these observations, it was proposed[200,210] that states from the highest-index edge channel hybridize with the localized states from the bulk Landau level of the same index when both types of states coexist at the Fermi level. Such a coexistence does not occur for the lower-index edge channels. A complete theoretical description is not available, and our present understanding of how fully adiabatic transport breaks down (as well as of the length scale on which this occurs) remains incomplete.

To avoid misunderstanding, it should be emphasized that the fact that the measured resistance can be expressed in terms of the transmission probabilities of edge states at the Fermi level does *not* imply that these few states carry a macroscopic current, *nor* does it imply that the current flows at the edges. A determination of the spatial current distribution, rather than just the total current, requires consideration of all the states below the Fermi level, which acquire a net drift velocity because of the Hall field. Within the range of validity of a linear response theory,[79] however, knowledge of the current distribution is not necessary to know the resistances. (See Ref. 18 for a more extensive discussion of this issue, which has caused some confusion in the literature.)

The outline of Part IV is as follows. Section 9 deals with anomalies in the quantum Hall effect due to the absence of local equilibrium at the current and voltage contacts.[10,34,198–200] The ideality of the contacts then affects the QHE in a fundamental way,[77,211] in contrast to the weak-field case where non-ideal contacts lead only to an uninteresting additive contact resistance (*cf.* Part II). An *ideal* contact in the QHE is one that establishes an equilibrium population among the outgoing edge channels by distributing the injected current equally among these propagating modes (*cf.* Section 2.b.ii). The selective population of edge channels by quantum point contacts[10] is discussed as an extreme example of a non-equilibrium population. Selective

backscattering within a single edge channel, and the resulting anomalies in the Shubnikov–de Haas effect,[11,12] are the subjects of Section 10. In Section 11, inter-edge channel tunneling in a quantum point contact is discussed, which can explain the unusual observation of Aharonov–Bohm magnetoresistance oscillations in a *singly connected* geometry.[16] By combining two such point contacts to form a disc-shaped cavity, it has been possible to study this effect in a highly controlled way.[17]

In this chapter, we restrict ourselves to the *integer* QHE, where the edge channels can be described by single-electron states. Recent theoretical[212] and experimental[213–216] work on adiabatic transport in the *fractional* QHE (which fundamentally is a many-body effect)[64,217,218] indicates that many of the phenomena discussed here have analogues in that context as well.

9. Anomalous Quantum Hall Effect

a. Ideal versus Disordered Contacts

The quantization of the Hall resistance was discovered in a four-terminal measurement.[62] Under conditions in which the Hall resistance is quantized, the longitudinal resistance vanishes. Since the two-terminal resistance is the sum of the Hall and longitudinal resistances, a two-terminal measurement also shows a quantized resistance in a strong magnetic field[115–117] (to within an experimental uncertainty on the order of one part in 10^6). Nevertheless, investigators interested in high-precision determinations of the quantized Hall resistance generally have preferred a four-terminal measurement,[219] under the assumption that one thereby eliminates the contact resistances (*cf.* Part II). That is correct if local equilibrium is established near the contacts. A surprising conclusion of the work described in this section and the next is that local equilibrium is *not* the rule in the QHE regime, and that the effects of contacts can persist on macroscopic-length scales. Four-terminal measurements then in general will not yield a more accurate determination of the quantized Hall resistance than a two-terminal measurement does.

A necessary condition for the accurate quantization of the two-terminal Hall resistance is that the source and drain contacts are ideal,[77,211] in the sense that the edge states at the Fermi level have unit transmission probability through the contacts. This condition also plays a central role in four-terminal geometries. However, in the latter case, the requirement of ideal contacts is not necessary if the edge channels close to the contacts are equilibrated (as a result of inelastic scattering).

i. Ideal contacts. We return to the four-terminal measurements on a quantum point contact considered in Section 4, but now in the QHE regime,

where the earlier assumption of local equilibrium near the contacts no longer is applicable in general. We assume strong magnetic fields, so that the four-terminal longitudinal resistance R_L of the quantum point contact is determined by the potential barrier in the constriction. Before we present the experimental results, we first describe briefly how the Landauer–Büttiker formalism[7] can be applied to multi-probe measurements in the QHE regime.[77] For a more detailed discussion of this formalism, we refer to Chapter 4 by Büttiker.

Consider the geometry of Fig. 37. Reservoirs at chemical potential μ_α are connected by leads to the conductor. The current I_α in the lead to reservoir α is related to these chemical potentials via the transmission probabilities $T_{\alpha\to\beta}$ (from reservoir α to β), and reflection probabilities R_α (from reservoir α back to the same reservoir). These equations have the form,[7]

$$\frac{h}{2e} I_\alpha = (N_\alpha - R_\alpha)\mu_\alpha - \sum_{\beta\neq\alpha} T_{\beta\to\alpha}\mu_\beta, \tag{39}$$

with N_α the number of propagating modes (or *quantum channels*) in lead α. In a strong magnetic field, N_α equals the number of edge channels at the Fermi energy E_F in lead α (which is the same as the number of bulk Landau levels below E_F in view of the one-to-one correspondence between edge channels and bulk Landau levels discussed in Section 8). As before, we denote the number of edge channels in the wide 2DEG and in the constriction by N_{wide} and N_{min}, respectively.

An *ideal* contact to the wide 2DEG has the property that all N_{wide} edge channels are fully transmitted into the contact, where they equilibrate.[77,211] Such a contact thus has $N_\alpha - R_\alpha = N_{\text{wide}}$. The constriction transmits only N_{min} channels, and the remaining $N_{\text{wide}} - N_{min}$ channels are reflected back

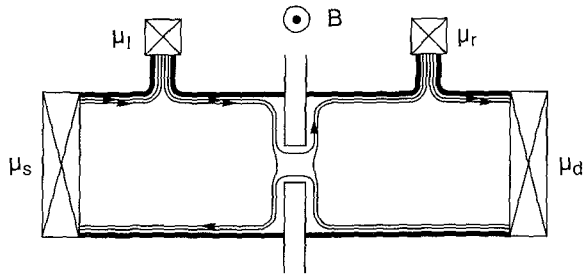

FIG. 37. Guiding center motion along equipotentials in the QHE regime, in a four-terminal geometry with a saddle-shaped potential formed by a split gate. Ideal contacts are assumed. The thin lines indicate the location of the edge channels at the Fermi level, with the arrows pointing in the direction of motion of edge channels that are populated by the contacts (crossed squares).

along the opposite 2DEG boundary (*cf* Fig 37) We denote by μ_l and μ_r the chemical potentials of adjacent voltage probes to the left and to the right of the constriction The current source is at μ_s, and the drain at μ_d Applying Eq (39) to this case, using $I_s = -I_d \equiv I$, $I_r = I_l = 0$, one finds for the magnetic field direction indicated in Fig 37,

$$\frac{h}{2e} I = N_{wide}\mu_s - (N_{wide} - N_{min})\mu_l, \tag{40a}$$

$$0 = N_{wide}\mu_l - N_{wide}\mu_s, \tag{40b}$$

$$0 = N_{wide}\mu_r - N_{min}\mu_l \tag{40c}$$

We have used the freedom to choose the zero level of chemical potential by fixing $\mu_d = 0$, so that we have three independent (rather than four dependent) equations The four-terminal longitudinal resistance $R_L \equiv (\mu_l - \mu_r)/eI$ that follows from Eq (40) is

$$R_L = \frac{h}{2e^2}\left(\frac{1}{N_{min}} - \frac{1}{N_{wide}}\right) \tag{41}$$

In reversed-field direction, the same result is obtained Equation (41), derived for ideal contacts without assuming local equilibrium near the contacts, is identical to Eq (10) Similarly, applying Eq (39) to a six-terminal geometry, one recovers the results from Eqs (12) and (13) for the four-terminal Hall and diagonal resistances

The fundamental reason that the local equilibrium approach of Section 4 (appropriate for weak magnetic fields) and the ideal contact approach of this section (for strong fields) yield identical answers is that an ideal contact attached to the wide 2DEG regions *induces* a local equilibrium by equipartitioning the outgoing current among the edge channels [77] (This is illustrated in Fig 37, where the current entering the voltage probe to the right of the constriction is carried by a *single* edge channel, while the equally large current flowing out of that probe is distributed equally over the *two* edge channels available for transport in the wide region) In weaker magnetic fields, when the cyclotron radius exceeds the width of the narrow 2DEG region connecting the voltage probe to the Hall bar, not all edge channels in the wide 2DEG region are transmitted into the voltage probe, which therefore is not effective in redistributing the current This is the reason that the weak-field analysis in Section 4 required the assumption of a local equilibrium in the wide 2DEG near the contacts

We now discuss some experimental results, which confirm the behavior predicted by Eq (41) in the QHE regime, to complement the weak-field ex-

periments discussed in Section 4. In this field regime, the split-gate point contact geometry of Fig. 37 essentially is equivalent to a geometry studied recently by several authors,[133–135,198,199] in which a potential barrier across the Hall bar is created by means of a narrow continuous gate. The quantization of the longitudinal conductance R_L^{-1} in *fractions* of $2e^2/h$ (for unresolved spin degeneracy), predicted by Eq. (41), is shown in Fig. 38 for a quantum point contact sample[125] at $T = 0.6$ K. (Similar point contact data is reported in Ref. 126.) The magnetic field is kept fixed at 1.4 T (such that $N_{\text{wide}} = 5$) and the gate voltage is varied (such that N_{min} ranges from 1 to 4). Conductance plateaus close to 5/4, 10/3, 15/2, and $20 \times (2e^2/h)$ (solid horizontal lines) are observed, in accord with Eq. (41). Spin-split plateaus (dashed lines) are barely resolved at this rather low magnetic field. Observations of such a *fractional* quantization due to the integer QHE were made before on wide Hall bars with regions of different electron density in series,[127,129] but the theoretical explanation[130] given at that time was less straightforward than Eq. (41).

ii. Disordered contacts. The validity of Eq. (41) in the QHE regime breaks down for non-ideal contacts, if local equilibrium near the contacts is not established. As discussed in Section 4, Eq. (41) implies that the Hall voltage over the wide 2DEG regions adjacent to the constriction (*cf.* Fig. 18) is

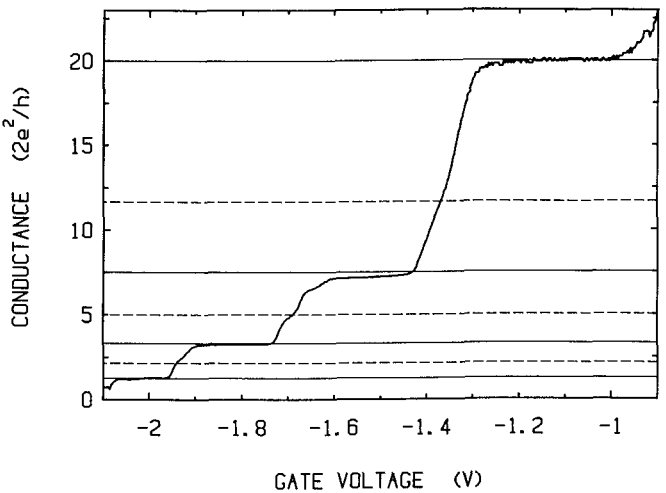

FIG. 38. *Fractional* quantization of the four-terminal longitudinal conductance R_L^{-1} of a point contact in a magnetic field of 1.4 T at $T = 0.6$ K. The solid horizontal lines indicate the quantized plateaus predicted by Eq. (41), with $N_{\text{wide}} = 5$ and $N_{\text{min}} = 1, 2, 3, 4$. The dashed lines give the location of the spin-split plateaus, which are not well-resolved at this magnetic field value. (From Ref. 125.)

unaffected by the presence of the constriction. In Fig. 39, we show the four-terminal longitudinal resistance R_L and Hall resistance R_H obtained on the sample considered in Section 4, but now over a wider field range, for both a small gate voltage (-0.3 V) and a large gate voltage (-2.5 V). In addition to the weak-field negative magnetoresistance discussed before in Section 4, a crossover to a positive magnetoresistance with superimposed Shubnikov–de Haas oscillations is seen.

