Coherent electron focusing with quantum point contacts in a two-dimensional electron gas

Houten, H. van; Beenakker, C.W.J.; Williamson, J.G.; Broekaart, M.E.I.; Loosdrecht, P.H.M. van; Wees, B.J. van; ... ; Harris, J.J.

Version:	Not Applicable (or Unknown)
License:	
Downloaded from:	https://hdl.handle.net/1887/3332

Note: To cite this publication please use the final published version (if applicable).

Coherent electron focusing with quantum point contacts in a two-dimensional electron gas

H. van Houten,* C. W. J. Beenakker, J G Williamson, M. E. I Broekaart, and P. H. M. van Loosdrecht[†‡]

Philips Research Laboratories, Postbox 80 000, NL-5600 JA Eindhoven, The Netherlands

B. J. van Wees and J E. Mooij

*Department of Applied Physics, Delft University of Technology,
Postbox 5046, NL-2600 GA Delft, The Netherlands*

C T Foxon and J J Harris

Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, United Kingdom

(Received 14 November 1988)

Transverse electron focusing in a two-dimensional electron gas is investigated experimentally and theoretically for the first time. A split Schottky gate on top of a GaAs-Al$_x$Ga$_{1-x}$As heterostructure defines two point contacts of variable width, which are used as injector and collector of ballistic electrons As evidenced by their quantized conductance, these are quantum point contacts with a width comparable to the Fermi wavelength At low magnetic fields, skipping orbits at the electron-gas boundary are directly observed, thereby establishing that boundary scattering is highly specular Large additional oscillatory structure in the focusing spectra is observed at low temperatures and for small point-contact size This new phenomenon is interpreted in terms of interference of coherently excited magnetic edge states in a two-dimensional electron gas A theory for this effect is given, and the relation with nonlocal resistance measurements in quantum ballistic transport is discussed It is pointed out, and experimentally demonstrated, that four-terminal transport measurements in the electron-focusing geometry constitute a determination of either a generalized longitudinal resistance or a Hall resistance At high magnetic fields the electron-focusing peaks are suppressed, and a transition is observed to the quantum Hall regime The anomalous quantum Hall effect in this geometry is discussed in light of a four-terminal resistance formula

I. INTRODUCTION

A magnetic field can be used to focus electron beams in vacuum The motion of Bloch electrons in the solid state on length scales small compared to the (elastic) mean free path l_c is similar to the motion in vacuum. One then speaks of ballistic transport Focusing of ballistic electrons has been studied extensively in metals In such experiments, point contacts small compared to l_c are employed to inject ballistic electrons at the Fermi level, so that an essentially monoenergetic ("monochromatic"), but divergent, electron beam is created. The focusing action of a magnetic field can then be detected in an elegant way by using a second point contact, which acts as a collector or voltage probe (drawing no net current). This technique, pioneered by Sharvin[1] and Tsoi,[2] is a powerful tool to obtain information on the shape of the Fermi surface,[3] on electron-phonon interaction,[4] and on surface scattering The surface can be a free surface of a crystal,[5] or a metal-superconductor interface[6] (Andreev reflection[7]) In metals, electron focusing is essentially a classical transport phenomenon, because of the small Fermi wavelength λ_F (typically 0.5 nm)

Due to the large Fermi wavelength in the two-dimensional electron gas (2D EG) in GaAs-Al$_x$Ga$_{1-x}$As heterostructures (typically 40 nm), the quantum ballistic transport regime has recently become accessible. We have fabricated 2D EG point contacts, with a variable size comparable to λ_F. The discovery of the quantized conductance of such *quantum point contacts* has been reported elsewhere.[8,9] Here we employ quantum point contacts as monochromatic point sources in an electron-focusing geometry (see Fig. 1). The current through the injector is kept fixed, while the collector voltage is measured as a function of the perpendicular magnetic field. Electrons injected in a direction perpendicular to the 2D EG boundary can reach the collector directly, or after specular reflections from the boundary, for magnetic fields B such that the point contact separation L is an integer multiple of the classical cyclotron diameter $2\hbar k_F/eB$ (with $k_F \equiv 2\pi/\lambda_F$ the Fermi wave vector) This occurs when B is an integer multiple of the focusing field B_{focus}, given by

$$B_{\text{focus}} = \frac{2\hbar k_F}{eL} . \tag{1}$$

Electrons injected within a small angle around the perpendicular direction are, for these field values, focused onto the collecting point contact, as indicated in Fig. 1. Trajectories at glancing angles do not contribute to the focusing, but are observed as a background signal. A plot of collector voltage versus field thus exhibits a series of equidistant peaks at $B = iB_{\text{focus}}$ ($i = 1, 2, 3, \ldots$), which is called a focusing spectrum. As we shall show, focusing experiments with quantum point contacts yield qualitatively different results than related experiments in the
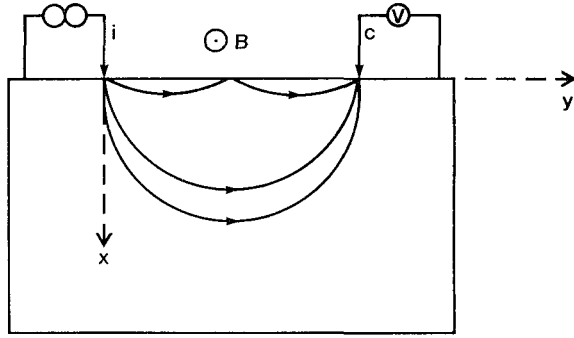
FIG. 1. Schematic arrangement for transverse electron focusing. Trajectories leaving the injector around the normal direction are focused onto the collecting point contact. Glancing angle trajectories contribute to the background.

classical ballistic transport regime characteristic for metals. This justifies a new name: *coherent electron focusing*.

From a different point of view, electron focusing is a quantum ballistic transport phenomenon, related to those observed in narrow multiprobe channels defined in a high-mobility 2D EG.[10] It has recently become clear that the voltage probes used to measure the potential drop along such channels have an inextricable influence on the transport. As we demonstrate in this paper, the injecting current contact is also of essential importance. Transport measurements in the electron-focusing geometry offer a striking demonstration of the fact that quantum ballistic transport is governed by the interaction of the relevant quantum states with current and voltage probes, rather than by a local resistivity, as in the classical diffusive transport regime.

We summarize our main results as follows. (1) Ballistic injection of electrons in a 2D EG is realized for the first time. (2) Skipping orbits are directly observed, thereby demonstrating conclusively that the scattering from the 2D EG boundary is predominantly specular. (3) Interference structure in the focusing spectra is observed at low temperatures and for small point-contact size, demonstrating the coherent nature of the focusing process. (4) A theoretical description of coherent electron focusing in a 2D EG is provided, reproducing the essential features of the experiment. (5) Four-terminal electron-focusing experiments are identified as either generalized longitudinal or Hall resistance measurements, as demonstrated by the transition from focusing peaks to quantum Hall plateaus at high magnetic fields. A four-terminal resistance formula is derived within the frame work of the Landauer-Büttiker[11] formalism, which relates the Hall resistance to the *contact resistances* of injector and collector. The anomalous quantum Hall effect in the electron-focusing geometry[12] is discussed in the light of this resistance formula.

The paper is organized as follows. In Sec. II details on the sample and its fabrication are described, and in Sec. III the properties of single quantum point contacts are briefly discussed. In Sec. IV the experimental results obtained in the electron-focusing geometry are presented. The theoretical analysis of classical and coherent electron

focusing in a 2D EG is the subject of Sec. V and Appendix A–D. A discussion of the results is given in Sec. VI, which concludes by giving indications for future extensions.

Some of our experimental and theoretical results have been briefly reported previously, in Refs. 13 and 14, respectively.

II. SAMPLES AND EXPERIMENTAL DETAILS

Conceptually, point contacts can be thought of as small orifices in a thin insulating layer, separating bulk metallic conductors (with l_e much larger than the size of the orifice). In practice, point contacts are usually[15] fabricated by pressing a metal needle on a metallic single crystal, followed by spot welding. Even though some surface damage is introduced, ballistic transport has successfully been studied in this way in a variety of metals.[16] One limitation of this technique is that the size of the point contacts is not continuously variable.

For a 2D EG the above-mentioned method cannot be used, since the electron gas is confined at the GaAs-$Al_xGa_{1-x}As$ interface in the sample interior. Our point contacts, defined by a split Schottky-gate lateral depletion technique,[17] are essentially short and narrow constrictions[18] in the 2D EG. The starting point for the fabrication is a GaAs-$Al_xGa_{1-x}As$ heterostructure grown by molecular-beam epitaxy. The carrier concentration as obtained from the Shubnikov–de Haas oscillations in the magnetoresistance is $n_s = 3.5 \times 10^{15}$ m^{-2}, and the mobility $\mu = 90$ m^2/V s, leading to a transport mean free path $l_e \approx 9$ μm. A standard mesa-etched Hall-bar geometry was subsequently defined by wet etching. The split-gate geometry on top of this Hall bar is schematically indicated in Fig. 2(a). Electron-beam lithography is used to write the fine details of the gate. A typical gate structure is shown in the scanning electron micrograph of Fig. 2(b). The actual 2D EG boundary is a depletion-potential wall underneath the gate pattern. Note that the depletion potential extends laterally beyond the gate pattern for high gate voltages (up to about 150 nm). As indicated in Fig. 2(a), the fine details of the gate are connected to broader gates, defined by optical lithograpy. These gates run over the wet-etched sides of the mesa towards a bonding pad (the mesa sides are inclined at an angle of 45° with respect to the substrate). No gate isolation is needed, even on the mesa sides, because of surface depletion of the 2D EG. (Leakage currents at low temperatures are below 10^{-10} A under normal operating conditions.) By increasing the (negative) voltage on the Schottky gate, the electron gas underneath the gate structure is depleted. 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 2D EG regions are thus electrically isolated from the rest of the 2D EG in the Hall bar, apart from the narrow constrictions or point contacts under the openings of the gate. We note that electron-beam-induced damage incurred during the fabrication process will be concentrated under the gates, where conduction does not take place. A further
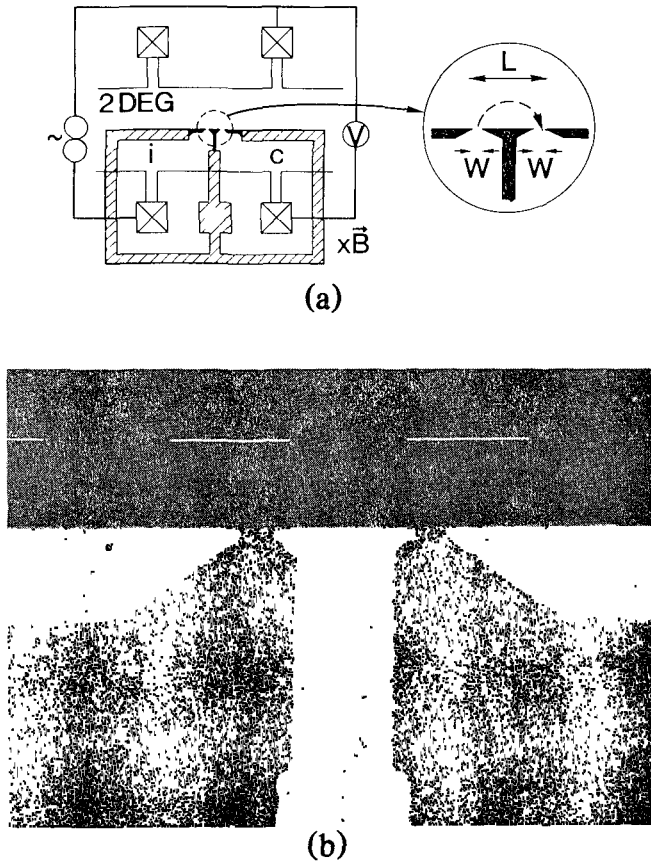
(a)

(b)

FIG. 2. (a) Schematic layout of the double-point-contact device for the electron-focusing experiments. The crossed squares are Ohmic contacts to the 2D EG. The split gate (shaded) separates injector (*i*) and collector (*c*) areas from the bulk 2D EG. The dashed line indicates an electron trajectory in a magnetic field. (b) Scanning electron micrograph of the fine details of the gate structure defining a double-point-contact device. The bar denotes a length of 1 μm. In this device the point-contact separation is 1.5 μm. Most experiments discussed in the text were performed on a device with a 3.0-μm point-contact separation.