The data for $V_g = -0.3$ V exhibits Shubnikov–de Haas oscillations with zero minima in the longitudinal resistance R_L, and the normal quantum Hall resistance $R_H = (h/2e^2)N_{wide}^{-1}$ is determined by the number of Landau levels occupied in the wide regions. (N_{wide} can be obtained from the quantum Hall effect measured in the absence of the constriction, or from the periodicity of the Shubnikov–de Haas oscillations.)

At the higher gate voltage $V_g = -2.5$ V, non-vanishing minima in R_L are seen in Fig. 39, as a result of the formation of a potential barrier in the constriction. At the minima, R_L has the fractional quantization predicted by Eq. (41) (See earlier.) For example, the plateau in R_L around 2.2 T for $V_g = -2.5$ V is observed to be at $R_L = 2.1\ \text{k}\Omega \approx (h/2e^2) \times (\frac{1}{2} - \frac{1}{3})$, in agreement with the fact that the two-terminal resistance yields $N_{min} = 2$, and the number of Landau levels in the wide regions $N_{wide} = 3$. In spite of this agreement, and in apparent conflict with the analysis leading to Eq. (41), it is seen in Fig. 39 that the Hall resistance R_H measured across a wide region for $V_g = -2.5$ V has *increased* over its value for small gate voltages. Indeed,

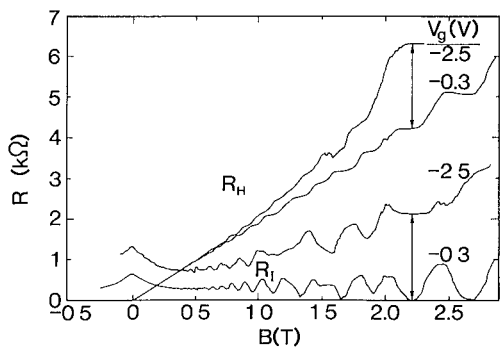

FIG 39 Non-vanishing Shubnikov–de Haas minima in the longitudinal resistance R_L and anomalous quantum Hall resistance R_H, measured in the point contact geometry of Fig. 18 at 50 mK. These experimental results are extensions to higher fields of the weak-field traces shown in Fig 15 The Hall resistance has been measured across the wide region, more than 100 μm away from the constriction, yet R_H is seen to increase if the gate voltage is raised from -0.3 V to -2.5 V The magnitude at $B = 2.2$ T of the deviation in R_H and the Shubnikov–de Haas minimum in R_L are indicated by arrows, which for both R_H and R_L have a length of $(h/2e^2)(\frac{1}{2} - \frac{1}{3})$, in agreement with the analysis given in the text (From Ref 34)

around 2.2 T, a Hall plateau at $R_H = 6.3\ k\Omega \approx (h/2e^2) \times \frac{1}{2}$ is found, as if the number of occupied Landau levels was given by $N_{min} = 2$ rather than by $N_{wide} = 3$. This unexpected deviation was noted in Ref. 34, but was not understood at the time. The temperature dependence of this effect has not been studied systematically, but it was found that the deviations in the Hall resistance persist at least up to 1.6 K. At higher magnetic fields (not shown in Fig. 39), the $N = 1$ plateau is reached, and the deviation in the Hall resistance vanishes. Following the similar experiment by Komiyama *et al.*,[198,199] and the demonstration of the anomalous QHE measured with quantum point contacts by the Delft–Philips collaboration,[10] the likely explanation of the data of Fig. 39 is that one or more of the ohmic contacts used as voltage probes are *disordered*, in the sense of Büttiker[77] that not all edge channels have unit transmission probability into the voltage probe. (Note that a point contact containing a potential barrier also is disordered in this sense.)

We now will demonstrate, following Refs. 77 and 211, how the anomalies in Fig. 39 can be accounted for nicely if one assumes that one of the probes used to measure the Hall voltage is disordered, because of a potential barrier in the probe with a height not below that of the barrier in the constriction. This is illustrated schematically in Fig. 40. A net current I flows through

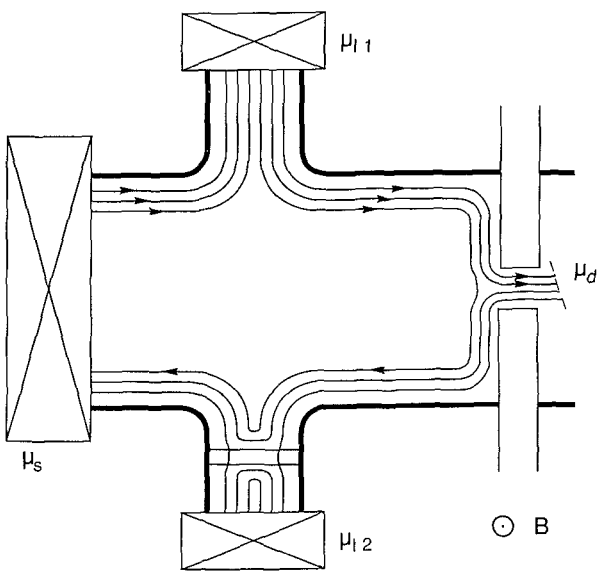

FIG. 40. Illustration of the flow of edge channels along equipotentials in a sample with a constriction (defined by the split gates) and a disordered voltage probe. (A potential barrier in the lower probe is indicated by the double bar.)

the constriction, determined by its two-terminal resistance according to $I = (2e/h)N_{\min}\mu_s$, with μ_s the chemical potential of the source reservoir (The chemical potential of the drain reservoir μ_d is taken as a zero reference) Equation (39), applied to the two opposite Hall probes l_1 and l_2 in Fig 40, takes the form (using $I_{l_1} = I_{l_2} = 0$, $\mu_s = (h/2e)I/N_{\min}$, and $\mu_d = 0$),

$$0 = N_{\text{wide}}\mu_{l_1} - T_{s\to l_1}\frac{h}{2e}\frac{I}{N_{\min}} - T_{l_2\to l_1}\mu_{l_2}, \tag{42a}$$

$$0 = N_{l_2}\mu_{l_2} - T_{s\to l_2}\frac{h}{2e}\frac{I}{N_{\min}} - T_{l_1\to l_2}\mu_{l_1}, \tag{42b}$$

where we have assumed that the disordered Hall probe l_2 transmits only $N_{l_2} < N_{\text{wide}}$ edge channels because of some barrier in the probe For the field direction shown in Fig 40, one has, under the assumption of no inter-edge channel scattering from constriction to probe l_2, $T_{s\to l_1} = N_{\text{wide}}$, $T_{s\to l_2} = T_{l_2\to l_1} = 0$, and $T_{l_1\to l_2} = \max(0, N_{l_2} - N_{\min})$ Equation (42) then leads to a Hall resistance $R_H \equiv (\mu_{l_1} - \mu_{l_2})/eI$ given by

$$R_H = \frac{h}{2e^2}\frac{1}{\max(N_{l_2}, N_{\min})} \tag{43}$$

In the opposite field direction, the normal Hall resistance $R_H = (h/2e^2)N_{\text{wide}}^{-1}$ is recovered

The assumption of a single disordered probe, plus absence of inter-edge channel scattering from constriction to probe, thus explains the observation in Fig 39 of an anomalously high quantum Hall resistance for large gate voltages, such that $N_{\min} < N_{\text{wide}}$ Indeed, the experimental Hall resistance for $V_g = -2\,5$ V has a plateau around 2 2 T, close to the value $R_H = (h/2e^2)N_{\min}^{-1}$ (with $N_{\min} = 2$), in agreement with Eq (43) if $N_{l_2} \le N_{\min}$ at this gate voltage This observation demonstrates the absence of inter-edge channel scattering over 100 μm (the separation of constriction and probe)—but only between the highest-index edge channel (with index $n = N_{\text{wide}} = 3$) and the two lower-index channels Since the $n = 1$ and $n = 2$ edge channels either are both empty or both filled (*cf* Fig 40, where these two-edge channels lie closest to the sample boundary), any scattering between $n = 1$ and 2 would have no measurable effect on the resistances As discussed in Section 8, we now know from the work of Alphenaar *et al* [200] that (at least in the present samples) the edge channels with $n \le N_{\text{wide}} - 1$, in fact, do equilibrate to a large extent on a length scale of 100 μm

In the absence of a constriction, or at small gate voltages (where the constriction is just defined), one has $N_{\min} = N_{\text{wide}}$ so that the normal Hall effect

is observed in both field directions. This is the situation realized in the experimental trace for $V_g = -0.3$ V in Fig. 39. In very strong fields such that $N_{min} = N_{l_2} = N_{wide} = 1$ (still assuming non-resolved spin-splitting), the normal result $R_H = h/2e^2$ would follow even if the contacts contain a potential barrier, in agreement with the experiment (not shown in Fig. 39). This is a more general result, which also holds for a barrier that only partially transmits the $n = 1$ edge channel.[12,77,220–222]

A similar analysis as the preceding predicts that the longitudinal resistance measured on the edge of the sample that contains ideal contacts retains its regular value, as in Eq. (41). The observation in the experiment of Fig. 39 for $V_g = -2.5$ V of a regular longitudinal resistance (in agreement with Eq. (41)), along with an anomalous quantum Hall resistance, thus is consistent with this analysis. On the opposite sample edge, the measurement would involve the disordered contact, and one finds instead

$$R_L = \frac{h}{2e^2}\left(\frac{1}{N_{min}} - \frac{1}{\max(N_{l_2}, N_{min})}\right) \tag{44}$$

The experiment[10] discussed in the following subsection is topologically equivalent to the geometry of Fig. 40, but involves quantum point contacts rather than ohmic contacts. This gives the possibility to populate and detect edge channels selectively, thereby enabling a study of the effects of a non-equilibrium population of edge channels in a controlled manner.

b. *Selective Population and Detection of Edge Channels*

In Section 6, we have seen in the coherent electron focusing experiment how a quantum point contact can inject a *coherent* superposition of edge channels at the 2DEG boundary. In that section, we restricted ourselves to weak magnetic fields. Here, we will show how in the QHE regime, the point contacts can be operated in a different way as *selective* injectors (and detectors) of edge channels.[10] We recall that electron focusing can be measured as a generalized Hall resistance, in which case the pronounced peaked structure due to mode interference is superimposed on the weak-field Hall resistance (*cf.* Fig. 25). If the weak-field electron focusing experiments are extended to stronger magnetic fields, a transition to the quantum Hall effect is observed, provided the injecting and detecting point contacts are not pinched off too strongly.[13] The oscillations characteristic of mode interference disappear in this field regime, suggesting that the coupling of the edge channels (which form the propagating modes from injector to collector) is suppressed and adiabatic transport is realized. It now is sufficient no longer

to model the point contacts by a point source/detector of infinitesimal width (as was done in Section 6), but a somewhat more detailed description of the electrostatic potential $V(x, y)$ defining the point contacts and the 2DEG boundary between them is required Schematically, $V(x, y)$ is represented in Fig 41 Fringing fields from the split gate create a potential barrier in the point contacts, so that V has a saddle form as shown The heights of the barriers E_1, E_c in the injector and collector are adjustable separately by means of the voltages on the split gates, and can be determined from the two-terminal conductances of the individual point contacts The point contact separation in the experiment of Ref 10 is small (1 5 μm), so that we can assume fully adiabatic transport from injector to collector in strong magnetic fields (The experiment has been repeated for much larger point contact separations (60 and 130 μm) by Alphenaar *et al*,[200] who find only partial absence of inter-edge channel scattering over these distances, See Section 8) As discussed in Section 8, the adiabatic transport is along equipotentials at the guiding center energy E_G Note that the edge channel with the smallest index n has the largest guiding center energy (according to Eq (38)) In the absence of inter-edge channel scattering, edge channels only can be transmitted through a point contact if E_G exceeds the potential barrier height (disregarding tunneling through the barrier) The injector thus injects

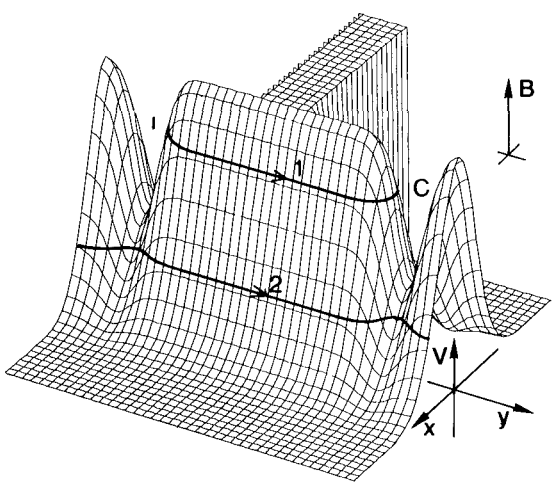

FIG 41 Schematic potential landscape showing the 2DEG boundary and the saddle-shaped injector and collector point contacts In a strong magnetic field, the edge channels are extended along equipotentials at the guiding center energy, as indicated here for edge channels with index $n = 1, 2$ (The arrows point in the direction of motion) In this case, a Hall conductance of $(2e^2/h)N$ with $N = 1$ would be measured by the point contacts—in spite of the presence of two occupied spin-degenerate Landau levels in the bulk 2DEG (From Ref 156)

$N_i - R_i \approx (E_F - E_i)/\hbar\omega_c$ edge channels into the 2DEG, while the collector is capable of detecting $N_c - R_c \approx (E_F - E_c/\hbar\omega_c$ channels. Along the boundary of the 2DEG, however, a larger number of $N_{wide} \approx E_F/\hbar\omega_c$ edge channels, equal to the number of occupied bulk Landau levels in the 2DEG, are available for transport at the Fermi level. The selective population and detection of Landau levels leads to deviations from the normal Hall resistance.

These considerations can be put on a theoretical basis by applying the Landauer–Büttiker formalism discussed in Section 9.a to the electron focusing geometry.[15] We consider a three-terminal conductor as shown in Fig. 2a, with point contacts in two of the probes (injector i and collector c), and a wide ideal drain contact d. The collector acts as a voltage probe, drawing no net current, so that $I_c = 0$ and $I_d = -I_i$. The zero of energy is chosen such that $\mu_d = 0$. One then finds from Eq. (39) the two equations,

$$0 = (N_c - R_c)\mu_c - T_{i\to c}\mu_i, \tag{45a}$$

$$\frac{h}{2e} I_i = (N_i - R_i)\mu_i - T_{c\to i}\mu_c, \tag{45b}$$

and obtains for the ratio of collector voltage $V_c = \mu_c/e$ (measured relative to the voltage of the current drain) to injected current I_i the result,

$$\frac{V_c}{I_i} = \frac{2e^2}{h} \frac{T_{i\to c}}{G_i G_c - \delta}. \tag{46}$$

Here, $\delta \equiv (2e^2/h)^2 T_{i\to c}\, T_{c\to i}$, and $G_i \equiv (2e^2/h)(N_i - R_i)$, $G_c \equiv (2e^2/h)\ (N_c - R_c)$ denote the conductances of injector and collector point contact. The injector–collector reciprocity demonstrated in Fig. 25 is present manifestly in Eq. (46), since G_i and G_c are even functions of B while[7] $T_{i\to c}(B) = T_{c\to i}(-B)$.

For the magnetic field direction indicated in Fig. 2a, the term δ in Eq. (46) can be neglected, since $T_{c\to i} \approx 0$. An additional simplification is possible in the adiabatic transport regime. We consider the case that the barrier in one of the two point contacts is sufficiently higher than in the other, to ensure that electrons are transmitted over the highest barrier will have a negligible probability of being reflected at the lowest barrier. Then $T_{i\to c}$ is dominated by the transmission probability over the highest barrier, $T_{i\to c} \approx \min(N_i - R_i, N_c - R_c)$. Substitution in Eq. (46) gives the remarkable result[10] that the *Hall conductance* $G_H \equiv I_i/V_c$ measured in the electron focusing geometry can be expressed entirely in terms of the *contact conductances* G_i and G_c,

$$G_H \approx \max(G_i, G_c). \tag{47}$$

Equation (47) tells us that quantized values of G_H occur not at $(2e^2/h)N_{wide}$, as one would expect from the N_{wide} populated Landau levels in the 2DEG—but at the smaller value of $(2e^2/h)\max(N_{min,i}, N_{min,c})$. As shown in Fig. 42, this indeed is observed experimentally.[10] Notice in particular how any deviation from quantization in $\max(G_i, G_c)$ is reproduced faithfully in G_H, in complete agreement with Eq. (47).

10. Anomalous Shubnikov–de Haas Effect

Shubnikov–de Haas oscillations periodic in $1/B$ occur in the longitudinal resistance of a 2DEG at low temperatures ($k_B T < \hbar\omega_c$), provided the mobility μ (or scattering time τ) is large enough to allow the formation of Landau levels ($\mu B \equiv \omega_c \tau > 1$, so that many cyclotron orbits occur on average between scattering events). In weak magnetic fields, where a theoretical description in

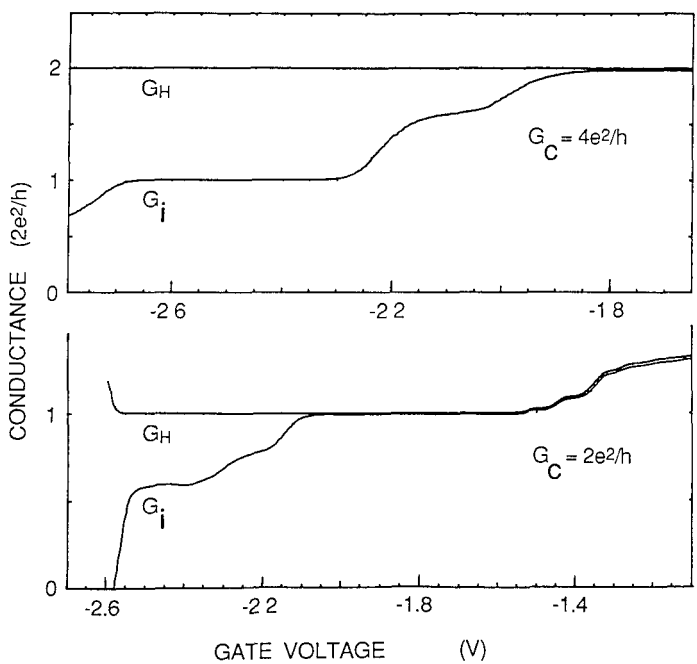

FIG. 42. Experimental correlation between the conductances G_i, G_c of injector and collector, and the Hall conductance $G_H \equiv I_i/V_c$, shown to demonstrate the validity of Eq. (47) ($T = 1.3$ K; point contact separation is 1.5 μm). The magnetic field was kept fixed (top: $B = 2.5$ T, bottom: $B = 3.8$ T, corresponding to a number of occupied bulk Landau levels $N = 3$ and 2, respectively) By increasing the gate voltage on one half of the split gate defining the injector, G_i was varied at constant G_c. (From Ref. 10.)

terms of a local resistivity tensor is meaningful, a satisfactory agreement between theory and experiment is obtained.[2] In the strong magnetic field regime of interest here, one has the complication that the concept of a local resistivity tensor may break down entirely because of the absence of local equilibrium. A theory of the Shubnikov–de Haas effect then has to take into account explicitly the properties of the contacts used for the measurement. The resulting anomalies seem not to have been anticipated in the theoretical literature. In Section 10.a, we discuss the anomalous Shubnikov–de Haas oscillations[11,12] in a geometry with a quantum point contact as a selective edge channel detector, which demonstrates the selective backscattering within a single Landau level and the absence of local equilibrium on a length scale of 250 μm. More general consequences of the absence of local equilibrium for the Shubnikov–de Haas effect are the subject of Section 10.b.

a. *Selective Backscattering*

To discuss the anomalous Shubnikov–de Haas effect, we consider the three-terminal geometry of Fig. 43, where a single voltage contact is present on the boundary between source and drain contacts. (An alternative two-terminal measurement configuration also is possible; see Ref. 11.) The voltage probe p is formed by a quantum point contact, while source s and drain d are normal ohmic contacts. (Note that *two* special contacts were required for the anomalous quantum Hall effect of Section 9.b.) One finds directly from Eq. (39) that the three-terminal resistance $R_{3t} \equiv (\mu_p - \mu_d)/eI$ measured between point

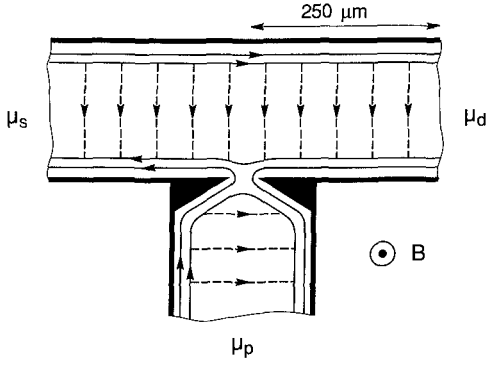

FIG. 43. *Illustration of the mechanism for the suppression of Shubnikov–de Haas oscillations due to selective detection of edge channels. The black area denotes the split gate point contact in the voltage probe, which is at a distance of 250 μm from the drain reservoir. Dashed arrows indicate symbolically the selective backscattering in the highest-index edge channel, via states in the highest-bulk Landau level that coexist at the Fermi level.*

contact probe and drain is given by

$$R_{3t} = \frac{h}{2e^2} \frac{T_{s\to p}}{(N_s - R_s)(N_p - R_p) - T_{p\to s}T_{s\to p}} \tag{48}$$

This three-terminal resistance corresponds to a generalized *longitudinal* resistance if the magnetic field has the direction of Fig 43 In the absence of backscattering in the 2DEG, one has $T_{s\to p} = 0$, so that R_{3t} vanishes, as it should for a longitudinal resistance in a strong magnetic field

Shubnikov–de Haas oscillations in the longitudinal resistance arise when backscattering leads to $T_{s\to p} \neq 0$ The resistance reaches a maximum when the Fermi level lies in a bulk Landau level, corresponding to a maximum probability for backscattering (which requires scattering from one edge to the other across the bulk of the sample, as indicated by the dashed lines in Fig 43) From the preceding discussion of the anomalous quantum Hall effect, we know that the point contact voltage probe in a high magnetic field functions as a selective detector of edge channels with index n less than some value determined by the barrier height in the point contact If, as discussed in Section 8, backscattering itself occurs selectively for the channel with the highest index $n = N_{wide}$, and if the edge channels with $n \leq N_{wide} - 1$ do not scatter to that edge channel, then a suppression of the Shubnikov–de Haas oscillations is to be expected when R_{3t} is measured with a point contact containing a sufficiently high potential barrier This indeed was observed experimentally,[11] as is shown in Fig 44 The Shubnikov–de Haas maximum at 5 2 T, for example, is found to disappear at gate voltages such that the point contact conductance is equal to or smaller than $2e^2/h$, which means that the point contact only transmits two spin-split edge channels The number of occupied spin-split Landau levels in the bulk at this magnetic field value is three This experiment thus demonstrates that the Shubnikov–de Haas oscillations result from the highest-index edge channel only, and that this edge channel does not scatter to the lower-index edge channels over the distance of 250 μm from point contact probe to drain

In Section 9 a, we discussed how an *ideal* contact at the 2DEG boundary *induces* a local equilibrium by equipartitioning the outgoing current equally among the edge channels The anomalous Shubnikov–de Haas effect provides a direct way to study this contact-induced equilibration by means of a second point contact between the point contact voltage probe p and the current drain d in Fig 43 This experiment is described in Ref 12 Once again, use was made of the double-split-gate point contact device (Fig 2), in this case with a 1 5 μm separation between point contact p and the second point contact It is found that the Shubnikov–de Haas oscillations in R_{3t} are suppressed only if the second point contact has a conductance of

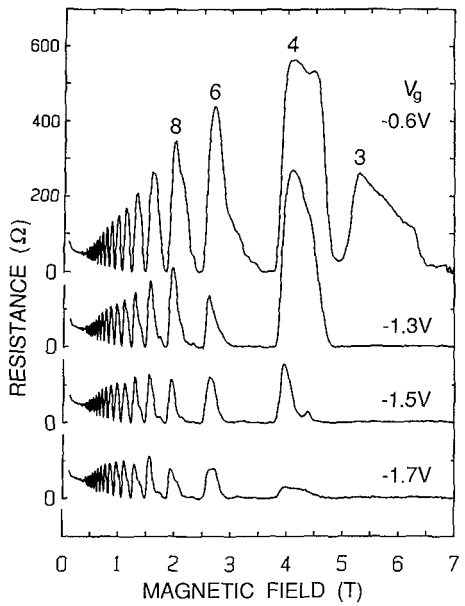

FIG. 44. Measurement of the anomalous Shubnikov–de Haas oscillations in the geometry of Fig. 43. The plotted longitudinal resistance is the voltage drop between contacts p and d divided by the current from s to d. At high magnetic fields, the oscillations increasingly are suppressed as the point contact in the voltage probe is pinched off by increasing the negative gate voltage. The number of occupied spin-split Landau levels in the bulk is indicated at several of the Shubnikov–de Haas maxima. (From Ref. 11.)