One of the contacts can be used as a ballistic electron injector, while the other point contact acts as a collector for the electrons which are focused by a magnetic field. (We remark that due to fabrication tolerances the injecting and collecting point contacts have a different width for a given gate voltage. One of the devices has been constructed in such a way that injector and collector can be adjusted separately.) A low-frequency ac lock-in technique is used to measure the ratio of the collector voltage to the injected current. Several Ohmic contacts (alloyed Au-Ge-Ni) on the sides of the Hall bar [see Fig. 2(a)] allow the electron focusing to be measured four-terminally as well as three-terminally. Both experiments have been performed (see Sec. IV). Here we already note that in three-terminal measurements of the focusing a background voltage is measured in series with the collector voltage, mainly because of the Ohmic contact resistance, and additionally because of a small diffusive background resistance (of order 100 Ω) originating in the wide 2D EG regions. Such a background resistance is also seen in two-terminal measurements on a single point contact. At temperatures around 1 K the Ohmic contact resistance is small, but upon lowering the temperature to the mK region an anomalous rise of the zero-field alloyed Ohmic contact resistance (up to 3 kΩ) is observed. This effect can be suppressed by a weak magnetic field (0.1 T) leading to a negative magnetoresistance in these measurements. We attribute this effect to quasi-one-dimensional localization[17,19] in the disordered Ohmic contact regions, presumably related to narrow meandering conduction paths in these regions. For fields beyond 0.1 T, or for temperatures above 300 mK, this effect does not influence the results. Moreover, it can be fully eliminated in four-terminal measurements of the focusing.

Most of the experiments were performed on the device with a point-contact separation L of 3.0 μm described above. Additional measurements have been made on a similar device with a smaller L of 1.5 μm, to check the expected $1/L$ dependence of the magnetic field scale. Unless stated otherwise, our results refer to the $L = 3.0$ μm device.

increase of the gate voltage forces both constrictions to become progressively narrower until they are fully pinched off. By this technique it is possible to define point contacts with variable width W. A nice feature is that the point-contact separation L remains approximately constant (3.0 μm) when the width is varied. We mention that 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 strong reverse biases lead to a shift in the depletion threshold. The details of the focusing spectra are as a consequence not reproducible after such changes, although they reproduce very well if the sample is kept cold, and the gate voltage is not strongly varied. We also remark that, as a secondary effect, the "mirror" constituted by the electron-gas boundary between the two point contacts shifts if the gate voltage is changed. Typically, this shift is of the order of 100 nm/V (as estimated from the pinch-off characteristics of the constrictions).

III. TRANSPORT THROUGH SINGLE QUANTUM POINT CONTACTS

The transport properties of the injecting and collecting point contacts are relevant to the electron-focusing experiments reported in this paper. In this section we therefore give a brief discussion of transport through single quantum point contacts.[8,9,20] For classical ballistic point contacts $(l_e \gg W \gg \lambda_F)$ in a 2D EG the two-terminal conductance G is given by[8]

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

for an infinite square-well confining potential in the point contact. An increase of the negative gate voltage leads to a decrease of both the point-contact width W and of the electron density $n_c = k_F^2/2\pi$ in the constriction. We would therefore expect a smooth decrease of G with increasing negative gate voltage. As reported in Refs. 8

and 9, the conductance of quantum point contacts deviates in an interesting way from this classical formula, in that plateaus in G as a function of gate voltage are observed. The point-contact conductance is approximately quantized in multiples of $2e^2/h$, due to quantization of the transverse momentum in the constriction. As can be shown semiclassically,[8] or by means of the two-terminal Landauer formula,[11,21,22] the conductance in a plateau region is given by

$$G = \frac{2e^2}{h} T_c = \frac{2e^2}{h} N_c , \qquad (3)$$

with T_c the transmission probability through the con-

striction and N_c the largest integer smaller than $k_F W/\pi$ The second equality in Eq (3) assumes that no backscattering occurs in the constriction If quantization can be ignored $(N_c \gg 1)$, the classical result (2) for G is recovered. Under the conditions of our electron-focusing experiment, N_c is a small number, so that the quantum nature of the point contacts is important.

We now turn to the effect of a perpendicular magnetic field B on the two-terminal conductance of a single point contact Equation (3) remains valid in a field, which has only the effect of reducing the number of occupied subbands N_c in the constriction Ignoring the discreteness of N_c, one finds, for an infinite confining potential,[20]

$$N_c(B) = \begin{cases} (k_F l_{cycl}/\pi)\{ \arcsin(W/2l_{cycl}) + (W/2l_{cycl})[1-(W/2l_{cycl})^2]^{1/2}\} & \text{if } W < 2l_{cycl} , \\ k_F l_{cycl}/2 \text{ if } W > 2l_{cycl} , \end{cases}$$

$$\qquad (4a)$$
$$\qquad (4b)$$

where $l_{cycl} \equiv \hbar k_F/eB$ is the cyclotron radius. A derivation of Eq. (4) is given in Appendix A. The magnetic depopulation of subbands begins at fields where the cyclotron diameter is of the order of the point-contact width W. Accordingly, the conductance of a single quantum point contact decreases stepwise if the magnetic field is increased, as found experimentally.[9,20] As shown in Ref. 20 these experimental data, together with Eq. (4), yield estimates for both the width W and the electron density n_c in the constriction.

In Fig. 3 results are given for the resistance as a function of magnetic field for a range of gate voltages V_g (at a temperature of 50 mK). As mentioned in Sec II, a small diffusive background resistance originating in the wide 2D EG regions is measured in series with the ballistic point-contact resistance. Shubnikov−de Haas oscillations in the background resistance can be observed, with a characteristic $1/B$ periodicity[23] (see, e.g., the curve for $V_g = -0.75$ V in Fig. 3 in the field region from 0.6 to 3

T). Notice also the negative magnetoresistance peak around zero field, which originates in the Ohmic contacts (see Sec II) In a range of gate voltages at high magnetic fields we see magnetoresistance oscillations which are periodic in B itself,[24] and are reminiscent of the Aharonov-Bohm effect in rings. These oscillations vanish for very narrow constrictions (cf. the curve for $V_g = -0.75$ V in Fig 3), and also at higher temperatures (1 K) or injection voltages As discussed in Ref. 24, a possible mechanism for this quantum-interference effect is tunneling between edge states across the point contact at the potential step at the entrance and exit of the constriction. (The potential step is a result of the reduced electron density in the constriction) The magnetoresistance data in Fig. 3 are for one of the point contacts of the device with $L = 3.0$ μm Other point contacts had a similar magnetoresistance, but without periodic oscillations.

For the low-field electron-focusing experiments, the most important conclusion to be drawn from Fig 3 is that the two-terminal resistance of a narrow point contact (corresponding to a large negative gate voltage) is essentially constant over a broad field range.

IV. ELECTRON FOCUSING: EXPERIMENTAL RESULTS

A. Temperature and gate-voltage dependence

In Fig 4 the collector voltage as a function of magnetic field is shown for the device with 3.0-μm point-contact separation, at temperatures between 7 K and 50 mK At the higher temperatures a clear set of equidistant peaks is observed, associated with multiple specular reflections from the 2D EG boundary. Classically, peaks in the collector voltage are expected to occur at values of the magnetic field such that the point-contact separation L is an integer multiple of the cyclotron orbit diameter $2\hbar k_F/eB$ Substituting $L = 3.0$ μm and $k_F = (2\pi n_s)^{1/2} = 1.5 \times 10^8$ m^{-1} (with n_s determined from the Shubnikov−de Haas oscillations), we expect focusing peaks at multiples of
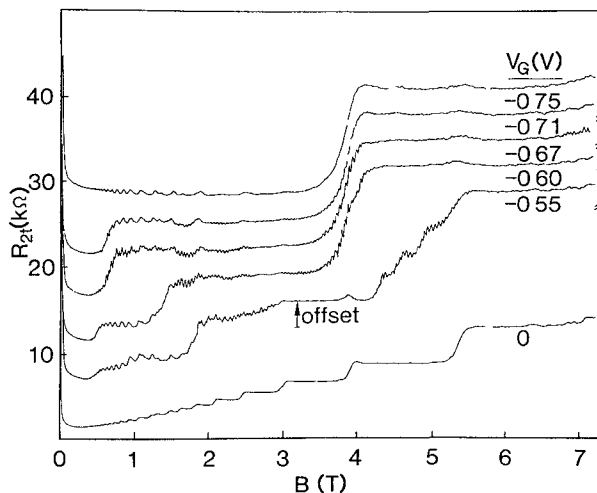
FIG 3 Two-terminal magnetoresistance of a single point contact at 50 mK for a series of gate voltages The curves have been offset vertically for clarity

$B_{\text{focus}} = 0.066$ T from Eq. (1). These field values are indicated by arrows in Fig. 4. The observed peak positions at the higher temperatures agree within the experimental uncertainties with this prediction. The observation of electron focusing establishes that ballistic injection of electrons in the 2D EG has been realized in this experiment. We recall that the *i*th peak is due to electrons which have made $i - 1$ specular reflections at the boundary. The large number of peaks observed thus demonstrates that the reflections from the 2D EG boundary are predominantly specular.

The experimental focusing spectrum shown in Fig. 4 differs in several ways from the classical focusing spectra familiar from similar experiments in metals. Classically, for purely specular boundary scattering the peak height is expected to be independent of the peak number *i*, and if the scattering is only partially specular the peak height should decrease with *i* (see Sec. V). In our experiment, however, the peak height depends nonmonotonously on *i*. Moreover, as shown in Fig. 4, at low temperatures a reproducible fine structure is superimposed on the classical focusing peaks. This fine structure is smeared if the injection voltage is increased. The data in Fig. 4 have been obtained with an injection voltage below $k_B T / e$, which for a temperature of 50 mK corresponds to 4 μV. The voltage measured on the collector was typically a factor of 10 or more lower than the voltage drop across the injector. The signal-to-noise ratio was still acceptable under these conditions. If the injector voltage is considerably increased above this value, the fine structure is

smeared, analogous to the effect of a temperature increase. A similar smearing of the spectra occurs if the point contacts are widened by reducing the negative gate voltage, as shown in Fig. 5. In this experiment the injection voltage was kept low, and the temperature was 50 mK. The position of the classical focusing peaks is unchanged, as expected, since the point-contact separation is essentially unaffected by the gate voltage. The width of the point contacts can, in principle, be determined from their quantized conductance in a magnetic field.[20] A difficulty with this analysis is that the electron density in the constriction also changes if the gate voltage is varied. In the experiments discussed above (on the $L = 3.0$ μm device), the width of the injecting point contact is estimated to be 20 nm at the highest negative gate voltage studied, while the width of the collector is appreciably wider (of the order of 80 nm). The ultimate resolution attained with this device is therefore limited by the finite point-contact size, rather than by the temperature or the injection voltage. This is substantiated by the fact that a limited temperature increase up to 300 mK, or a corresponding increase in injection voltage, have a negligible influence on the focusing spectra.