$(2e^2/h)(N_{\text{wide}} - 1)$ or smaller. At larger conductances, the oscillations in R_{3t} return, because this point contact now can couple to the highest-index edge channel and distribute the backscattered electrons over the lower-index edge channels. The point contact positioned between contacts p and d thus functions as a controllable *edge channel mixer.*

b. *Anomalous Scaling of the Shubnikov–de Haas Effect*

The conclusions of the previous subsection have interesting implications for the Shubnikov–de Haas oscillations in the strong-field regime even if measured with contacts that do *not* detect selectively certain edge channels only. Consider again the geometry of Fig. 43, in the low-gate voltage limit where the point contact voltage probe transmits all edge channels with unit probability. (This is the case of an *ideal* contact; *cf.* Section 9.a.) To simplify the expression, Eq. (48), for the three-terminal longitudinal resistance R_{3t}, we note that the transmission and reflection probabilities $T_{s\to p}$, R_s, and R_p

contact probe and drain is given by

$$R_{3t} = \frac{h}{2e^2} \frac{T_{s\to p}}{(N_s - R_s)(N_p - R_p) - T_{p\to s}T_{s\to p}}. \tag{48}$$

This three-terminal resistance corresponds to a generalized *longitudinal* resistance if the magnetic field has the direction of Fig. 43. In the absence of backscattering in the 2DEG, one has $T_{s\to p} = 0$, so that R_{3t} vanishes, as it should for a longitudinal resistance in a strong magnetic field.

Shubnikov–de Haas oscillations in the longitudinal resistance arise when backscattering leads to $T_{s\to p} \neq 0$. The resistance reaches a maximum when the Fermi level lies in a bulk Landau level, corresponding to a maximum probability for backscattering (which requires scattering from one edge to the other across the bulk of the sample, as indicated by the dashed lines in Fig. 43). From the preceding discussion of the anomalous quantum Hall effect, we know that the point contact voltage probe in a high magnetic field functions as a selective detector of edge channels with index n less than some value determined by the barrier height in the point contact. If, as discussed in Section 8, backscattering itself occurs selectively for the channel with the highest index $n = N_{wide}$, and if the edge channels with $n \leq N_{wide} - 1$ do not scatter to that edge channel, then a suppression of the Shubnikov–de Haas oscillations is to be expected when R_{3t} is measured with a point contact containing a sufficiently high potential barrier. This indeed was observed experimentally,[11] as is shown in Fig. 44. The Shubnikov–de Haas maximum at 5.2 T, for example, is found to disappear at gate voltages such that the point contact conductance is equal to or smaller than $2e^2/h$, which means that the point contact only transmits two spin-split edge channels. The number of occupied spin-split Landau levels in the bulk at this magnetic field value is three. This experiment thus demonstrates that the Shubnikov–de Haas oscillations result from the highest-index edge channel only, and that this edge channel does not scatter to the lower-index edge channels over the distance of 250 μm from point contact probe to drain.

In Section 9.a, we discussed how an *ideal* contact at the 2DEG boundary *induces* a local equilibrium by equipartitioning the outgoing current equally among the edge channels. The anomalous Shubnikov–de Haas effect provides a direct way to study this contact-induced equilibration by means of a second point contact between the point contact voltage probe p and the current drain d in Fig. 43. This experiment is described in Ref. 12. Once again, use was made of the double-split-gate point contact device (Fig. 2), in this case with a 1.5 μm separation between point contact p and the second point contact. It is found that the Shubnikov–de Haas oscillations in R_{3t} are suppressed only if the second point contact has a conductance of

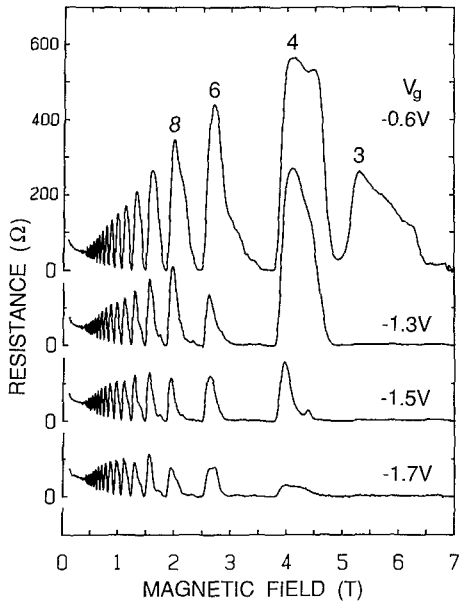

FIG. 44. Measurement of the anomalous Shubnikov–de Haas oscillations in the geometry of Fig. 43. The plotted longitudinal resistance is the voltage drop between contacts p and d divided by the current from s to d. At high magnetic fields, the oscillations increasingly are suppressed as the point contact in the voltage probe is pinched off by increasing the negative gate voltage. The number of occupied spin-split Landau levels in the bulk is indicated at several of the Shubnikov–de Haas maxima. (From Ref. 11.)

$(2e^2/h)(N_{\text{wide}} - 1)$ or smaller. At larger conductances, the oscillations in R_{3t} return, because this point contact now can couple to the highest-index edge channel and distribute the backscattered electrons over the lower-index edge channels. The point contact positioned between contacts p and d thus functions as a controllable *edge channel mixer.*

b. Anomalous Scaling of the Shubnikov–de Haas Effect

The conclusions of the previous subsection have interesting implications for the Shubnikov–de Haas oscillations in the strong-field regime even if measured with contacts that do *not* detect selectively certain edge channels only. Consider again the geometry of Fig. 43, in the low-gate voltage limit where the point contact voltage probe transmits all edge channels with unit probability. (This is the case of an *ideal* contact; *cf.* Section 9.a.) To simplify the expression, Eq. (48), for the three-terminal longitudinal resistance R_{3t}, we note that the transmission and reflection probabilities $T_{s\to p}$, R_s, and R_p

refer to the highest-index edge channel only (with index $n = N_{\text{wide}}$), under the assumptions of selective backscattering and absence of scattering to lower-index edge channels discussed before As a consequence, $T_{s\to p}$, R_s, and R_p each are at most equal to 1, so that up to corrections smaller by a factor N_{wide}^{-1}, we may put these terms equal to zero in the denominator on the right-hand side of Eq (48) In the numerator, the transmission probability $T_{s\to p}$ may be replaced by the backscattering probability $t_{bs} \leq 1$, which is the probability that the highest-index edge channel injected by the source contact reaches the point contact probe following scattering across the wide 2DEG (dashed lines in Fig 43) With these simplifications, Eq (48) takes the form (assuming spin degeneracy),

$$R_{3t} = \frac{h}{2e^2} \frac{t_{bs}}{N_{\text{wide}}^2} \times (1 + \text{order } N_{\text{wide}}^{-1}) \tag{49}$$

Only if $t_{bs} \ll 1$ may the backscattering probability be expected to scale linearly with the separation of the two contacts p and d (between which the voltage drop is measured) If t_{bs} is not small, then the upper limit $t_{bs} \leq 1$ leads to the novel prediction of a *maximum* possible amplitude,

$$R_{\max} = \frac{h}{2e^2} \frac{1}{N_{\text{wide}}^2} \times (1 + \text{order } N_{\text{wide}}^{-1}), \tag{50}$$

of the Shubnikov–de Haas resistance oscillations in a given large magnetic field, independently of the length of the segment over which the voltage drop is measured—provided equilibration does not occur on this segment Equilibration might result, for example, from the presence of additional contacts between the voltage probes, as discussed in Sections 9 a and 10 a One easily verifies that the high-field Shubnikov–de Haas oscillations in Fig 44 at $V_g = -0\,6$ V (when the point contact is just defined, so that the potential barrier is small) lie well below the upper limit, from Eq (50) For example, the peak around 2 T corresponds to the case of four occupied spin-degenerate Landau levels, so that the theoretical upper limit is $(h/2e^2) \times \frac{1}{16} \approx 800\ \Omega$, well above the observed peak value of about 350 Ω The prediction of a maximum longitudinal resistance implies that the linear scaling of the amplitude of the Shubnikov–de Haas oscillations with the distance between voltage probes found in the weak-field regime, and expected on the basis of a description in terms of a local resistivity tensor,[2] breaks down in strong magnetic fields Anomalous scaling of the Shubnikov–de Haas effect has been observed experimentally,[188 223 224] and also has been interpreted[225] recently in terms of a non-equilibrium between the edge channels

Selective backscattering and the absence of local equilibrium have con-

sequences as well for the two-terminal resistance in strong magnetic fields. In weak fields, one usually observes in two-terminal measurements a superposition of the Shubnikov–de Haas longitudinal resistance oscillations and the quantized Hall resistance (*cf.* Section 9.a). This superposition shows up as a characteristic *overshoot* of the two-terminal resistance as a function of magnetic field, as it increases from one quantized Hall plateau to the next. (The plateaus coincide with minima of the Shubnikov–de Haas oscillations.) In the strong-field regime (in the absence of equilibration between source and drain contacts), no such superposition is to be expected. Instead, the two-terminal resistance would increase monotonically from $(h/2e^2)N_{\text{wide}}^{-1}$ to $(h/2e^2)(N_{\text{wide}} - 1)^{-1}$ as the transmission probability from source to drain decreases from N_{wide} to $N_{\text{wide}} - 1$. We are not aware of an experimental test of this prediction. In Chapter 4, Büttiker discusses an experiment by Fang *et al.* on the Shubnikov–de Haas oscillations in the short-channel regime, to which our analysis does not apply.

11. AHARONOV–BOHM OSCILLATIONS AND INTER-EDGE CHANNEL TUNNELING

a. *Aharonov–Bohm Effect in a Singly Connected Geometry*

The *Aharonov–Bohm effect* is a fundamental manifestation of the influence of a magnetic field on the phase of the electron wave function.[226,227] In the solid state, Aharonov–Bohm magnetoresistance oscillations have been studied extensively in metal rings and cylinders,[228,229] and more recently with a much larger amplitude in 2DEG rings.[230,231] (See Chapter 3 by Timp.) In such experiments, oscillations in the resistance of the ring are observed as a function of the applied perpendicular magnetic field B. The oscillations are periodic in B, with a fundamental period $\Delta B = h/eA$ determined by the area A of the ring. Their origin is the field-induced phase difference between the two paths (one clockwise, one counterclockwise), which take an electron from one side of the ring to the other. Normally, a *multiply*-connected geometry (a ring) is required to see the effect. It thus came as a surprise when a magnetoresistance oscillation periodic in B was discovered in a quantum point contact,[16] which constitutes a *singly* connected geometry. A similar effect more recently has been reported by Wharam *et al.*[232]

The magnetic field dependence of the two-terminal resistance in the experiment of Ref. 16 is shown in Fig. 45. The periodic oscillations occur predominantly between quantum Hall plateaus, in a range of gate voltages only, and only at low temperatures. (In Fig. 45, $T = 50$ mK; the effect has disappeared at 1 K.) The fine structure is very reproducible if the sample is kept in the cold, but changes after cycling to room temperature. As one can see

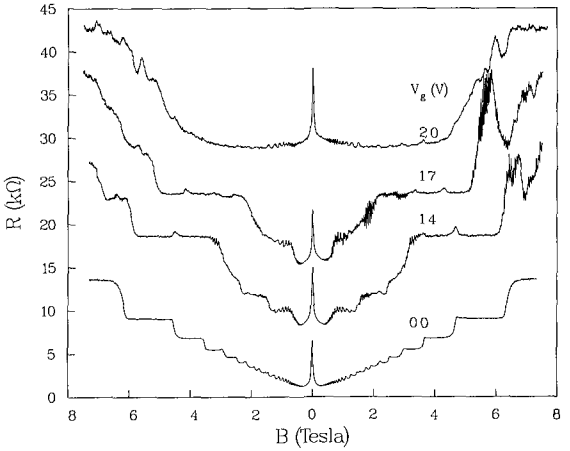

FIG 45 Two-terminal magnetoresistance of a point contact for a series of gate voltages at $T = 50$ mK, showing oscillations that are periodic in B between the quantum Hall plateaus The second, third and fourth curves from the bottom have offsets, respectively, of 5, 10, and 15 kΩ The rapid oscillations below 1 T are Shubnikov–de Haas oscillations periodic in $1/B$, originating from the wide 2DEG regions The sharp peak around $B = 0$ T originates from the ohmic contacts (From Ref 16)

from the enlargements in Fig 46, a splitting of the peaks occurs in a range of magnetic fields, presumably as spin-splitting becomes resolved A curious aspect of the effect (which has remained unexplained) is that the oscillations have a much larger amplitude in one field direction than in the other (Fig 45), in apparent conflict with the $\pm B$ symmetry of the two-terminal resistance required by the reciprocity relation[7 158] in the absence of magnetic impurities Other devices of the same design did not show oscillations of well-defined periodicity, and had a two-terminal resistance that was approximately $\pm B$ symmetric

Figure 47 illustrates the tunneling mechanism for the periodic magnetoresistance oscillations as it was originally proposed[16] to explain the observations Because of the presence of a barrier in the point contact, the electrostatic potential has a saddle form Equipotentials at the guiding center energy (*cf* Section 8) are drawn schematically in Fig 47 (Arrows indicate the direction of motion along the equipotential) An electron that enters the constriction at a can be reflected back into the broad region by tunneling to the opposite edge, either at the potential step at the entrance of the constriction (from a to b) or at its exit (from d to c) These two tunneling paths acquire an Aharonov–Bohm phase difference[186 223] of $eBA/\hbar$ (where A is the enclosed area $abcd$), leading to periodic magnetoresistance oscillations (Note that the periodicity ΔB may differ[17 208] somewhat from the usual expression $\Delta B = h/eA$, if A itself is B-dependent due to the B-dependence of the guiding

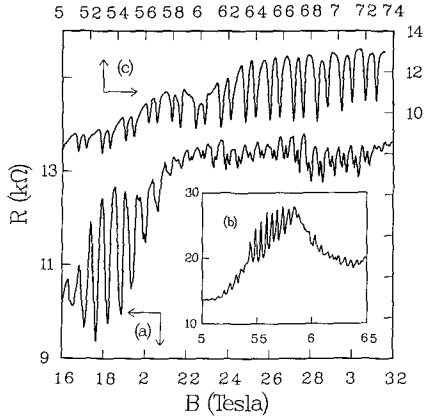

FIG. 46. Curves *a* and *b* are close-ups of the curve for $V_g = -1.7$ V in Fig. 45. Curve *c* was measured three months earlier on the same device. (Note the different field scale due to a change in electron density in the constriction.) (From Ref. 16.)

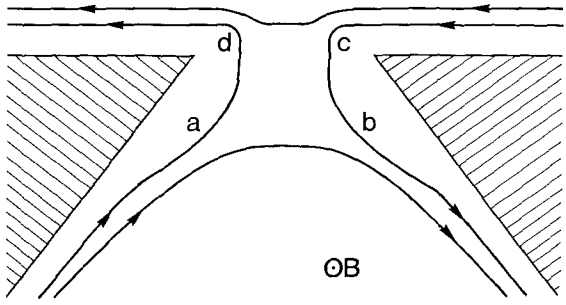

FIG. 47. Equipotentials at the guiding center energy in the saddle-shaped potential created by a split gate (shaded). Aharonov–Bohm oscillations in the point contact magnetoresistance result from the interference of tunneling paths *ab* and *adcb*. Tunneling from *a* to *b* may be assisted by an impurity at the entrance of the constriction. (From Ref. 16.)

center energy from Eq. (38).) This mechanism shows how an Aharonov–Bohm effect is possible in principle in a singly connected geometry: The point contact behaves as if it were multiply connected, by virtue of the spatial separation of edge channels moving in opposite directions. (Related mechanisms, based on circulating edge currents, have been considered for Aharonov–Bohm effects in small conductors.)[221,222,234–236] Unlike in the original Aharonov–Bohm effect[226,227,237] in vacuum, where only the effect of the magnetic field on the *phase* of the wave function matters, the *Lorentz force* plays an essential role here. This explains why the oscillations periodic in B are observed only at large magnetic fields in Fig. 45 (above about 1 T; the oscillations at lower fields are Shubnikov–de Haas oscillations periodic in $1/B$, due to the series

resistance of the wide 2DEG regions). At low magnetic fields, the spatial separation of edge channels responsible for the Aharonov–Bohm effect is not effective yet. The spatial separation also can be destroyed by a large negative gate voltage (top curve in Fig. 45) when the width of the point contact becomes so small that the wave functions of edge states at opposite edges overlap.

Although the mechanism illustrated in Fig. 47 is attractive because it is an intrinsic consequence of the point contact geometry, the observed well-defined periodicity of the magnetoresistance oscillations requires that the potential induced by the split gate varies rapidly over a short distance (to have a well-defined area A). A smooth saddle potential seems more realistic. Moreover, one would expect the periodicity to vary more strongly with gate voltage than the small 10% variation observed experimentally as V_g is changed from -1.4 to -1.7 V. Glazman and Jonson[208] have proposed an alternative *impurity-assisted* tunneling mechanism as a more likely (but essentially equivalent) explanation of the experiment.[16] In their picture, one of the two tunneling processes (from a to b in Fig. 47) is mediated by an impurity outside but close to the constriction. The combination of impurity and point contact introduces a well-defined area even for a smooth saddle potential, which moreover will not be strongly gate voltage-dependent. We note that the single sample in which the effect was observed was special in having an anomalously small pinch-off voltage. (See the curves labeled 3 in Fig. 14.) This well may be due to the accidental presence of an impurity in the immediate vicinity of the constriction. To study the Aharonov–Bohm effect due to inter-edge channel tunneling under more controlled conditions, a different device geometry is necessary,[17] as discussed in the following subsection.

b. Tunneling through an Edge State Bound in a Cavity

i. Experiment. The geometry of Ref. 17 is shown schematically in Fig. 48. A cavity with two opposite point contact openings is defined in the 2DEG by split gates A and B. The diameter of the cavity is approximately 1.5 μm. The conductances G_A and G_B of the two point contacts can be measured independently (by grounding one set of gates), with the results plotted in Figs. 49a,b (for $V_g = -0.35$ V on either gate A or B). The conductance G_C of the cavity (for $V_g = -0.35$ V on both the split gates) is plotted in Fig. 49c. A long series of periodic oscillations is observed between two quantum Hall plateaus. Similar series of oscillations (but with a different periodicity) have been observed between other quantum Hall plateaus. The oscillations are suppressed on the plateaus themselves. The amplitude of the oscillations is comparable to that observed in the experiment on a single point contact[16] (discussed in Section 11.a), but the period is much smaller (consistent with a larger effective area in the double point contact device), and also no splitting of

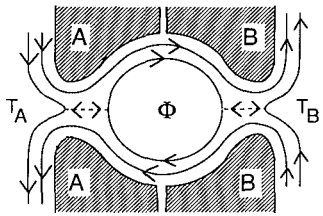

FIG. 48. Tunneling through an edge state bound in a cavity. The cavity (of 1.5 μm diameter) is defined by a double set of split gates A and B. For large negative gate voltages, the 2DEG region under the narrow gap between gates A and B is fully depleted, while transmission remains possible over the potential barrier in the wider openings at the left and right of the cavity. Tunneling through the left and right barrier (as indicated by dashed lines) occurs with transmission probabilities T_A and T_B, which are adjustable separately by means of the voltages on gates A and B. On increasing the magnetic field, resonant tunneling through the cavity occurs periodically each time the flux Φ enclosed by the circulating edge state increases by one flux quantum h/e. (From Ref. 17.)

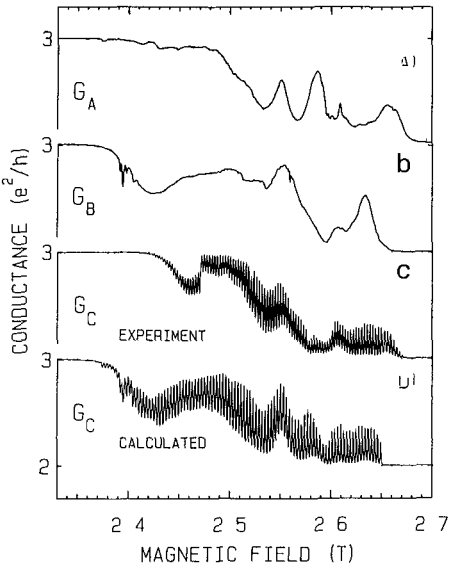

FIG. 49. Magnetoconductance experiments on the device of Fig. 48 at 6 mK, for a fixed gate voltage of -0.35 V. (*a*) Conductance of point contact A, measured with gate B grounded. (*b*) Conductance of point contact B (gate A grounded). (*c*) Measured conductance of the entire cavity. (*d*) Calculated conductance of the cavity, obtained from Eqs. (51) and (52) with the measured G_A and G_B as input. (From Ref. 17.)

the peaks is observed (presumably due to a fully resolved spin degeneracy). No gross $\pm B$ asymmetries were found in the present experiment, although an accurate test of the symmetry on field reversal was not possible because of difficulties with the reproducibility. The oscillations are quite fragile, disappearing when the temperature is raised above 200 mK or when the voltage across the device exceeds 40 μV. (The data in Fig. 49 was taken at 6 mK and 6 μV.) The experimental data is well described by resonant transmission through a circulating edge state in the cavity,[17] as illustrated in Fig. 48 and described in detail shortly. Aharanov–Bohm oscillations due to resonant transmission through a similar structure have been reported more recently by Brown *et al.*[238]

ii. Model. As discussed in Sections 9.b and 11.a, the electrostatic potential defining each point contact has a saddle shape, due to the combination of the lateral confinement and the potential barrier. The height of the barrier can be adjusted by means of the gate voltage. An edge state with a guiding center energy below the barrier height forms a bound state in the cavity formed by two opposite point contacts, as is illustrated in Fig. 48. Tunneling of edge channels through the cavity via this bound state occurs with transmission probability T_{AB}, which for a single edge channel is given by[221,239]

$$T_{AB} = \left| \frac{t_A t_B}{1 - r_A r_B \exp(i\Phi e/\hbar)} \right|^2 = \\ = \frac{T_A T_B}{1 + R_A R_B - 2(R_A R_B)^{1/2} \cos(\phi_0 + \Phi e/\hbar)}. \tag{51}$$

Here, t_A and r_A are the transmission and reflection probability amplitudes through point contact A, $T_A \equiv |t_A|^2$ and $R_A \equiv |r_A|^2 = 1 - T_A$ are the transmission and reflection probabilities, and t_B, r_B, T_B, R_B denote the corresponding quantities for point contact B. In Eq. (51), the phase acquired by the electron on one revolution around the cavity is the sum of the phase ϕ_0 from the reflection probability amplitudes (which can be assumed to be only weakly B-dependent) and the Aharanov–Bohm phase $\Phi \equiv BA$, which varies rapidly with B. (Φ is the flux through the area A enclosed by the equipotential along which the circulating edge state is extended.) Resonant tunneling occurs periodically with B, whenever $\phi_0 + \Phi e/\hbar$ is a multiple of 2π.

In the case that only a single (spin-split) edge channel is occupied in the 2DEG, the conductance $G_C = (e^2/h)T_{AB}$ of the cavity follows directly from Eq. (51). The transmission and reflection probabilities can be determined independently from the individual point contact conductances, $G_A = (e^2/h)T_A$ (and similarly for G_B)—at least if one may assume that the presence of the cavity has no effect itself on T_A and T_B (but only on the total transmission

probability T_{AB}). If $N > 1$ spin-split edge channels are occupied, we assume that the $N - 1$ lowest-index edge channels are fully transmitted and do not modify the transmission probability of the Nth edge channel, so that we can write[17]

$$G_C = \frac{e^2}{h}(N - 1 + T_{AB}),\ G_A = \frac{e^2}{h}(N - 1 + T_A),\ G_B = \frac{e^2}{h}(N - 1 + T_B). \quad (52)$$

This simple model is compared with the experiment in Fig. 49. The trace in Fig. 49d has been calculated from Eqs. (51) and (52) using the individual point contact conductances in Figs. 49a,b as input for T_A and T_B. The flux Φ has been adjusted to the experimental periodicity of 3 mT, and the phase ϕ_0 in Eq. (51) has been ignored (since that only would amount to a phase shift of the oscillations). Energy averaging due to the finite temperature and voltage has been taken into account in the calculation.[17] The agreement with the experimental trace (Fig. 49c) is quite satisfactory.