In the low-magnetic-field range of Fig. 4, the focusing spectra are characterized by classical focusing peaks with superimposed fine structure. At higher fields (beyond about 0.4 T) the collector voltage shows oscillations with a much larger amplitude than the low-field focusing peaks, and the resemblance to the classical focusing spectrum is lost. This is shown in Fig. 6, for two gate voltages. Notice that, although the spectra are well reproducible, they depend sensitively on the gate voltage. A Fourier transform of the spectra (see inset of Fig. 6) shows that the large-amplitude high-field oscillations have a dominant periodicity of 0.06 ± 0.01 T, which is approximately the same as the periodicity $B_{\text{focus}} = 0.066$ T [Eq. (1)] of the low-field focusing peaks. This dominant
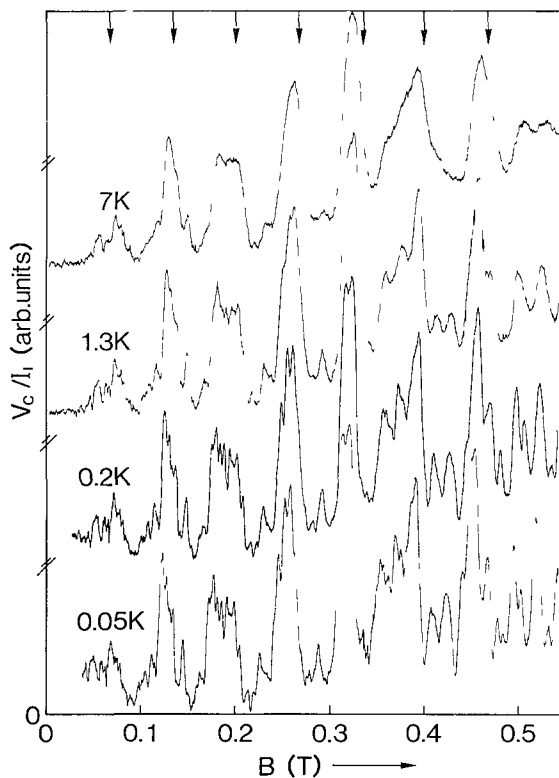

FIG. 4. Typical (three-terminal) electron-focusing spectra at temperatures between 7 K and 50 mK. Peak positions predicted by Eq. (1) are indicated by arrows.
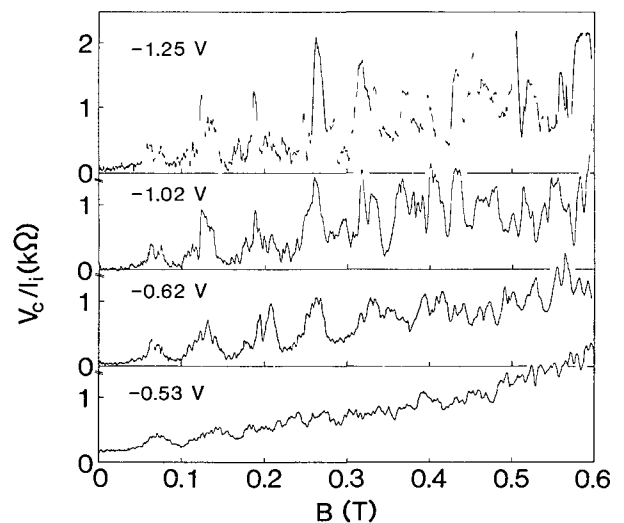

FIG. 5. Dependence of the electron-focusing spectra on the gate voltage (and thus on the point-contact width) at 50 mK. A large negative gate voltage corresponds to narrow injector and collector point contacts.
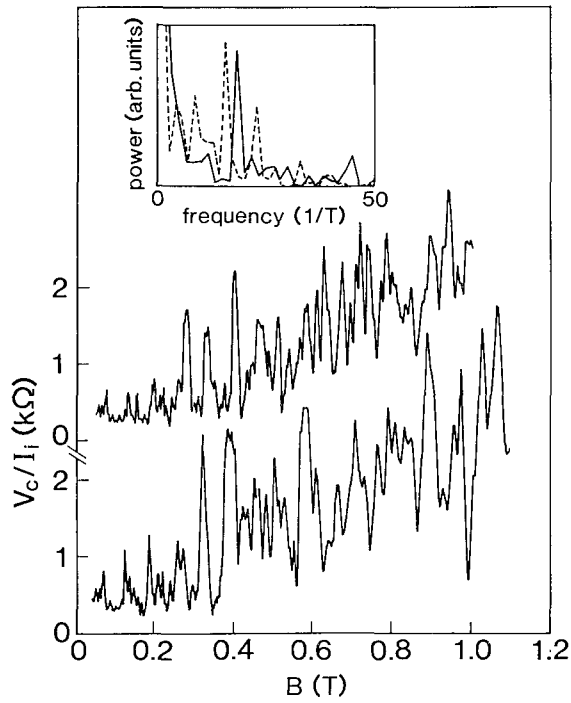
FIG. 6. Electron-focusing spectra at 50 mK for two gate voltages $V_g = -1.53$ V (lower trace) and -1.22 V (upper trace). The inset gives the Fourier-transform power spectrum of V_c for $B > 0.4$ T (dashed curve, $V_g = -1.53$ V; solid curve, $V_g = -1.22$ V).

FIG. 7. Electron-focusing spectrum at 50 mK for the device with 1.5-μm point-contact separation (all other data are for the $L = 3.0$ μm device). The inset gives the Fourier-transform power spectrum of V_c for $B > 0.8$ T.

periodicity is insensitive to changes in gate voltage. In Sec. V we will explain its origin in terms of quantum interference of coherently excited edge states in the 2D EG.

These data were obtained on the device with a point-contact separation $L = 3.0$ μm. In order to check the expected scaling of the periodicity with $1/L$, we also studied a device with $L = 1.5$ μm (and carrier density $n_s = 3.9 \times 10^{15}$ m^{-2}, estimated from the Shubnikov–de Haas oscillations). The focusing spectrum for this second device is shown in Fig. 7. The characteristic features above for the $L = 3.0$ μm device (Fig. 6) are reproduced in the $L = 1.5$ μm device, but on a field scale which is larger by approximately a factor of 2. From the first two focusing peaks we estimate $B_{\text{focus}} = 0.11 \pm 0.01$ T, which is somewhat smaller than the value of 0.14 T predicted for this device by Eq. (1). This discrepancy may be due in part to the uncertainty in the effective point-contact separation of the order of the split-gate opening (250 nm), which in this device is relatively large compared to the nominal point-contact separation (1.5 μm, which follows if we assume that the centers of both point contacts are in the middle of the openings in the gate). A Fourier transform of the high-field oscillations (inset in Fig. 7) shows that these have the same dominant periodicity as the low-field focusing peaks, consistent with the results obtained for the $L = 3.0$ μm device. The $L = 1.5$ μm device had a slightly different design, which allowed the injector and collector widths to be adjusted separately (the results of Fig. 7 were obtained with both point contacts having the smallest quantized conductance of $2e^2/h$). The increased resolution is most likely the reason for the much larger peak height in Fig. 7, compared with Fig. 6. (An
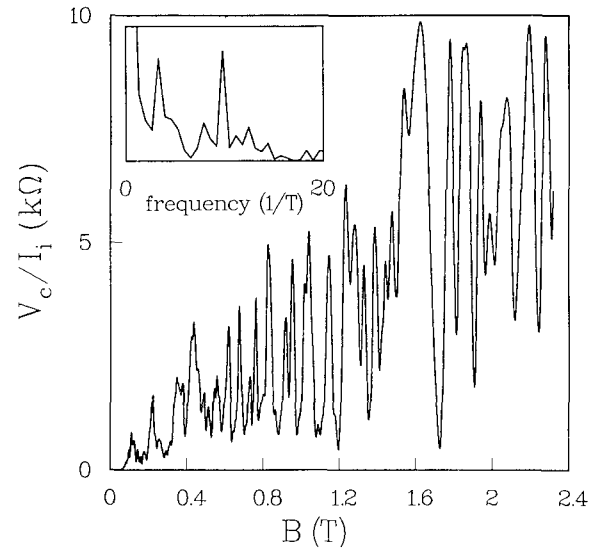
additional reason might be that the point contacts are closer with respect to the electron phase coherence length.) Notice that up to 95% modulations of the collector voltage are realized in this quantum-interference device.

B. Relation to nonlocal resistance measurements

The quantity measured in the electron-focusing experiments is the voltage difference between the collector and one of the Ohmic contacts attached to the wide 2D EG region, divided by the injected current. Depending on how the Ohmic contacts are connected [see Fig. 2(a)] this corresponds to a nonlocal Hall- or longitudinal-resistance measurement. This correspondence was never manifest in electron focusing in metals, presumably because of the more complicated 3D geometry. We therefore discuss this in some detail (cf. also Appendix D).

A three-terminal measurement in a 2D EG ideally yields for one magnetic field direction a purely longitudinal resistance, while for the other field direction the Hall resistance is obtained. In practice, also the voltage drop across the common current-carrying contact is measured in series. In three-terminal measurements of electron focusing this additional voltage drop, related to the Ohmic contact resistance, causes the negative magnetoresistance peak around $B = 0$ seen in Fig. 8(a). The origin of this effect has been discussed in Sec. II. This contact resistance is eliminated in four-terminal resistance measurements. A four-terminal measurement of electron focusing can be characterized as a *generalized Hall-resistance* measurement if an imaginary line connecting the voltage probes crosses a line connecting the current source and drain contacts, or a *generalized longitudinal-resistance* measurement if this is not the case. The data in Figs. 8(b) and 8(c) demonstrate this experimentally.

We now give a more detailed discussion of these re-

sults. The data in Fig. 8(a) have been obtained in a three-terminal configuration. For one field direction the focusing signal is seen to be superimposed on a rising background, with a slope corresponding to the Hall resistance (see below). For reverse fields beyond 0.4 T weak Shubnikov–de Haas oscillations are observable, arising from the 2D EG background resistance, which are characteristic for a longitudinal-resistance measurement. Apart from these oscillations, the reverse field signal is essentially independent of the magnetic field, because the 2D EG has no classical magnetoresistance [see also Fig. 8(b)].

Figure 8(b) is a four-terminal measurement of the generalized longitudinal resistance. As expected, the rising background associated with the Hall effect, and also the magnetoresistance peak around $B = 0$, have disappeared. Figure 8(c) shows the generalized Hall resistance. The straight line in the reverse field part of the focusing plot in Fig. 8(c) is indeed the normal classical Hall effect. From the slope of this line we find a Hall ratio of 1780 Ω/T, in close agreement with a calculated value using the

carrier density obtained from Shubnikov–de Haas data (from $n_s = 3.5 \times 10^{15}$ m^{-2} we find a Hall ratio $1/n_s e = 1790$ Ω/T). At $B = 0$ a sudden transition is seen from a linear to an approximately quadratic B dependence of the resistance (cf. the calculation of this transition in Sec. V). The electron-focusing experiment in this configuration is a *nonlocal* Hall-resistance measurement in the ballistic transport regime. Classically, the nonlocality arises because the collector is less than a mean free path away from the point-contact injector. Quantum mechanically, the resistance measurement is, in addition, nonlocal because the point-contact separation is less than the phase coherence length, which can appreciably exceed the mean free path. The Hall resistance measured in our electron-focusing experiment is alternatingly both larger and smaller than the classical Hall resistance, as a consequence of the electron focusing (cf. Sec. V). This also explains why alternatingly positive and negative voltages are seen in Fig. 8(b), because it is equivalent to Fig. 8(c) after subtraction of the classical Hall resistance. The possibility, in principle, of negative resistances ("uphill voltages") in a four-terminal measurement was emphasized by Büttiker[11] and Landauer.[21] The present experiment provides one simple physical mechanism for such an effect (an explicit calculation is given in Sec. V).

C. Reciprocity of injector and collector

In the diffusive transport regime, where a local resistivity can be defined, Onsager-Casimir[25] relations describe the symmetry of the components of the *resistivity* tensor in the presence of a magnetic field. The origin of these symmetries is microscopic time-reversal invariance. Also symmetry relations for *resistances* can be found, as discussed, for example, by van der Pauw.[26] In the present case of ballistic transport no local resistivity exists, but instead only resistances have a meaning. Büttiker[11] has derived a reciprocity relation for resistances, which holds in the nonlocal quantum transport regime of interest in this paper. He shows that

$$R_{12,34}(B) = R_{34,12}(-B) , \qquad (5)$$

where the two pairs of indices refer to the current and voltage leads respectively. This relation describes the reciprocity of resistances with interchanged current and voltage leads. As discussed above, electron focusing is a generalized resistance measurement, and accordingly Eq. (5) should also hold for these experiments. Indeed this is found to be the case as demonstrated by the data in Fig. 9, which were obtained after interchanging the roles of injector and collector. This experiment beautifully demonstrates the reciprocity relation (5) for the four-terminal phase-coherent resistance in the ballistic transport regime (see Ref. 27 for other experimental confirmations).