The results shown in Fig. 49 demonstrate once again how some textbook quantum mechanics becomes reality when one studies transport through quantum point contacts. This seems an appropriate place to conclude our review of the subject. From initial work on the properties of single quantum point contacts, we have moved on to structures involving two adjacent point contacts (as in the electron focusing geometry), or two opposite point contacts (as in the cavity geometry considered before). The basic principles of the ballistic and adiabatic transport regime are now well-understood, and have been demonstrated experimentally. Future work in other transport regimes and in more complicated structures undoubtedly will reveal more interesting and potentially useful phenomena. It is our belief that the quantum point contact, because of its simplicity and fundamental nature, will continue to be used as a versatile tool and building block in such investigations.

Acknowledgments

We thank our collaborators from the Delft University of Technology, and the Philips Research Laboratories in Eindhoven and Redhill for their generous permission to reproduce results from our joint publications.[4,10–13,15–17,57–61] Reference 34 has been part of a collaboration between Philips and the University of Cambridge. The stimulating support of Professors J. A. Pals and M. F. H. Schuurmans from the Philips Research Laboratories is gratefully acknowledged. Technological support has been provided by the Delft Center for Submicron Technology and the Philips Mask Center. The work at Delft is supported financially by the Stichting voor Fundamenteel Onderzoek der Materie (F.O.M.).

REFERENCES

1 R Landauer, in *Localization, Interaction, and Transport Phenomena* (B Kramer, G Bergmann, and Y Bruynseraede, eds), Springer, Berlin (1985)
2 T Ando, A B Fowler, and F Stern, *Rev Mod Phys* **54**, 437 (1982)
3 J J Harris, J A Pals, and R Woltjer, *Rep Prog Phys* **52**, 1217 (1989)
4 B J van Wees, H van Houten, C W J Beenakker, J G Williamson, L P Kouwenhoven, D van der Marel, and C T Foxon, *Phys Rev Lett* **60**, 848 (1988)
5 D A Wharam, T J Thornton, R Newbury, M Pepper, H Ahmed, J E F Frost, D G Hasko, D C Peacock, D A Ritchie, and G A C Jones, *J Phys C* **21**, L209 (1988)
6 R Landauer, *IBM J Res Dev* **1**, 223 (1957)
7 M Buttiker, *Phys Rev Lett* **57**, 1761 (1986)
8 C W J Beenakker and H van Houten, *Phys Rev B* **39**, 10445 (1989)
9 D A Wharam, M Pepper, H Ahmed, J E F Frost, D G Hasko, D C Peacock, D A Ritchie, and G A C Jones, *J Phys C* **21**, L887 (1988)
10 B J van Wees, E M M Willems, C J P M Harmans, C W J Beenakker, H van Houten, J G Williamson, C T Foxon, and J J Harris, *Phys Rev Lett* **62**, 1181 (1989)
11 B J van Wees, E M M Willems, L P Kouwenhoven, C J P M Harmans, J G Williamson, C T Foxon, and J J Harris, *Phys Rev B* **39**, 8066 (1989)
12 B J van Wees, L P Kouwenhoven, E M M Willems, C J P M Harmans, J E Mooij, H van Houten, C W J Beenakker, J G Williamson, and C T Foxon, *Phys Rev B* **43**, 12431 (1991)
13 H van Houten, B J van Wees, J E Mooij, C W J Beenakker, J G Williamson, and C T Foxon, *Europhys Lett* **5**, 721 (1988)
14 C W J Beenakker, H van Houten, and B J van Wees, *Europhys Lett* **7**, 359 (1988)
15 H van Houten, C W J Beenakker, J G Williamson, M E I Broekkaart, P H M van Loosdrecht, B J van Wees, J E Mooij, C T Foxon, and J J Harris, *Phys Rev B* **39**, 8556 (1989)
16 P H M van Loosdrecht, C W J Beenakker, H van Houten, J G Williamson, B J van Wees, J E Mooij, C T Foxon, and J J Harris, *Phys Rev B* **38**, 10162 (1988)
17 B J van Wees, L P Kouwenhoven, C J P M Harmans, J G Williamson, C E T Timmering, M E I Broekaart, C T Foxon, and J J Harris, *Phys Rev Lett* **62**, 2523 (1989)
18 C W J Beenakker and H van Houten, *Solid State Physics* **44**, 1 (1991)
19 H Heinrich, G Bauer, and F Kuchar, eds , *Physics and Technology of Submicron Structures*, Springer, Berlin (1988)
20 M Reed and W P Kirk, eds , *Nanostructure Physics and Fabrication*, Academic Press, New York (1989)
21 S P Beaumont and C M Sotomayor-Torres, eds , *Science and Engineering of* 1- *and* 0-*Dimensional Semiconductors*, Plenum, London (1990)
22 J M Chamberlain, L Eaves, and J C Portal, eds , *Electronic Properties of Multilayers and Low-Dimensional Semiconductor Structures*, Plenum, New York (1990)
23 Yu V Sharvin, *Zh Eksp Teor Fiz* **48**, 984 (1965) [*Sov Phys JETP* **21**, 655 (1965)]
24 V S Tsoi, *Pis'ma Zh Eksp Teor Fiz* **19**, 114 (1974) [*JETP Lett* **19**, 701 (1974)], *Zh Eksp Teor Fiz* **68**, 1849 (1975) [*Sov Phys JETP* **41**, 927 (1975)]
25 I K Yanson, *Zh Eksp Teor Fiz* **66**, 1035 (1974) [*Sov Phys JETP* **39**, 506 (1974)]
26 A G M Jansen, A P van Gelder, and P Wyder, *J Phys C* **13**, 6073 (1980)
27 R Trzcinski, E Gmelin, and H J Queisser, *Phys Rev B* **35**, 6373 (1987)
28 T J Thornton, M Pepper, H Ahmed, D Andrews, and G J Davies, *Phys Rev Lett* **56**, 1198 (1986)
29 H Z Zheng, H P Wei, D C Tsui, and G Weimann, *Phys Rev B* **34**, 5635 (1986)

30 J P Kirtley, Z Schlesinger, T N Theis, F P Milliken, S L Wright, and L F Palmateer, *Phys Rev B* **34**, 5414 (1986)
31 H van Houten, B J van Wees, M G J Heijman, and J P Andre, *Appl Phys Lett* **49**, 1781 (1986)
32 Y Hirayama, T Saku, and Y Horikoshi, *Phys Rev B* **39**, 5535 (1989), Y Hirayama and T Saku, *Appl Phys Lett* **54**, 2556 (1989)
33 A D Wieck and K Ploog, *Surf Sci* **229**, 252 (1990), *Appl Phys Lett* **56**, 928 (1990)
34 H van Houten, C W J Beenakker, P H M van Loosdrecht, T J Thornton, H Ahmed, M Pepper, C T Foxon, and J J Harris, *Phys Rev B* **37**, 8534 (1988), and unpublished
35 J C Maxwell, *A Treatise on Electricity and Magnetism*, Clarendon, Oxford (1904)
36 R Holm, *Electrical Contacts Handbook*, Springer, Berlin (1958)
37 Yu V Sharvin and N I Bogatina, *Zh Eksp Teor Fiz* **56**, 772 (1969) [*Sov Phys JETP* **29**, 419 (1969)]
38 R A R Tricker, *Proc Camb Phil Soc* **22**, 454 (1925)
39 A B Pippard, *Magnetoresistance in Metals*, Cambridge University Press, Cambridge (1989)
40 P C van Son, H van Kempen, and P Wyder, *Phys Rev Lett* **58**, 1567 (1987)
41 E E Vdovin, A Yu Kasumov, Ch V Kopetskii, and I B Levinson, *Zh Eksp Teor Fiz* **92**, 1026 (1987) [*Sov Phys JETP* **65**, 582 (1987)]
42 H M Swartjes, A P van Gelder, A G M Jansen, and P Wyder, *Phys Rev B* **39**, 3086 (1989)
43 J K Gimzewski and R Moller, *Phys Rev B* **36**, 1284 (1987)
44 N D Lang, *Phys Rev B* **36**, 8173 (1987)
45 J Ferrer, A Martin-Rodero, and F Flores, *Phys Rev B* **38**, 10113 (1988)
46 N D Lang, *Comm Cond Matt Phys* **14**, 253 (1989)
47 N D Lang, A Yacoby, and Y Imry, *Phys Rev Lett* **63**, 1499 (1989)
48 N Garcia and H Rohrer, *J Phys Condens Matter* **1**, 3737 (1989)
49 A C Warren, D A Antoniadis, and H I Smith, *Phys Rev Lett* **56**, 1858 (1986)
50 K K Choi, D C Tsui, and S C Palmateer, *Phys Rev B* **33**, 8216 (1986)
51 G Timp, A M Chang, P Mankiewich, R Behringer, J E Cunningham, T Y Chang, and R E Howard, *Phys Rev Lett* **59**, 732 (1987)
52 M L Roukes, A Scherer, S J Allen, Jr, H G Craighead, R M Ruthen, E D Beebe, and J P Harbison, *Phys Rev Lett* **59**, 3011 (1987)
53 A M Chang, G Timp, T Y Chang, J E Cunningham, P M Mankiewich, R E Behringer, and R E Howard, *Solid State Comm* **67**, 769 (1988)
54 Y Takagaki, K Gamo, S Namba, S Ishida, S Takaoka, K Murase, K Ishibashi, and Y Aoyagi, *Solid State Comm* **68**, 1051 (1988)
55 M L Roukes, T J Thornton, A Scherer, J A Simmons, B P van der Gaag, and E D Beebe, in *Science and Engineering of 1- and 0-Dimensional Semiconductors* (S P Beaumont and C M Sotomayer-Torres, eds), Plenum, London (1990)
56 G Timp, H U Baranger, P de Vegvar, J E Cunningham, R E Howard, R Behringer, and P M Mankiewich, *Phys Rev Lett* **60**, 2081 (1988)
57 B J van Wees, L P Kouwenhoven, H van Houten, C W J Beenakker, J E Mooij, C T Foxon, and J J Harris, *Phys Rev B* **38**, 3625 (1988)
58 L W Molenkamp, A A M Staring, C W J Beenakker, R Eppenga, C E Timmering, J G Williamson, C J P M Harmans, and C T Foxon, *Phys Rev B* (**41**, 1274 (1990)
59 A A M Staring, L W Molenkamp, C W J Beenakker, L P Kouwenhoven, and C T Foxon, *Phys Rev B* **41**, 8461 (1990)
60 L P Kouwenhoven, B J van Wees, C J P M Harmans, J G Williamson, H van Houten, C W J Beenakker, C T Foxon, and J J Harris, *Phys Rev B* **39**, 8040 (1989)
61 J G Williamson, H van Houten, C W J Beenakker, M E I Broekaart, L I A Spendeler, B J van Wees, and C T Foxon, *Phys Rev B* **41**, 1207 (1990), *Surf Sci* **229**, 303 (1990)

62 K von Klitzing, G Dorda, and M Pepper, *Phys Rev Lett* **45**, 494 (1980)
63 G Timp R Behringer, S Sampere, J E Cunningham, and R E Howard, in *Nanostructure Physics and Fabrication* (M A Reed and W P Kirk, eds), Academic Press, New York (1989)
64 R E Prange and S M Girvin, eds, *The Quantum Hall Effect*, Springer, New York (1987)
65 H van Houten, B J van Wees, and C W J Beenakker, in *Physics and Technology of Submicron Structures* (H Heinrich, G Bauer, and F Kuchar, eds), Springer, Berlin (1988)
66 H van Houten and C W J Beenakker, in *Nanostructure Physics and Fabrication* (M A Reed and W P Kirk, eds), Academic Press, New York (1989), see also M Knudsen, *The Kinetic Theory of Gases*, Methuen, London (1934)
67 D S Fisher and P A Lee, *Phys Rev B* **23**, 6851 (1981), E N Economou and C M Soukoulis, *Phys Rev Lett* **46**, 618 (1981)
68 Y Imry, in *Directions in Condensed Matter Physics*, Vol 1 (G Grinstein and G Mazenko, eds), World Scientific Publishing Co Singapore (1986)
69 A D Stone and A Szafer, *IBM J Res Dev* **32**, 384 (1988)
70 D J Thouless, *Phys Rev Lett* **47**, 972 (1981)
71 R Landauer, *Phys Lett A* **85**, 91 (1981)
72 H L Engquist and P W Anderson, *Phys Rev B* **24**, 1151 (1981)
73 D C Langreth and E Abrahams, *Phys Rev B* **24**, 2978 (1981)
74 P W Anderson, D J Thouless, E Abrahams, and D S Fisher, *Phys Rev B* **22**, 3519 (1980)
75 M Buttiker, Y Imry, R Landauer, and S Pinhas, *Phys Rev B* **31**, 6207 (1985)
76 M Ya Azbel, *J Phys C* **14**, L225 (1981)
77 M Buttiker, *Phys Rev B* **38**, 9375 (1988)
78 J Kucera and P Streda *J Phys C* **21**, 4357 (1988)
79 H U Baranger and A D Stone, *Phys Rev B* **40**, 8169 (1989)
80 Y Isawa, *J Phys Soc Japan* **57**, 3457 (1988)
81 B Kramer and J Masek, *J Phys C* **21**, L1147 (1988)
82 R Johnston and L Schweitzer, *J Phys C* **21**, L861 (1988)
83 R Landauer, *J Phys Condens Matter* **1**, 8099 (1989), also in *Nanostructure Physics and Fabrication* (M A Reed and W P Kirk, eds), Academic Press, New York (1989)
84 L Escapa and N Garcia, *J Phys Condens Matter* **1**, 2125 (1989)
85 E G Haanappel and D van der Marel, *Phys Rev B* **39**, 5484 (1989), D van der Marel and E G Haanappel, *Phys Rev B* **39**, 7811 (1989)
86 G Kirczenow, *Solid State Comm* **68**, 715 (1988), *J Phys Condens Matter* **1**, 305 (1989)
87 A Szafer and A D Stone, *Phys Rev Lett* **62**, 300 (1989)
88 E Tekman and S Ciraci, *Phys Rev B* **39**, 8772 (1989), *Phys Rev B* **40**, 8559 (1989)
89 Song He and S Das Sarma, *Phys Rev B* **40**, 3379 (1989)
90 N Garcia and L Escapa, *Appl Phys Lett* **54**, 1418 (1989)
91 E Castaño and G Kirczenow, *Solid State Comm* **70**, 801 (1989)
92 A Kawabata, *J Phys Soc Japan* **58**, 372 (1989)
93 I B Levinson, *Pis'ma Zh Eksp Teor Fiz* **48**, 273 (1988) [*JETP Lett* **48**, 301 (1988)]
94 A Matulis and D Segzda, *J Phys Condens Matter* **1**, 2289 (1989)
95 M Buttiker, *Phys Rev B* **41**, 7906 (1990)
96 H van Houten and C W J Beenakker, in *Analogies in Optics and Microelectronics* (W van Haeringen and D Lenstra, eds), Kluwer, Deventer, (1990)
97 L I Glazman, G B Lesovik, D E Khmel nitskii, R I Shekhter, *Pis'ma Zh Teor Fiz* **48**, 218 (1988) [*JETP Lett* **48**, 238 (1988)]
98 R Landauer, *Z Phys B* **68**, 217 (1987)
99 C S Chu and R S Sorbello, *Phys Rev B* **40**, 5941 (1989)
100 J Masek, P Lipavsky, and B Kramer, *J Phys Condens Matter* **1**, 6395 (1989)
101 P F Bagwell and T P Orlando, *Phys Rev B* **40**, 1456 (1989)

102 R J Brown, M J Kelly, R Newbury, M Pepper, B Miller, H Ahmed, D G Hasko, D C Peacock, D A Ritchie, J E F Frost, and G A C Jones, *Solid State Electron* **32**, 1179 (1989)

103 Y Hirayama, T Saku, and Y Horikoshi, *Jap J Appl Phys* **28**, L701 (1989)

104 B L Al'tshuler, *Pis'ma Zh Eksp Teor Fiz* **41**, 530 (1985) [*JETP Lett* **41**, 648 (1985)]

105 P A Lee and A D Stone, *Phys Rev Lett* **55**, 1622 (1985), P A Lee, A D Stone, and H Fukuyama, *Phys Rev B* **35**, 1039 (1987)

106 G Timp, in *Mesoscopic Phenomena in Solids* (P A Lee, R A Webb, and B L Al'tshuler, eds), Elsevier Science Publ , Amsterdam (1989)

107 J H Davies and J A Nixon, *Phys Rev B* **39**, 3423 (1989), J H Davies in *Nanostructure Physics and Fabrication* (M A Reed and W P Kirk, eds), Academic Press, New York (1989)

108 T J Thornton, M L Roukes, A Scherer, and B van der Gaag, *Phys Rev Lett* **63**, 2128 (1989)

109 H van Houten, C W J Beenakker, M E I Broekaart, M G H J Heijman, B J van Wees, H E Mooij, and J P Andre, *Acta Electronica*, **28**, 27 (1988)

110 L D Landau and E M Lifshitz, *Quantum Mechanics*, Pergamon, Oxford, (1977)

111 K -F Berggren, G Roos, and H van Houten, *Phys Rev B* **37**, 10118 (1988)

112 L I Glazman and A V Khaetskii, *J Phys Condens Matter* **1**, 5005 (1989)

113 Y Avishai and Y B Band, *Phys Rev B* **40**, 3429 (1989)

114 K B Efetov, *J Phys Condens Matter* **1**, 5535 (1989)

115 F F Fang and P J Stiles, *Phys Rev B* **27**, 6487 (1983)

116 T G Powell, C C Dean, and M Pepper, *J Phys C* **17**, L359 (1984)

117 G L J A Rikken, J A M M van Haaren, W van der Wel, A P van Gelder, H van Kempen, P Wyder, J P Andre, K Ploog, and G Weimann, *Phys Rev B* **37**, 6181 (1988)

118 K F Berggren, T J Thornton, D J Newson, and M Pepper, *Phys Rev Lett* **57**, 1769 (1986)

119 S B Kaplan and A C Warren, *Phys Rev B* **39**, 1346 (1986)

120 H van Houten, B J van Wees, J E Mooij, G Roos, and K -F Berggren, *Superlattices and Microstructures*, **3**, 497 (1987)

121 J F Weisz and K -F Berggren, *Phys Rev B* **40**, 1325 (1989)

122 S E Laux, D J Frank, and F Stern, *Surf Sci* **196**, 101 (1988)

123 A Kumar, S E Laux, and F Stern, *Appl Phys Lett* **54**, 1270 (1989), *Bull Am Phys Soc* **34**, 589 (1989)

124 D A Wharam, U Ekenberg, M Pepper, D G Hasko, H Ahmed, J E F Frost, D A Ritchie, D C Peacock, and G A C Jones, *Phys Rev B* **39**, 6283 (1989)

125 L P Kouwenhoven, Master Thesis, Delft University of Technology, Delft, The Netherlands (1988)

126 B R Snell, P H Beton, P C Main, A Neves, J R Owers-Bradley, L Eaves, M Henini, O H Hughes, S P Beaumont, and C D W Wilkinson, *J Phys Condens Matter* **1**, 7499 (1989)

127 K von Klitzing and G Ebert, in *Two-Dimensional Systems, Heterostructures, and Superlattices* (G Bauer, F Kuchar, and H Heinrich, eds), Springer, Heidelberg, Germany (1984), **61**, 1001 (1988)

128 F F Fang and P J Stiles, *Phys Rev B* **29**, 3749 (1984)

129 D A Syphers, F F Fang, and P J Stiles, *Surf Sci* **142**, 208 (1984)

130 D A Syphers and P J Stiles, *Phys Rev B* **32**, 6620 (1985)

131 A B Berkut, Yu V Dubrovskii, M S Nunuparov, M I Reznikov, and V I Tal'yanskii, *Pis'ma Zh Teor Fiz* **44**, 252 (1986) [*JETP Lett* **44**, 324 (1986)]

132 G L J A Rikken, J A M M van Haaren, A P van Gelder, H van Kempen, P Wyder, H -U Habermeier, and K Ploog, *Phys Rev B* **37**, 10229 (1988)

133 R J Haug, A H MacDonald, P Streda, and K von Klitzing, *Phys Rev Lett* **61**, 2797 (1988)

134 S Washburn, A B Fowler, H Schmid, and D Kern, *Phys Rev Lett* **61**, 2801 (1988)

135 H Hirai, S Komiyama, S Hiyamizu, and S Sasa, in *Proc 19th International Conference on the Physics of Semiconductors* (W Zawadski, ed), Inst of Phys , Polish Acad Sci (1988)
136 B E Kane, D C Tsui, and G Weimann, *Phys Rev Lett* **61**, 1123 (1988)
137 R J Haug, J Kucera, P Streda, and K von Klitzing, *Phys Rev B* **39**, 10892 (1989)
138 Y Zhu, J Shi, and S Feng, *Phys Rev B* **41**, 8509 (1990)
139 Y Hirayama and T Saku, *Solid State Commun* **73**, 113 (1990), **41**, 2927 (1990)
140 E Castaño and G Kirczenow, *Phys Rev B* **41**, 5055 (1990)
141 Y Avishai, M Kaveh, S Shatz, and Y B Band, *J Phys Condens Matter* **1**, 6907 (1989)
142 C G Smith, M Pepper, R Newbury, H Ahmed, D G Hasko, D C Peacock, J E F Frost, D A Ritchie, G A C Jones, and G Hill, *J Phys Condens Matter* **1**, 6763 (1989)
143 H U Baranger and A D Stone, *Phys Rev Lett* **63**, 414 (1989)
144 C W J Beenakker and H van Houten, *Phys Rev Lett* **63**, 1857 (1989), and in *Electronic Properties of Multilayers and Low-Dimensional Semiconductor Structures* (J M Chamberlain, L Eaves, and J C Portal, eds), Plenum, New York (1990)
145 C J B Ford, S Washburn, M Buttiker, C M Knoedler, and J M Hong, *Phys Rev Lett* **62**, 2724 (1989)
146 A M Chang, T Y Chang, and H U Baranger, *Phys Rev Lett* **63**, 996 (1989)
147 L D Landau and E M Lifshitz, *Mechanics*, Pergamon, Oxford, (1976)
148 N S Kapany, in *Concepts of Classical Optics* (J Strong, ed), Freeman, San Francisco (1958)
149 M Buttiker, *Phys Rev B* **33**, 3020 (1986)
150 M Buttiker, *IBM J Res Dev* **32**, 63 (1988)
151 P H Beton, B R Snell, P C Main, A Neves, J R Owers-Bradley, L Eaves, M Henini, O H Hughes, S P Beaumont, and C D W Wilkinson, *J Phys Condens Matter* **1**, 7505 (1989)
152 L P Kouwenhoven, B J van Wees, W Kool, C J P M Harmans, A A M Staring, and C T Foxon, *Phys Rev B* **40**, 8083 (1989)
153 A D Benoit, C P Umbach, R B Laibowitz, and R A Webb, *Phys Rev Lett* **58**, 2343 (1987)
154 J Spector, H L Stormer, K W Baldwin, L N Pfeiffer, and K W West, *Surf Sci* **228**, 283 (1990)
155 K Nakamura, D C Tsui, F Nihey, H Toyoshima, and T Itoh, *Appl Phys Lett* **56**, 385 (1990)
156 C W J Beenakker, H van Houten, and B J van Wees, *Festkorperprobleme/Advances in Solid State Physics* **29**, 299 (1989)
157 P A M Benistant, G F A van de Walle, H van Kempen, and P Wyder, *Phys Rev B* **33**, 690 (1986)
158 M Buttiker, *IBM J Res Dev* **32**, 317 (1988)
159 M Buttiker, *Phys Rev B* **38**, 12724 (1988)
160 R E Prange and T -W Nee, *Phys Rev* **168**, 779 (1968), M S Khaikin, *Av Phys* **18**, 1 (1969)
161 A M Kosevich and I M Lifshitz, *Zh Eksp Teor Fiz* **29**, 743 (1955) [*Sov Phys JETP* **2**, 646 (1956)]
162 J R Barker and D K Ferry, *Solid State Electron* **23**, 519 (1980), *ibid* , **23**, 531 (1980), D K Ferry and J R Barker, *Solid State Electron* **23**, 545 (1980)
163 W Hansch and M Mima-Mattausch, *J Appl Phys* **60**, 650 (1986)
164 E M Conwell and M O Vassell, *IEEE Trans Electron Devices* **ED-13**, 22 (1966)
165 J M Shannon, *IEEE J Solid State Electron Devices*, **3**, 142 (1979)
166 J R Hayes, A F J Levi, and W Wiegman, *Phys Rev Lett* **54**, 1570 (1985)
167 J R Hayes and A F J Levi, *IEEE J Quantum Electr* **22**, 1744 (1986)
168 M Heiblum, M I Nathan, D C Thomas, and C M Knoedler, *Phys Rev Lett* **55**, 2200 (1985)
169 A Palevski, M Heiblum, C P Umbach, C M Knoedler, A N Broers, and R H Koch, *Phys Rev Lett* **62**, 1776 (1989)