FIG. 8. Electron focusing at 50 mK for three measurement configurations, depicted in the insets. (a) Three-terminal measurement; (b) four-terminal generalized longitudinal-resistance measurement; (c) four-terminal generalized Hall-resistance measurement.

D. Transition to the quantum Hall regime

The analogy between electron-focusing and Hall-resistance measurements has a further interesting aspect, which we discuss in this subsection on the basis of results
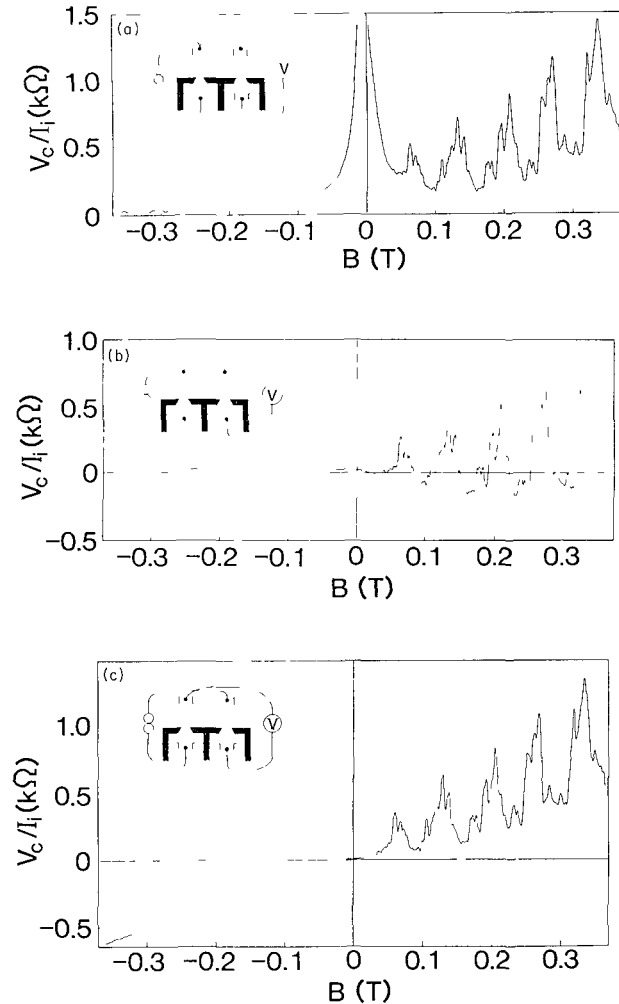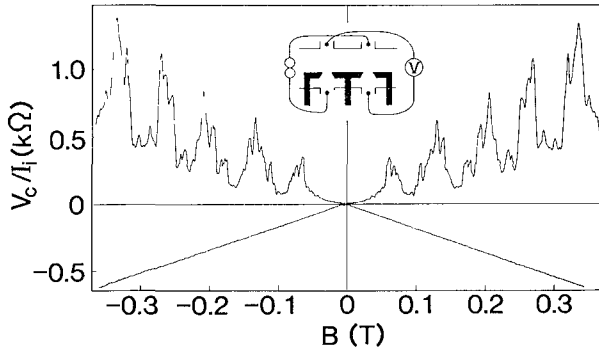
FIG. 9. Electron focusing in the generalized Hall-resistance configuration, as in Fig. 8(c). The two traces correspond to interchanged current and voltage leads, and demonstrate the injector-collector reciprocity which follows from Eq. (5).

obtained over a wider field range (up to 5 T) shown in Fig. 10. For $B > 2$ T focusing peaks are no longer observed, and instead quantum Hall plateaus[28] appear. The basic reason for this transition is that at such high fields the resolution necessary to distinguish subsequent focusing peaks is lost, since the cyclotron diameter is smaller than the point-contact size (an explicit calculation of the smearing of the focusing peaks is given in Sec. V). One might therefore expect the sample to be equivalent to a normal Hall bar with wide current and voltage probes, and consequently to observe the quantum Hall effect (in the generalized Hall-resistance configuration, or for three-terminal measurements with the appropriate field direction). Note that for typical point contacts in metals this limit is beyond reach, because of the much larger Fermi velocity, and the correspondingly larger cyclotron radius. Although the similarity between the result shown in Fig. 10 and the normal quantum Hall resistance is suggestive, significant deviations occur. In contrast to the quantum Hall effect observed in normal Hall-bar
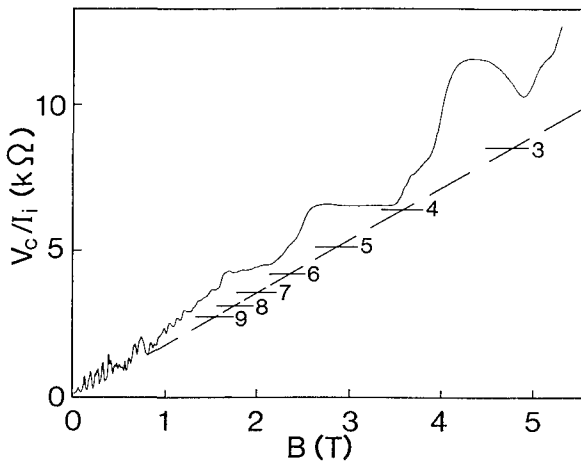

FIG. 10. Transition from weak-field electron focusing to high-field quantum Hall effect (in a three-terminal measurement). Quantum Hall-plateau values corresponding to h/ie^2 are indicated for $i = 3, 4, \ldots, 9$. The observed plateaus occur consistently at lower-magnetic-field values than expected from the classical Hall resistance of 1790 Ω/T, indicated by the dashed line.

geometries (where the plateaus are centered at the classical Hall resistance), the plateaus seen in Fig. 10 systematically occur at lower-magnetic-field values. Furthermore, at fields around 5 T an unusual oscillation is seen in the Hall resistance.

In order to understand the origin of such deviations, one has to take into account the potential barrier in the point contacts, resulting from the reduced local electron concentration. The argument[12] is as follows. The current-carrying states in high magnetic fields are edge states at the 2D EG boundary with Fermi energy E_F. The edge states with Landau level index n (referred to collectively as an edge *channel*) can only be transmitted across a potential barrier if their guiding center energy

$$E_G = E_F - (n - \tfrac{1}{2})\hbar\omega_c \tag{6}$$

exceeds the potential-barrier height (disregarding tunneling through the barrier). Here $\omega_c = eB/m$ is the cyclotron frequency, and the Zeeman spin splitting is ignored for simplicity. In the injector (in which the barrier has a height E_i), this condition is met for $N_i \approx (E_F - E_i)/\hbar\omega_c$ edge channels, while the collector (with barrier height E_c) is capable of transmitting $N_c \approx (E_F - E_c)/\hbar\omega_c$ channels. At the boundary of the 2D EG, however, a larger number of $N_L \approx E_F/\hbar\omega_c$ edge channels, equal to the number of bulk Landau levels in the 2D EG, are available for the current transport. The key point necessary for an understanding of the deviations seen in Fig. 10 is that, in the absence of inter-edge-channel scattering, the current along the 2D EG boundary from injector to collector is carried by only the first N_i of the N_L available edge channels. The selective population, and detection, of edge channels leads to deviations from the normal Hall resistance, consistent with the argument of Büttiker[29] that equilibration of edge channels by inelastic scattering plays a crucial role in establishing the quantum Hall effect.

These considerations can be put on a theoretical basis using the general Landauer-Büttiker formalism,[11] which, as shown in Appendix D, predicts for the electron-focusing geometry that the generalized Hall resistance $\mathcal{R}_H \equiv V_c/I_i$ is given by

$$\mathcal{R}_H = \frac{2e^2}{h} T_{i \to c} \frac{1}{G_i G_c} . \tag{7}$$

Here, $T_{i \to c}$ is the transmission probability from injector to collector, and G_i, G_c are the two-terminal conductances of the injecting and collecting point contacts. (It should be noted that "two-terminal" in this context refers to an idealized situation, where series resistance contributions due to the alloyed Ohmic contacts, or to finite resistivity in the wide 2D EG regions, is eliminated, as opposed to real two-terminal measurements such as shown in Fig. 3.) In high magnetic fields the guiding center energy is an approximate constant of the motion, provided the electrostatic potential varies slowly on the scale of the magnetic length $l_m \equiv (\hbar/eB)^{1/2}$. In this field regime Eq. (7) takes a very simple form if the barrier in one of the two point contacts is much higher than in the other, so that electrons which are transmitted over the highest barrier will have a negligible probability of being reflected at

the lowest barrier (in the absence of other sources of scattering in the 2D EG, which due to the close proximity of injector and collector is not an unrealistic assumption, as shown throughout this paper). Then $T_{l \rightarrow c}$ is dominated by the transmission probability over the highest barrier, and therefore by the smallest of the two-terminal conductances, $(2e^2/h)T_{l \rightarrow c} \approx \min\{G_l, G_c\}$ [cf. Eq. (3)]. Substitution into Eq. (7) gives the remarkable result,

$$\mathcal{R}_H \approx \frac{1}{\max\{G_l, G_c\}} , \qquad (8)$$

that *the high-field Hall resistance in the electron-focusing geometry can be expressed entirely in terms of contact resistances*. In particular, Eqs. (3) and (8) tell us that quantized values of $\mathcal{R}_H$ occur not at $h/2e^2 N_L$, as one would expect from the N_L Landau levels in the 2D EG, but at the larger value of $h/(2e^2 \max\{N_l, N_c\})$, determined by the largest number of edge channels above the barrier in either the injector or the collector. As demonstrated in Ref. 12, this anomalous behavior of the Hall resistance is indeed observed experimentally.

The enhancement of the Hall resistance predicted by Eq. (8) explains the shift of plateaus in Fig. 10 to lower magnetic field values. To put it differently, due to the reduced density in the point contacts, depopulation of the edge states occurs at lower magnetic fields than in the bulk of the 2D EG. The identification of the Hall resistance with a contact resistance in Eq. (8) also explains the large deviations from exact quantization. The slow oscillation in Fig. 10 at 5 T is indeed similar to that seen in two-terminal measurements of the magnetoresistance of the point contacts.[24]

The mechanism responsible for this slow oscillation is unclear, but may be related to spin splitting, as we now briefly discuss. The guiding center energy differs for spin-up and spin-down electrons by the Zeeman energy $g\mu_B B$. We note that the odd-integer plateaus are barely resolved in Fig. 10. In normal Hall-bar geometries such plateaus are fully resolved for fields above 4 T. This suggests that the g factor in perpendicular magnetic fields is much less enhanced in the point contact than in the wide 2D EG. It is quite conceivable that this would affect the point-contact magnetoresistance. We note in this connection that in *parallel* fields the spin splitting requires even larger fields exceeding 10 T to be resolved.[9] This effect is similar to that noted by Smith *et al.*[30] in capacitance measurements on narrow 2D EG channels in parallel fields.