170 A A Palevski, C P Umbach, and M Heiblum, *Appl Phys Lett* **55**, 1421 (1989)
171 U Sivan, M Heiblum, and C P Umbach, *Phys Rev Lett* **63**, 992 (1989)
172 P Hawrylak, G Eliasson, and J J Quinn, *Phys Rev B* **37**, 10187 (1988)
173 R Jalabert and S Das Sarma, *Solid State Electron* **32**, 1259 (1989), *Phys Rev B* **40**, 9723 (1989)
174 L Esaki and R Tsu, *IBM J Res Dev* **14**, 61 (1970)
175 L L Chang, L Esaki, and R Tsu, *Appl Phys Lett* **24**, 593 (1974)
176 L Esaki, *Rev Mod Phys* **46**, 237 (1974)
177 S Y Chou, D R Allee, R F W Pease, and J S Harris, Jr, *Appl Phys Lett* **55**, 176 (1989)
178 R Landauer, in *Nonlinearity in Condensed Matter*, Springer, Berlin (1987)
179 C B Duke, *Tunneling in Solids*, Academic Press, New York (1969)
180 R F Kazarinov and S Luryi, *Appl Phys Lett* **38**, 810 (1981)
181 P Hu, *Phys Rev B* **35**, 4078 (1987)
182 D Lenstra and R T M Smokers, *Phys Rev B* **38**, 6452 (1988)
183 L I Glazman and A V Khaetskii, *Pis'ma Zh Teor Fiz* **48**, 546 (1988) [*JETP Lett* **48**, 591 (1988)], *Europhys Lett* **9**, 263 (1989)
184 R J Brown, M J Kelly, M Pepper, H Ahmed, D G Hasko, D C Peacock, J E F Frost, D A Ritchie, and G A C Jones, *J Phys Condens Matter* **1**, 6285 (1989)
185 M J Kelly, *J Phys Condens Matter* **1**, 7643 (1989)
186 J K Jain and S A Kivelson, *Phys Rev B* **37**, 4276 (1988)
187 L Bliek, E Braun, G Hein, V Kose, J Niemeyer, G Weimann, and W Schlapp, *Semiconductor Sci Techn* **1**, 1101 (1986)
188 B E Kane, D C Tsui, and G Weimann, *Phys Rev Lett* **59**, 1353 (1987)
189 P G N de Vegvar, A M Chang, G Timp, P M Mankiewich, J E Cunningham, R Behringer, and R E Howard, *Phys Rev B* **36**, 9366 (1987)
190 P M Mensz and D C Tsui, preprint
191 P C van Son, G H Kruithof, and T M Klapwijk, *Surf Sci* **229**, 57 (1990), P C van Son and T M Klapwijk, *Eur Phys Lett* **12**, 429 (1990)
192 R Landauer, *Phil Mag* **21**, 863 (1970)
193 R Landauer, *Z Phys B* **21**, 247 (1975)
194 R Landauer, *IBM J Res Dev* **32**, 306 (1988)
195 J R Kirtley, S Washburn, and M J Brady, *Phys Rev Lett* **60**, 1546 (1988)
196 P C van Son, H van Kempen, and P Wyder, *J Phys F* **17**, 1471 (1987)
197 L I Glazman and M Jonson, *J Phys Condens Matter* **1**, 5547 (1989)
198 S Komiyama, H Hirai, S Sasa, and T Fuji, *Solid State Comm* **73**, 91 (1990)
199 S Komiyama, H Hirai, S Sasa, and S Hiyamizu, *Phys Rev B* **40**, 12566 (1989)
200 B W Alphenaar, P L McEuen, R G Wheeler, and R N Sacks, *Phys Rev Lett* **64**, 677 (1990)
201 B I Halperin, *Phys Rev B* **25**, 2185 (1982)
202 A H MacDonald and P Streda, *Phys Rev B* **29**, 1616 (1984), P Streda, J Kucera, and A H MacDonald, *Phys Rev Lett* **59**, 1973 (1987)
203 R B Laughlin, *Phys Rev B* **23**, 5632 (1981)
204 R F Kazarinov and S Luryi, *Phys Rev B* **25**, 7626 (1982), S Luryi, in *High Magnetic Fields in Semiconductor Physics* (G Landwehr, ed), Springer, Berlin (1987)
205 S V Iordansky, *Solid State Comm* **43**, 1 (1982)
206 R Joynt and R E Prange, *Phys Rev B* **29**, 3303 (1984)
207 R E Prange, in *The Quantum Hall Effect* (R E Prange and S M Girvin, eds), Springer, New York (1987)
208 L I Glazman and M Jonson, *Phys Rev B* **41**, 10686 (1990)
209 T Martin and S Feng, *Phys Rev Lett* **64**, 1971 (1990)
210 J K Jain, unpublished
211 S Komiyama and H Hirai, *Phys Rev B* **40**, 7767 (1989)
212 C W J Beenakker, *Phys Rev Lett* **64**, 216 (1990)

213 A M Chang and J E Cunningham, *Solid State Comm* **72**, 651 (1989)
214 L P Kouwenhoven, B J van Wees, N C van der Vaart, C J P M Harmans, C E Timmering, and C T Foxon, *Phys Rev Lett* **64**, 685 (1990)
215 J A Simmons, H P Wei, L W Engel, D C Tsui, and M Shayegan, *Phys Rev Lett* **63**, 1731 (1989)
216 G Timp, R Behringer, J E Cunningham, and R E Howard, *Phys Rev Lett* **63**, 2268 (1989)
217 D C Tsui, H L Stormer, and A C Gossard, *Phys Rev Lett* **48**, 1559 (1982), D C Tsui and H L Stormer, *IEEE J Quantum Electr* **22**, 1711 (1986), T Chakraborty and P Pietilainen, *The Fractional Quantum Hall Effect*, Springer, Berlin (1988)
218 R B Laughlin, *Phys Rev Lett* **50**, 1395 (1983)
219 M E Cage, in *The Quantum Hall Effect* (R E Prange and S M Girvin, eds), Springer, New York (1987)
220 M Buttiker, *Surf Sci* **229**, 201 (1990)
221 U Sivan, Y Imry, and C Hartzstein, *Phys Rev B* **39**, 1242 (1989)
222 U Sivan, and Y Imry, *Phys Rev Lett* **61**, 1001 (1988)
223 H Z Zheng, K K Choi, D C Tsui, and G Weimann, *Phys Rev Lett* **55**, 1144 (1985)
224 K von Klitzing, G Ebert, N Kleinmichel, H Obloh, G Dorda, and G Weimann, *Proc 17th International Conference on the Physics of Semiconductors* (J D Chadi and W A Harrison, eds), Springer, New York (1985)
225 R J Haug and K von Klitzing, *Europhys Lett* **10**, 489 (1989)
226 Y Aharonov and D Bohm, *Phys Rev* **115**, 485 (1959)
227 Y Imry and R A Webb, *Sci Am*, **259**, 56 (1989)
228 S Washburn and R A Webb, *Adv Phys* **35**, 375 (1986)
229 A G Aronov and Yu V Sharvin, *Rev Mod Phys* **59**, 755 (1987)
230 G Timp, A M Chang, J E Cunningham, T Y Chang, P Mankiewich, R Behringer, and R E Howard, *Phys Rev Lett* **58**, 2814 (1987), A M Chang, K Owusu-Sekyere, and T Y Chang, *Solid State Comm* **67**, 1027 (1988), A M Chang, G Timp, J E Cunningham, P M Mankiewich, R E Behringer, R E Howard, and H U Baranger, *Phys Rev B* **37**, 2745 (1988), G Timp, P M Mankiewich, P DeVegvar, R Behringer, J E Cunningham, R E Howard, H U Baranger, and J K Jain, *Phys Rev B* **39**, 6227 (1989)
231 C J B Ford, T J Thornton, R Newbury, M Pepper, H Ahmed, C T Foxon, J J Harris, and C Roberts, *J Phys C* **21**, L325 (1988)
232 D A Wharam, M Pepper, R Newbury, H Ahmed, D G Hasko, D C Peacock, J E F Frost, D A Ritchie, and G A C Jones, *J Phys Condens Matter* **1**, 3369 (1989)
233 J K Jain and S Kivelson, *Phys Rev B* **37**, 4111 (1988)
234 E N Bogachek and G A Gogadze, *Zh Eksp Teor Fiz* **63**, 1839 (1972) [*Sov Phys JETP* **36**, 973 (1973)]
235 N B Brandt, D V Gitsu, A A Nikolaevna, and Ya G Ponomarev, *Zh Eksp Teor Fiz* **72**, 2332 (1977) [*Sov Phys JETP* **45**, 1226 (1977)], N B Brandt, D B Gitsu, V A Dolma and Ya G Ponomarev, *Zh Eksp Teor Fiz* **92** 913 (1987) [*Sov Phys JETP* **65**, 515 (1987)]
236 Y Isawa, *Surf Sci* **170**, 38 (1986)
237 A Tonomura, *Rev Mod Phys* **59**, 639 (1987)
238 R J Brown, C G Smith, M Pepper, M J Kelly, R Newbury, H Ahmed, D G Hasko, J E F Frost, D C Peacock, D A Ritchie, and G A C Jones, *J Phys Condens Matter* **1**, 6291 (1989)
239 J K Jain, *Phys Rev Lett* **60**, 2074 (1988)

CHAPTER 3

When Does a Wire Become an Electron Waveguide?

G Timp

AT&T BELL LABORATORIES
HOLMDEL NEW JERSEY

I INTRODUCTION 113
II TWO-TERMINAL RESISTANCE OF AN ELECTRON WAVEGUIDE 117
1 *How to Make an Electron Waveguide* 117
2 *Measurements of the Two-Terminal Resistance* 123
3 *Theoretical Estimate of the Two-Terminal Resistance* 127
4 *Discussion of the Two-Terminal Results* 132
III THREE TERMINAL RESISTANCE 136
5 *Theoretical Estimates of the Three-Terminal Resistance* 136
6 *Measurements of the Three-Terminal Resistance* 137
7 *Discussion of the Three-Terminal Results* 142
IV FOUR-TERMINAL RESISTANCE 143
8 *Theoretical Estimates of the Four-Terminal Magnetoresistance* 143
9 *Four-Terminal Resistance Measurements for* $HeW^2/hc < 1$ 147
10 *Four-Terminal Resistance Measurements for* $1 < HeW^2/hc < 10$*—the Integer Quantized Hall Effect* 164
11 *The Magnetoresistance of an Annulus in the Quantized Hall Regime* 167
12 *Four-Terminal Resistance Measurements for* $HeW^2/hc > 10$*—the Fractional Quantized Hall Effect* 174
V SUMMARY AND OUTLOOK 182
ACKNOWLEDGMENTS 185
REFERENCES 185

I. Introduction

The low-temperature electrical resistance of a high-mobility, semiconducting wire 100 nm wide, 10 nm thick, and 200 nm long is nonlinear, it does not scale with length, it can be quantized as a function of the width, it depends upon the leads used to measure it, and it is nonlocal, i e , the current at one point in the wire depends not only on the electric field at that point, but on electric fields micrometers away Classical models for the resistance cannot

ISBN 0 12 752135 6