A further discussion of these effects is beyond the scope of this paper, and we refer to Ref. 12 for a systematic study of the quantum Hall effect in a geometry with separately adjustable injecting and collecting point contacts. We note that the effects described above are related to the observations of backscattering of edge states in four-terminal experiments on a constriction or wire containing a potential barrier.[31,32]

V. THEORY OF ELECTRON FOCUSING

A. Classical electron focusing

Before turning to interference effects, we first consider the focusing spectrum in a 2D EG as it would follow

from classical mechanics. We start with the simplest case of a point injector and collector, and put the finite contact size in afterwards. For simplicity, we assume in the following calculations that the electron density in the point contacts is the same as that in the broad 2D EG regions (the effect of a reduced density in the point-contact region is discussed below). The 2D EG boundary is modeled as an infinite potential wall, causing purely specular boundary scattering. Consider a current flux tube which leaves the injector at an angle α with the x axis (see Fig. 1 for our choice of axes), with an infinitesimal angular opening $d\alpha$. The current through the flux tube is $dI = \frac{1}{2} \cos\alpha\, I_l\, d\alpha$, with I_l the total injected current. We assume that the electrons are injected under all angles $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$, weighed by $\cos\alpha$. This is what one expects classically in zero magnetic field for an injector modeled by a "hole in a screen." A nonzero field will only affect this angular distribution appreciably if the cyclotron diameter is comparable to the injector width W_l, which regime is considered below. The flux tube reaches the 2D EG boundary at separation $s = 2pl_{\text{cycl}} \cos\alpha$ from the injector, after $p - 1$ specular reflections ($p = 1, 2, \ldots$). We denote the cyclotron radius by $l_{\text{cycl}} \equiv \hbar k_F/eB$. The collector is at $s = L$, and has an infinitesimal width W_c. For each $p \geq L/2l_{\text{cycl}}$, two flux tubes with $\cos\alpha = L/2pl_{\text{cycl}}$ (one for positive and one for negative α) are incident on the collector. The current through the collector due to one such flux tube is $(W_c\, dI/d\alpha)|\partial s/\partial\alpha|^{-1}$. The total incident current I_c is

$$I_c = 2 \sum_{p \geq L/2l_{\text{cycl}}} \frac{1}{2} \cos\alpha\, I_l\, W_c |2pl_{\text{cycl}} \sin\alpha|^{-1}$$

$$= I_l \frac{W_c}{L} \sum_{p \geq L/2l_{\text{cycl}}} (L/2pl_{\text{cycl}})^2$$

$$\times [1 - (L/2pl_{\text{cycl}})^2]^{-1/2} . \qquad (9)$$

A similar expression for the incident current is given for the three-dimensional electron-focusing geometry by Benistant.[33] The collector voltage V_c will adjust itself so that the same amount of current flows back into the 2D EG and no net current is drawn. This implies $V_c = I_c/G_c$, where G_c is the conductance of the collector and V_c is measured relative to a grounded Ohmic contact which is also the drain for I_l [a three-terminal measurement, or equivalently a four-terminal measurement in the generalized Hall-resistance configuration (see Sec. IV and Appendix D)]. Classically, G_c is given by Eq. (2), provided W_c is much smaller than the cyclotron diameter. Combination with Eq. (9) gives

$$\frac{V_c}{I_l} = \frac{h}{2e^2} \frac{\pi}{k_F L} \sum_{p \geq L/2l_{\text{cycl}}} (L/2pl_{\text{cycl}})^2$$

$$\times [1 - (L/2pl_{\text{cycl}})^2]^{-1/2} . \qquad (10)$$

The divergencies at $L = 2pl_{\text{cycl}}$ are a consequence of assuming an infinitesimal width of both point contacts. These divergencies disappear if the finite contact width is accounted for by replacing L in Eq. (10) by $L + y_c - y_l$, and then averaging y_l, y_c over the respective widths

(a) Hall resistance
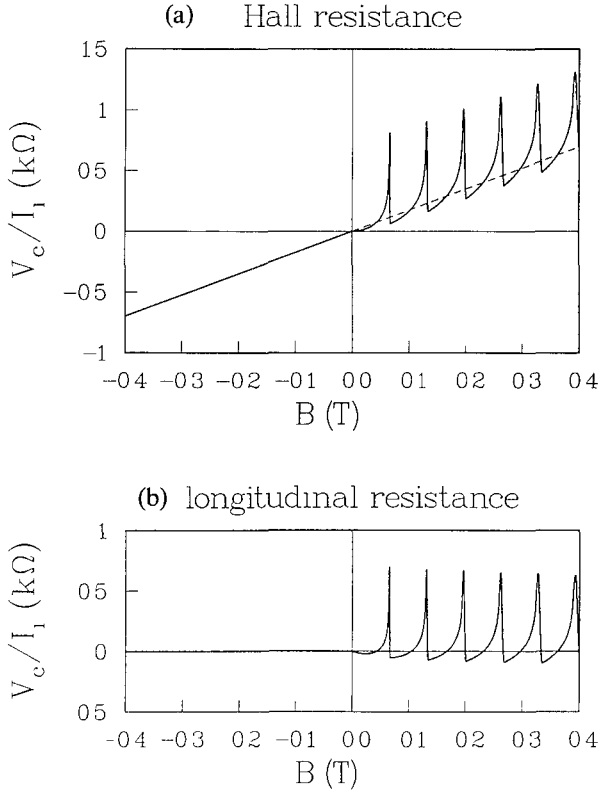

(b) longitudinal resistance

FIG. 11. Classical focusing spectrum, calculated from Eq. (10) with $W_i = W_c = 50$ nm, for two measurement configurations corresponding to Figs. 8(b) and 8(c). The dashed line in (a) is the extrapolation of the classical Hall resistance seen in reverse fields.

W_i, W_c of the injector and collector. (The resulting expression is rather lengthy, and not recorded here.)

A plot of the classical focusing spectrum in the generalized Hall-resistance configuration obtained from Eq. (10) is shown in Fig. 11(a) for the experimental parameters $L = 3.0$ μm and $k_F = 1.5 \times 10^8$ m^{-1}. The spectrum consists of a series of equidistant peaks at magnetic fields which are multiples of $B_{focus} = 0.066$ T [Eq. (1)]. With respect to the monotonously increasing baseline, these focusing peaks are of approximately equal height, which increases upon reducing the point-contact widths (Fig. 11

is for $W_i = W_c = 50$ nm). Such a classical focusing spectrum is commonly observed in metals albeit with a decreasing height of subsequent peaks because of partially diffuse scattering at the metal surface.

The resistance V_c/I_i is alternatingly larger and smaller than the classical Hall resistance $R_H = B/en_s$ [dashed line in Fig. 11(a)], as a consequence of the focusing effect. This was also found experimentally [cf. Sec. IV and Fig. 8(c)]. At very small fields the resistance is suppressed below R_H,

$$\frac{V_c}{I_i} = \frac{h}{2e^2} \frac{\pi}{k_F L} \sum_{p=1}^{\infty} (L/2pl_{cycl})^2$$

$$= \frac{\pi^3}{48} \frac{B}{en_s} \frac{L}{l_{cycl}} \quad \text{if } W_c, W_i \ll L \ll l_{cycl} , \tag{11}$$

vanishing as B^2 rather than B. Note in Fig. 11(a) that for reverse fields the normal Hall resistance occurs, leading to a discontinuity in $V_c/I_i B$ at $B = 0$. This behavior is evident in the experimental data, cf. Fig. 8(c). The electron-focusing spectrum in the generalized longitudinal-resistance configuration is obtained from Eq. (10) by simply subtracting the classical Hall resistance R_H. A plot is shown in Fig. 11(b). As a result of the focusing effect, we see *negative* longitudinal resistances—in agreement with the experiment [cf. Fig. 8(b)].

At very large fields, such that *either* W_c or W_i is larger than the cyclotron diameter, the resistance V_c/I_i becomes identical to R_H, because the resolution required for the observation of the focusing effect is lost. (If both point contacts have a reduced density, the Hall resistance at large fields deviates from its normal value of B/en_s, as discussed in Ref. 12 and Sec. IV; this case is not considered here.) The transition from focusing peaks to normal Hall resistance can be studied most easily for the case in which one of the point contacts has an infinitesimal width, much smaller than both l_{cycl} and L. Which of the two point contacts has the smallest width is irrelevant in view of the injector-collector reciprocity (Sec. IV and Appendix D), but to be definite let us assume that $W_i \ll l_{cycl}, L$. The angular distribution of injected electrons is then unaffected by the magnetic field. The total incident current on the collector is given by

$$I_c = I_i \int_0^{W_c} dy_c \frac{1}{L+y_c} \sum_{p \geq (L+y_c)/2l_{cycl}}^{p \leq (L+y_c)/y_c} [(L+y_c)/2pl_{cycl}]^2 \{1 - [(L+y_c)/2pl_{cycl}]^2\}^{-1/2}$$

$$= I_i \sum_{p \geq L/2l_{cycl}} \{[1 - (L/2pl_{cycl})^2]^{1/2} - [1 - (L_{min}/2pl_{cycl})^2]^{1/2}\} , \tag{12}$$

with the definition $L_{min} \equiv \min\{L + W_c, pL/(p-1), 2pl_{cycl}\}$. Equation (12) is obtained from Eq. (9) upon carrying out the average over the collector width, and adding the restriction $py_c \leq L + y_c$ to the summation over p. This restriction, which was unnecessary in the case $W_i, W_c \ll l_{cycl}$ considered above, avoids the multiple counting of electron trajectories with chord lengths

smaller than the collector width. To obtain the collector voltage V_c, we divide I_c by the collector conductance G_c, modified by the magnetic field according to Eq. (4). The result is plotted in Fig. 12, for $W_c = 200$ nm. The transition to the normal Hall resistance (dashed line in Fig. 12) occurs at $2l_{cycl} = W_c$, which corresponds to a field of 1 T. Beyond this field all injected electrons enter the collector,
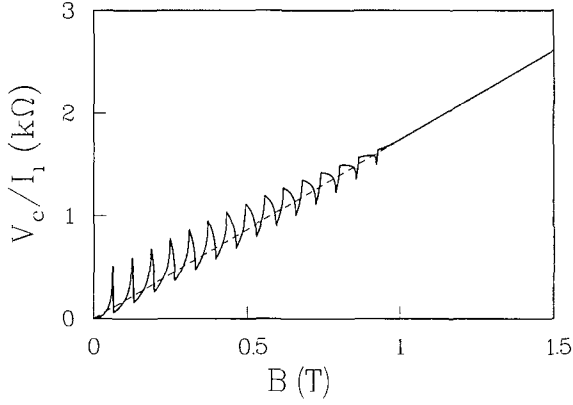
FIG. 12. Classical focusing spectrum in the generalized Hall-resistance configuration, calculated from Eq. (12) with $W_c = 200$ nm, showing the transition from electron focusing to the normal Hall effect (dashed line).

so that the focusing can no longer be detected.

We mention one more interesting regime. If the point-contact separation is very much larger than their widths, there is an intermediate-field regime $W_c, W_i \ll l_{\text{cycl}} \ll L$ in which we may approximate

$$\frac{V_c}{I_i} = \frac{h}{2e^2} \frac{\pi}{k_F L} \int_{L/2l_{\text{cycl}}}^{\infty} dp \, (L/2pl_{\text{cycl}})^2$$

$$\times [1 - (L/2pl_{\text{cycl}})^2]^{-1/2}$$

$$= \frac{\pi^2}{8} \frac{B}{en_s} \quad \text{if } W_c, W_i \ll l_{\text{cycl}} \ll L . \tag{13}$$

In this regime the resistance is approximately linear in B, just as the normal Hall resistance, but with an anomalously large slope.

In the above calculation we have disregarded the reduced electron density in the point-contact region. The electric fields induced by the carrier depletion *collimate* the electron beam, thereby enhancing the electron-focusing peaks, as we now briefly discuss. In the point contact the bottom of the conduction band is raised relative to the wide 2D EG. We model this by potential barriers of height E_i and E_c in the injector and collector. The corresponding reduced densities are $n_i = m(E_F - E_i)/\pi\hbar^2$ and $n_c = m(E_F - E_c)/\pi\hbar^2$, whereas the density in the wide 2D EG is given by $n_s = mE_F/\pi\hbar^2$. To be specific we first assume that the injector has the highest barrier. To overcome this barrier the energy of motion in the x direction should be larger than E_i, so that the injected electrons have velocity directions restricted to a cone of allowed angles α defined by $E_F \cos^2\alpha \geq E_i$. This restriction changes the normalization factor in the expression for the current through a flux tube, which is now given by $dI = \frac{1}{2}\cos\alpha(1 - E_i/E_F)^{-1/2}I_i d\alpha$, for α in the allowed cone. Since we have assumed that $E_i \geq E_c$, the barrier in the collector does not affect the incident current, but enters only in the expression for the collector conductance, Eq. (2), by reducing the Fermi wave vector,

$$G_c = (2e^2/h)(1 - E_c/E_F)^{1/2}k_F W_c/\pi .$$

Here, $k_F = (2mE_F)^{1/2}/\hbar$ denotes the Fermi wave vector in the wide regions. These two modifications combine to increase the height of the focusing peaks by a factor

$$(1 - E_i/E_F)^{-1/2}(1 - E_c/E_F)^{-1/2} = n_s(n_i n_c)^{-1/2} .$$

The injector-collector reciprocity discussed in Sec. IV and Appendix D implies that the above result is valid regardless of the relative height of E_i and E_c. We note that a horn-shaped constriction also tends to collimate the electron beam, and has a similar sharpening effect on the electron-focusing peaks as a barrier.

The results in this subsection explicitly demonstrate that large deviations from the normal Hall effect can occur in the ballistic transport regime due to classical electron focusing. Deviations similar to those of Eq. (13) can result from a reduced electron-gas density in the point contacts (or current and voltage probes in general), cf. Ref. 12 and Sec. IV. Their common origin is the absence of equilibrium among current-carrying electrons along the 2D EG boundary, due to a lack of inelastic scattering.[29] These anomalies in the Hall effect in a broad 2D EG are to be distinguished from the anomalies in narrow 2D EG channels, which have recently been the subject of extensive experimental[34] and theoretical[35,36] investigation.

B. Coherent electron focusing

To explain all of the experimental observations, it is necessary to go beyond the classical description of electron focusing given above. We first present a simple qualitative argument. Quantum ballistic transport along the 2D EG boundary takes place via magnetic edge channels,[37–40] which are the propagating modes of this problem. Some well-known results on magnetic edge states are reviewed in Appendix A. The modes at the Fermi level are labeled by a quantum number $n = 1, 2, \ldots, n_{\text{max}}$, with the number of edge channels n_{max} being equal to the number of occupied bulk Landau levels. If the injector has a width below λ_F, it excites these modes coherently. Since injector and collector are separated by less than a phase-coherence length (at least at low temperatures), interference between these modes can be of importance. In this subsection we demonstrate that such interference is responsible for large structure in the focusing spectra, if injector and collector are sufficiently narrow. For $k_F L \gg 1$ the interference of modes at the collector is determined by the phase factors $\exp(ik_n L)$, which vary rapidly as a function of n. The wave number k_n in the y direction (along the 2DEG boundary, see Fig. 1) corresponds classically to the x coordinate of the center of the cyclotron orbit, which is a conserved quantity upon specular reflection at the boundary.[39] In the gauge $\mathbf{A} = (0, Bx, 0)$ this correspondence may be written as [cf. Eq. (A2)] $k_n = k_F \sin\alpha_n$, where α is the angle with the x axis under which the cyclotron orbit is reflected from the boundary $(-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi)$. The quantized values α_n follow in this semiclassical description from the Bohr-Sommerfeld quantization rule[38–40] [cf. Eq. (A4)] that the flux enclosed by the cyclotron orbit and the boundary equals $(n - \frac{1}{4})h/e$ (for an infinite barrier potential). Sim-

ple geometry shows that this requires that

$$\tfrac{1}{2}\pi - \alpha_n - \tfrac{1}{2}\sin 2\alpha_n = \frac{2\pi}{k_F l_{\text{cycl}}}(n - \tfrac{1}{4}) ,$$

$$n = 1, 2, \ldots, n_{\text{max}} \qquad (14)$$

with n_{max} the largest integer smaller than $\tfrac{1}{2}k_F l_{\text{cycl}} + \tfrac{1}{4}$ As illustrated in Fig 13, the dependence on n of the phase $k_n L$ is close to linear in a broad interval Expansion of Eq (14) around $\alpha_n = 0$ gives

$$k_n L = \text{const} - 2\pi n \frac{B}{B_{\text{focus}}}$$

$$+ k_F L O([(n_{\text{max}} - 2n)/n_{\text{max}}]^3) \qquad (15)$$

It follows from this expansion that if B/B_{focus} is an integer, a fraction of order $(1/k_F L)^{1/3}$ of the n_{max} edge channels interfere constructively at the collector Because of the $\tfrac{1}{3}$ power, this is a substantial fraction even for the large $k_F L \sim 450$ of the experiment. The relevant states have quantum number n in an interval centered around $n_{\text{max}}/2$, corresponding to $\alpha_n = 0$. (The edge states outside the domain of linear n dependence of the phase give rise to additional interference structure which, however, does not have a simple periodicity.) The resulting mode-interference oscillations with B_{focus} periodicity can become much larger than the classical focusing peaks. To demonstrate this, we now calculate the wave function ψ in the Wentzel-Kramers-Brillouin (WKB) approximation.

We consider a point-dipole injector and determine $|\partial\Psi/\partial x|^2$ at the coordinates $(x,y)=(0,L)$ of the collector. We assume Ψ is unperturbed by the presence of the collector. The dipolar distribution

$$\Psi = (2I_i/\pi v_F r)^{1/2}\exp(ik_F r)\cos\alpha$$

at a small separation r from the injector (with I_i the injected current and v_F the Fermi velocity) is chosen instead of isotropic injection, because of the boundary condition $\Psi = 0$ at $x = 0$. The current I_c through a collector

with a width of the order of λ_F is determined in a first approximation by the unperturbed probability density at an infinitesimal distance from the collector. Since for an infinite barrier potential both Ψ and $\partial\Psi/\partial y$ vanish at $x = 0$, this density is proportional to our calculated $|\partial\Psi/\partial x|^2$, so that we can write

$$I_c = \varepsilon \left| \frac{\partial\Psi}{\partial x}(0,L) \right|^2 \qquad (16)$$

(with ε an undetermined parameter). The conductance of the collector is given in the same approximation by (see Appendix B)

$$G_c = \frac{2e^2}{h}\varepsilon k_F^3/v_F \qquad (17)$$

We thus find for the collector voltage $V_c = I_c/G_c$ the expression

$$V_c = \frac{h}{2e^2}(v_F/k_F^3)\left| \frac{\partial\Psi}{\partial x}(0,L) \right|^2 , \qquad (18)$$

which is independent of the parameter ε.

In the WKB approximation[41] the wave function $\Psi(x,y)$ is the sum over all classical trajectories from injector to the point (x,y) of an amplitude factor times a phase factor $\exp(i\phi)$. The amplitude factor is inversely proportional to the square root of the cross section of a particle flux tube containing the trajectory, as required by current conservation. The phase increment ϕ acquired along the trajectory is the sum of four terms (1) A path-length term $k_F l$, with l the length of the trajectory. (2) The Aharonov-Bohm phase $(-e/\hbar)\int dl \cdot \mathbf{A}$, given by the integral of the vector potential along the trajectory. In the gauge $\mathbf{A} = (0, Bx0)$ this term equals $-eB\mathcal{O}/\hbar$, with $\mathcal{O}$ the area between the trajectory and the boundary at $x = 0$. (3) A phase shift of π for each specular reflection at the boundary. (4) A phase shift of $-\tfrac{1}{2}\pi$ for each passage through a caustic, which is a point at which the cross section of the flux tube is reduced to zero (see Appendix C). These various terms are calculated in Appendix B. The final answer is

$$\frac{\partial\Psi}{\partial x}(0,L) = C \sum_{p \geq B/B_{\text{focus}}} \beta_p^2(1-\beta_p^2)^{-1/4}(e^{i\phi_p^+} + e^{i\phi_p^-}) ,$$

$$(19)$$

with $C \equiv -2ik_F(2I_i/\pi v_F L)^{1/2}$ a B-independent prefactor, $\beta_p \equiv B/pB_{\text{focus}} = L/2pl_{\text{cycl}}$ the reduced magnetic field, and the phases given by

$$\phi_p^+ = \frac{k_F L}{4\beta_p}[\pi - 2\arccos\beta_p + 2\beta_p(1-\beta_p^2)^{1/2}]$$

$$+ (p-1)\pi - (p-1)\tfrac{1}{2}\pi , \qquad (20a)$$

$$\phi_p^- = \frac{k_F l}{4\beta_p}[\pi + 2\arccos\beta_p - 2\beta_p(1-\beta_p^2)^{1/2}]$$

$$+ (p-1)\pi - p\tfrac{1}{2}\pi . \qquad (20b)$$

The index p labels the trajectories incident on the collector after $p - 1$ specular reflections. For each value of p
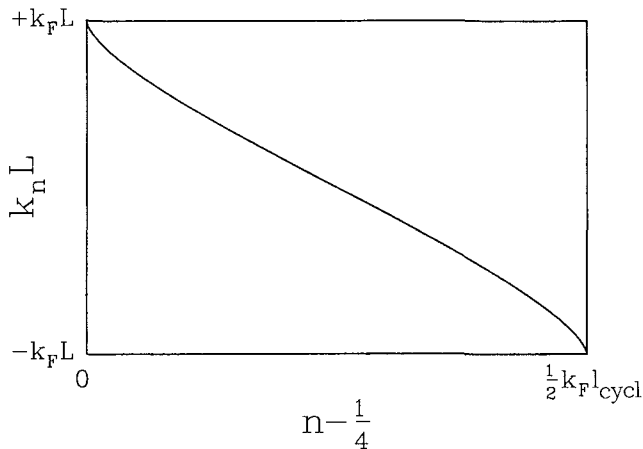

FIG 13 Phase $k_n L = k_F L \sin\alpha_n$ of the edge states, calculated from Eq (14) Note the domain of approximately linear n dependence of the phase, discussed in the text

there are two such trajectories, leaving the injector at an angle with the x axis given by $\alpha = \pm \arccos \beta_p$. Both trajectories have the same amplitude factor, but different phase increments ϕ^+ and ϕ^-.

In view of the long transport mean free path $l_e \sim 9$ μm in the experiment, we do not include the effects of impurity scattering in our calculation. We have found that taking into account impurity scattering in an averaged way, by weighing the contribution of trajectories of length l to Ψ with a factor $\exp(-l/2l_e)$, does not significantly affect our results. It is quite possible that the actual mean free path between collisions is considerably smaller than the transport mean free path obtained from the mobility, since the latter is insensitive to forward scattering. Timp *et al.*[27] have argued for a reduction by a factor of 10, on the basis of their nonlocal resistance measurements in a narrow 2D EG wire. However, if forward scattering would play an important role in the electron-focusing experiment, we would not expect to measure oscillations with a well-defined periodicity, since the scattering would scramble the phases of the edge states. The present experiment and theoretical analysis therefore seem to indicate that the actual mean free path is not much less than the point-contact separation, which is 3 μm.

In the above treatment we have neglected spin splitting, since the Zeeman energy is much smaller than the Fermi energy E_F in the field range considered. We also neglect a possible B dependence of k_F. In the absence of Landau-level broadening in the bulk of the 2D EG, pinning of E_F at Landau levels would give rise to a modulation of k_F periodic in $1/B$ by up to 10% at 1 T. In practice, the amplitude of the modulation is much reduced by Landau-level broadening. Moreover, this effect does not lead to a definite B periodicity of the collector voltage.

The magnetic field dependence of V_c/I_1 resulting from Eqs. (18)–(20) is shown in Fig. 14 (bottom), for the experimental values $L = 3.0$ μm and $k_F = 1.5 \times 10^8$ m^{-1}. The most rapid oscillations were eliminated by averaging L over an interval of 100 nm. This corresponds roughly to
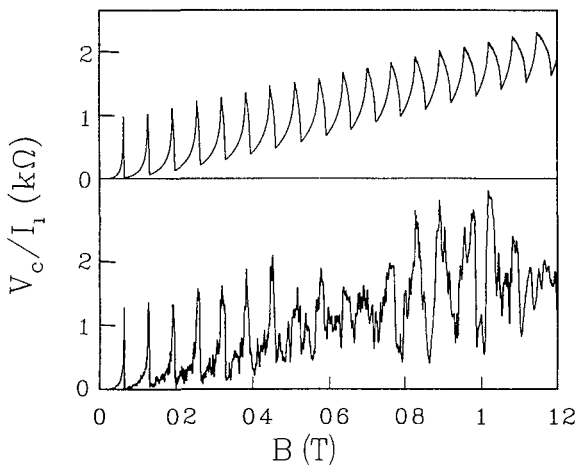

FIG. 14. The bottom theoretical curve shows the magnetic field dependence of the collector voltage calculated from Eqs. (18)–(20). The top curve results if only the incoherent contributions are retained (no interference effects).

the combined width of the point contacts—but is otherwise not intended to be a realistic description of the effect of a finite contact size, which remains a subject for further investigation. Also plotted in Fig. 14 (top) is the incoherent contribution to the collector voltage, without the interference of different trajectories, which shows simply the peaks from classical electron focusing at multiples of B_{focus}. Interference effects give rise to fine structure on the focusing peaks at low magnetic fields, which grows in amplitude with increasing field. It is apparent from Fig. 14 (and confirmed by Fourier transform) that *the large-amplitude high-field oscillations have the same periodicity as the smaller low-field focusing peaks* —as observed experimentally, and consistent with the mode-interference argument given above. This is the main result of our calculation, which we have found to be insensitive to details of the point-contact modeling. (Insensitive, for example, to assuming isotropic instead of dipolar injection.) The above calculation is for $T = 0$. We have investigated the effect of energy averaging (over an interval of $k_B T$ around E_F) at finite T, and found that temperatures of the order of 10 K are necessary to smear out most of the interference structure. This is a weaker temperature dependence than observed experimentally (Fig. 4), possibly due to inelastic scattering limiting the phase coherence at finite temperatures.

The relation between Eq. (19) and the edge states can be made explicit, if one transforms the sum over trajectories into a sum over modes by means of the method of stationary phase. This is done in Appendix B, with the result

$$\frac{\partial \Psi}{\partial x}(0,L) = C' \sum_{n=1}^{n_{\max}} e^{ik_n L} , \tag{21}$$

plus corrections from evanescent waves [which a numerical comparison with Eq. (19) has shown to be small]. The prefactor is

$$C' \equiv -2ik_F(I_t/v_F k_F l_{\text{cycl}}^2)^{1/2} \exp(-\tfrac{3}{4}\pi i) .$$

The phases $k_n L = k_F L \sin \alpha_n$ of the modes are the same as determined earlier in Eq. (14). [This is as expected, since the Bohr-Sommerfeld quantization rule used to derive Eq. (14) and the WKB approximation which leads to Eq. (21) are equivalent levels of approximation.] The alternative representations (19) and (21) of this quantum-mechanical transport problem are the analogues of the classical ray and mode descriptions of propagation in a waveguide.[42] In this context the edge states correspond to Lord Rayleigh's "whispering gallery" waves.[43] The present theory also has many similarities with the theory of radio-wave propagation through the atmosphere, where focusing and mode interference are well-known phenomena.[44]

We note that Tsoi[45] (to explain a fine structure in the first focusing peak in bismuth) has proposed that an *individual* edge state n would cause a peak in the collector voltage whenever L is an integer multiple of the chord length $2l_{\text{cycl}} \cos \alpha_n$ of the corresponding skipping orbit. We do not see how this can be reconciled with the fact that the probability density $|\Psi_n|^2$ of an individual edge

state is y independent [since $\Psi_n(x,y)=f_n(x)\exp(ik_ny)$].

The accuracy of our theoretical treatment of coherent electron focusing is limited by our use of the WKB approximation, which, in principle, restricts the theory to treating the effects of edge states with large quantum numbers. The theory should be accurate at low fields when a large number of edge channels are populated, but we expect our main result of the fundamental periodicity to hold at higher fields as well. We surmise that an *exact* calculation of the focusing spectrum is feasible in the point-contact limit $W \ll \lambda_F$, since one can then use the unperturbed wave functions in the 2D EG which are known exactly (Weber functions).

VI. DISCUSSION

Electron focusing in metals is a technique which is widely used to obtain information on the Fermi surface, on surface scattering, and on other scattering processes. Such information can be obtained from electron-focusing experiments in a 2D EG as well. The experiments reported in this paper demonstrate conclusively that scattering of electrons by the 2D EG boundary is predominantly specular. This conclusion is of importance for the interpretation of galvanomagnetic size effects in narrow 2D EG channels.[19,46] The Fermi surface in the 2D EG in a GaAs-Al$_x$Ga$_{1-x}$As heterostructure [on a (100)-oriented GaAs surface] is simply a circle. It would be of interest to perform similar experiments on heterostructures with more complicated Fermi surfaces. The prerequisite for such studies is a sufficiently high mobility.

Our electron-focusing experiments in a 2D EG have also yielded results of a different nature, not previously found in metals. The shape and amplitude of the focusing peaks, and especially the fine structure observed at low temperatures, are signatures of a new phenomenon: *coherent electron focusing*. A quantitative comparison between theory and experiment requires a more detailed analysis of the point contacts and gate potential than attempted in Sec. V. Such a calculation would have to take into account the reduced electron density in the point-contact region and along the 2D EG boundary formed by the gate potential. The appearance of high-field oscillations with the focusing periodicity, but with much larger amplitude, is, however, characteristic for the mode-interference mechanism proposed in this paper. Indeed, this is the feature of the experimental focusing spectra which is *insensitive to small changes in the gate voltage* and which is present in both the devices studied. This novel quantum-interference effect in ballistic transport may also play a role in the multiprobe "electron waveguides" of current interest.[10] Voltage fluctuations with a well-defined periodicity were found in such a device by Chang *et al.*,[47] albeit in the regime where the transport was not fully ballistic.

Electron focusing is in essence a nonlocal[48] transport measurement in the quantum ballistic regime, in the most simple geometry conceivable. In contrast to the usual channel geometry, here electrons interact with a single boundary only, while current injection and detection is done by means of point contacts comparable in size to the

electron wavelength. This allows for a simple solution of the transport problem, as shown in Sec. V. The interpretation of four-terminal measurements of electron focusing as generalized Hall- and longitudinal-resistance measurements shows that even in the weak-magnetic-field regime large and interesting deviations from the Hall and longitudinal resistance in diffusive transport result, as a consequence of the finite size of current and voltage probes. It would be of interest to extend both theory and experiment to narrow channel geometries, with point contacts as current and voltage probes. The present experiments and theoretical analysis point the way to the correct modeling of such probes.

At high magnetic fields a transition from electron focusing to the quantum Hall effect is observed. In this regime the measured four-terminal resistance becomes essentially identical to the two-terminal resistance of injector or collector—whichever is largest.[12] Thus quantized plateaus as well as quantum-interference effects originating in a single point contact show up in the Hall resistance. These observations can be understood on the basis of the expression (7) for the collector voltage, derived from the Landauer-Büttiker[11] formalism in Appendix D.

We mention a few possibilities for future extensions of the electron-focusing technique in a 2D EG. Ballistic injection of hot electrons can be realized by voltage-biasing a point contact. This can be observed as a shift in the focusing peaks. Such experiments are in progress. Diffraction of electrons on a periodically corrugated 2D EG boundary might be investigated with electron focusing if a structured gate is used to define the boundary.[49] The classical or quantum-mechanical localization of electrons by a strong magnetic field could be studied, in principle, by defining an obstacle (by means of a gate) in the space between collector and injector.

The experimental results presented in this paper demonstrate the feasibility of coherent electron optics in the solid state. Quantum point contacts as monochromatic point sources of ballistic electrons, and the 2D EG boundary as a mirror, constitute the first proven building blocks for this new field.

APPENDIX A: MAGNETIC EDGE STATES

In this appendix we review some known results on magnetic edge states, which are used in Sec. V. We first discuss the classical motion. A skipping orbit (Fig. 1 or

inset of Fig. 15) consists of a series of translated circular arcs. The position (x,y) of the electron on the circle with center coordinates (X, Y) can be expressed in terms of its velocity **v** by

$$x = X + v_y/\omega_c, \quad y = Y - v_x/\omega_c \ , \tag{A1}$$

with $\omega_c = eB/m$ the cyclotron frequency. Note that the separation X of the center from the boundary is constant on a skipping orbit; only the center coordinate Y parallel to the boundary changes at each specular reflection. The canonical momentum of the electron is $\mathbf{p} = m\mathbf{v} - e\,\mathbf{A}$. In the Landau gauge $\mathbf{A} = (0,Bx,0)$ we have

$$p_x = mv_x, \quad p_y = -eBX \ , \tag{A2}$$

which tells us that p_y is a constant of the motion.

The motion projected on the x axis is periodic, so that we can apply the quasiclassical Bohr-Sommerfeld quantization rule

$$\frac{1}{\hbar} \oint p_x dx + \gamma = 2\pi n \ . \tag{A3}$$

The integral is over one period of the motion, n is an integer, and γ is the sum of the phase shifts acquired at the two turning points of the motion. The phase shift upon reflection at the boundary is π (for an infinite barrier potential); the other turning point is a caustic (a point at which classically the particle density becomes infinite), which gives a phase shift of $-\frac{1}{2}\pi$ (see Appendix B). This
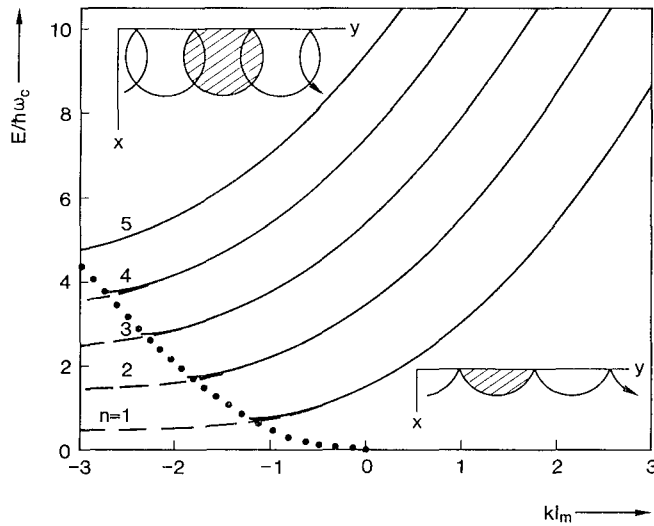

FIG. 15. Energy spectrum $E_n(k)$ of magnetic edge states at a single boundary represented by an infinite barrier potential. The energy E is scaled by $\hbar\omega_c$, and the wave number k by the reciprocal of the magnetic length $l_m \equiv (\hbar/eB)^{1/2}$. The insets show classical skipping orbits for positive and negative k. Note that $kl_m = -X/l_m$, with X the separation of the orbital center from the boundary. In the quasiclassical approximation the magnetic flux through the shaded areas is quantized. The resulting edge-state–energy spectrum (A6) (solid curves) is indistinguishable from the exact solution (dashed curves, from Ref. 51), unless k is within $1/l_m$ of the transition from skipping to cyclotron orbits (dotted curve).

totals to $\gamma = \frac{1}{2}\pi$. Using Eqs. (A1) and (A2) we may thus write Eq. (A3) in the form

$$\frac{eB}{\hbar} \oint (Y - y)dx = 2\pi(n - \tfrac{1}{4}), \quad n = 1,2,\dots \ . \tag{A4}$$

This quantization rule has the simple geometrical interpretation that the flux enclosed by one arc of the skipping orbit and the boundary equals $n - \frac{1}{4}$ times the flux quantum h/e. The above derivation of this old result[38–40] is more direct than the usual one, which proceeds from the quantization of the orbits in momentum space to coordinate space.

We note that the quantization rule (A4) holds also for bulk Landau levels (corresponding to circular cyclotron orbits which do not collide with the 2D EG boundary)— provided the coefficient $n - \frac{1}{4}$ on the right-hand side (rhs) is replaced by $n - \frac{1}{2}$. [The replacement of $\frac{1}{4}$ by $\frac{1}{2}$ accounts for the fact that the reflection at the boundary is replaced by a caustic turning point, so that the phase shift is $\gamma = -\frac{1}{2}\pi - \frac{1}{2}\pi = \pi$ (mod 2π) instead of $\gamma = \frac{1}{2}\pi$.] Similarly, Eq. (A4) also holds in a narrow-channel 2D EG for states which interact with both boundaries, corresponding to traversing trajectories[35,50] which move from one channel wall to the opposite one. The coefficient $n - \frac{1}{4}$ on the rhs is then to be replaced by n [corresponding to a phase shift $\gamma = \pi + \pi = 0$ (modulo 2π) for two reflections at the boundary]. In this case the geometrical interpretation of Eq. (A4) is that n flux quanta are contained in the area bounded by the channel walls and a circle of cyclotron radius mv/eB centered at X. In the limit $B \to 0$ this area equals $2W(mv/eB)\cos\alpha$ (with α the angle of the trajectory with the x axis and W the channel width), so that the usual zero-field quantization condition $(mv/\hbar)\cos\alpha = n\pi/W$ is recovered.

Equation (A4) determines, for a given magnetic field, the energy $E = \frac{1}{2}mv^2$ as a function of the quantum number n and the wave number $k \equiv p_y/\hbar = -(eB/\hbar)X$. To carry out the integration in Eq. (A4) we express y in terms of x by means of Eq. (A1),

$$(Y - y)dx = [E(\tfrac{1}{2}m\omega_c^2)^{-1} - (x - X)^2]^{1/2}|dx| \ . \tag{A5}$$

The resulting energy spectrum $E_n(k)$ is given by

$$\frac{2E}{\hbar\omega_c}[\arccos\xi - \xi(1 - \xi^2)^{1/2}] = 2\pi(n - \tfrac{1}{4}) \ ,$$

$$\xi \equiv \hbar k(2mE)^{-1/2} \ , \tag{A6}$$

and is plotted in Fig. 15 (solid curves). Also plotted in Fig. 15 is the exact solution of the Schrödinger equation (dashed curves, taken from Ref. 51). The quasiclassical approximation (A6) is indistinguishable on this scale from the exact solution, except just before the transition from skipping orbits to bulk cyclotron orbits at $X = mv/eB$ (dotted curve in Fig. 15).

The quantization rule (A4) can also be used to find, for a given energy E, the number of occupied subbands N as a function of magnetic field B in a 2D EG channel of width W. This number N is the largest possible value of the quantum number n in Eq. (A4). In view of the geometrical interpretation of the quantization rule, N is

determined by the maximum number of flux quanta contained in an area bounded by the channel walls and a circle of cyclotron radius $l_{cycl} \equiv mv/eB = (2mE)^{1/2}/eB$. Note that a maximal enclosed flux is obtained by centering the circle on the channel axis. We therefore have

$$N = \text{Int} \left[\frac{\gamma}{2\pi} + \frac{Be}{h} \int_0^{\min\{W/2, l_{cycl}\}} dx \, 4(l_{cycl}^2 - x^2)^{1/2} \right] ,$$

(A7)

with $\gamma = 0$ if $l_{cycl} > W/2$, and $\gamma = \pi$ if $l_{cycl} < W/2$. "Int" denotes truncation to an integer. Evaluation of the integral gives the required result,

$$N = \begin{cases} \text{Int}\left[\dfrac{2}{\pi}(E/\hbar\omega_c)[\arcsin u + u(1-u^2)^{1/2}] \right] \\ \qquad\qquad\qquad\qquad \text{if } u < 1 , \quad \text{(A8a)} \\ \text{Int}[\frac{1}{2} + E/\hbar\omega_c] \text{ if } u > 1 , \qquad \text{(A8b)} \end{cases}$$

with $u \equiv W/2l_{cycl}$. The above expression for the number of occupied subbands has a discontinuity at $l_{cycl} = W/2$, originating from the discontinuity in the phase shift γ at that magnetic field value. This is an artifact of the semiclassical approximation, which ignores the extension of the wave function beyond the classical orbit[52]. Since the discontinuity in N is at most ± 1, it is unimportant in many applications.

APPENDIX B: DETAILS OF THE CALCULATION IN SEC. V

1. Ray calculation

The required derivative $\partial\Psi/\partial x$ of the unperturbed wave function at the collector coordinates $(0,L)$ is calculated as follows. The WKB expression[41] for Ψ at a point $(\delta x, L)$ close to the collector is a sum over all classical trajectories from injector to $(\delta x, L)$ of an amplitude factor A times a phase factor $e^{i\phi}$. Consider the trajectories which contain $p-1$ specular reflections and reach the endpoint just before hitting the boundary (this is the incident wave; the reflected wave also contributes to the unperturbed wave function and is dealt with below). Two trajectories meet this requirement (provided $2pl_{cycl} \geq L$), one leaving the injector at a positive angle α with the x axis, the other at a negative angle. Both have the same amplitude factor,

$$A_p = \left[\frac{2I_i}{\pi v_F} \right]^{1/2} \beta_p [L(1-\beta_p^2)^{1/2}]^{-1/2}[1 + O(\delta x/L)] ,$$

(B1)

where $\beta_p \equiv B/pB_{focus} = L/2pl_{cycl}$ is the reduced magnetic field. The amplitude factor is the product of a B-independent normalization constant, the angular weight factor $\beta_p = \cos\alpha + O(\delta x/L)$ for dipolar injection, and the inverse square root of the cross section at the collector of a particle flux tube containing the trajectory—as re-

quired by current conservation. The phase factors follow from

$$\phi^{\pm} = \frac{1}{\hbar} \int^{\pm} \mathbf{p} \cdot d\mathbf{l} + n_r^{\pm}\pi - n_c^{\pm}\tfrac{1}{2}\pi ,$$

(B2)

where $\mathbf{p} = m\mathbf{v} - e\mathbf{A}$ is the canonical momentum (to be integrated along the classical trajectory, with $|\mathbf{v}| = v_F$), n_r is the number of reflections at the boundary, and n_c is the number of passages through a caustic. The $+$ and $-$ refers to the trajectory with positive and negative α, respectively. In the gauge $\mathbf{A} = (0, Bx, 0)$ one finds, to linear order in δx,

$$\frac{1}{\hbar} \int^{+} \mathbf{p} \cdot d\mathbf{l} = \frac{k_F L}{4\beta_p}[\pi - 2\arccos\beta_p + 2\beta_p(1-\beta_p^2)^{1/2}]$$
$$- \beta_p k_F \, \delta x ,$$

(B3a)

$$\frac{1}{\hbar} \int^{-} \mathbf{p} \cdot d\mathbf{l} = \frac{k_F L}{4\beta_p}[\pi + 2\arccos\beta_p - 2\beta_p(1-\beta_p^2)^{1/2}]$$
$$- \beta_p k_F \, \delta x .$$

(B3b)

Each of the $n_r^+ = n_r^- = p - 1$ reflections gives an additional phase shift of π, which is what is needed to make Ψ vanish at the boundary (since, then, incident and reflected waves cancel). Finally, each passage through a caustic (a point at which the cross section of the flux tube vanishes) retards the phase by $\frac{1}{2}\pi$. Physically, this phase shift is the result of diffraction setting a lower limit to the flux-tube cross section[53]. At a small separation R from the caustic the cross section is proportional to R, so that the amplitude factor $A \propto R^{-1/2}$. The sign change of R upon passing through the caustic then formally leads to the phase factor $(-1)^{-1/2} = \exp(-i\frac{1}{2}\pi)$. (A more satisfactory derivation for the case of zero magnetic field is given in Ref. 41, Chap. 17.) The location of the caustics in the 2D EG is determined in Appendix C. Here we only need the result for the number of passages $n_c^+ = p - 1$, $n_c^- = p$.

So far we have considered only the incident wave. The reflected wave gives identical contributions, but with an extra phase shift of π and with the term $-\beta_p k_F \, \delta x$ in Eq. (B3) replaced by $+\beta_p k_F \delta x$. Collecting results, we find

$$\Psi(\delta x, L) = -2ik_F \, \delta x \sum_{p \geq B/B_{focus}} \beta_p A_p (e^{i\phi_p^+} + e^{i\phi_p^-}) ,$$

(B4)

with the phases ϕ_p^+ and ϕ_p^- defined in Eq. (20). Consistent with the WKB approximation, we have in Eq. (B4) neglected terms of order $\delta x/L$, which are a factor $k_F L$ smaller than the terms retained. Upon division of $\Psi(\delta x, L)$ by the small increment δx, we thus obtain the required result (19) for $\partial\Psi/\partial x$ at $(0,L)$.

2. Mode calculation

The sum over trajectories in Eq. (B4) can be transformed into a sum over edge channels (or modes) by means of the method of stationary phase[54]. For $k_F L \gg 1$ the phase ϕ_p varies rapidly as a function of p, so that the dominant contributions to the sum (B4) are those which for subsequent p values close to some value p_n (determined below) differ by approximately $2\pi n$ in phase

$(n = 1, 2, \ldots)$. To apply the usual method of stationary phase, we replace ϕ_p by the equivalent $\tilde{\phi}_p \equiv \phi_p - 2\pi np$, so that the phase shift of these subsequent terms is approximately zero rather than $2\pi n$. The point p_n of stationary phase follows from $\partial\tilde{\phi}_{p_n}/\partial p_n = 0$, and need not be an integer. Upon carrying out the differentiation, we find that p_n is determined by $\beta_{p_n} = \cos\alpha_n$, where the quantized angle α_n is the same as defined in Eq. (14). Note that, depending on whether α_n is positive or negative, the contributing terms of stationary phase in Eq. (B4) are either those with ϕ^+ or with ϕ^-.

We proceed by expanding $\tilde{\phi}_p$ around p_n,

$$\tilde{\phi}_p = \tilde{\phi}_{p_n} + \tfrac{1}{2}(p - p_n)^2 \frac{\partial^2\tilde{\phi}_{p_n}}{\partial p_n^2} + \cdots$$

$$= \varepsilon^{\pm} + k_F L \sin\alpha_n - \tfrac{1}{2}k_F L \frac{\cos^2\alpha_n}{\sin\alpha_n}(p/p_n - 1)^2 + \cdots .$$

(B5)

The phase increments ε^+ and ε^- are used for α_n positive or negative, respectively, and are defined by $\varepsilon^+ \equiv -\tfrac{1}{2}\pi$ and $\varepsilon^- \equiv -\pi$. We now substitute the expansion (B5) into Eq. (B4),

$$\Psi(\delta x, L) = -2ik_F \, \delta x \sum_{n=1}^{n_{\max}} \beta_{p_n} A_{p_n} \exp(i\varepsilon^{\pm} + ik_F L \sin\alpha_n) \int_{-\infty}^{+\infty} dp \, \exp\left[-i\tfrac{1}{2}k_F L \frac{\cos^2\alpha_n}{\sin\alpha_n}(p/p_n - 1)^2 \right] .$$

(B6)

Here we have made the usual approximations of replacing β_p and A_p by their values at $p = p_n$, and the series in p by an integration. Evaluating this Gaussian integral, we find

$$\Psi(\delta x, L) = -2ik_F \, \delta x \sum_{n=1}^{n_{\max}} \beta_{p_n} A_{p_n} \exp(i\varepsilon^{\pm} + ik_F L \sin\alpha_n) \left[\frac{2\pi \sin\alpha_n p_n^2}{ik_F L \cos^2\alpha_n} \right]^{1/2} ,$$

(B7)

which reduces to the result (21) discussed in Sec. V, if one uses the identity $\exp(i\varepsilon^{\pm})(-i\sin\alpha_n)^{1/2} = \exp(-\tfrac{3}{4}\pi i)(1 - \beta_{p_n}^2)^{1/4}$.

3. Point-contact conductance

We calculate the two-terminal conductance G of a point contact which is narrower than λ_F, under the assumption of Sec. V that the current through the point contact is given by $\varepsilon|\partial\Psi/\partial x|^2$. Here, ε is an undetermined parameter (with dimensions of $m^4 s^{-1}$), and Ψ is the wave function which is unperturbed by the presence of the point contact. The derivative normal to the 2D EG boundary is evaluated at the position of the point contact.

In the WKB approximation the wave function Ψ_n of the nth edge channel close to the 2D EG boundary can be expanded in an incident (in) and a reflected (ref) plane wave, $\Psi_n = \Psi^{\text{in}} + \Psi^{\text{ref}}$, with

$$\Psi^{\text{in}} = \text{const} \times \exp[ik_F(y \sin\alpha_n - x \cos\alpha_n)]$$

and

$$\Psi^{\text{ref}} = \text{const} \times \exp[ik_F(y \sin\alpha_n + x \cos\alpha_n) + i\pi] .$$

Here, α_n is the angle with the x axis under which the corresponding skipping orbit is reflected from the boundary. It follows from this expansion that (at $x = 0$)

$$\left| \frac{\partial\Psi_n}{\partial x} \right|^2 = 4 \left| \frac{\partial\Psi^{\text{in}}}{\partial x} \right|^2$$

$$= 4(k_F \cos\alpha_n)^2 |\Psi^{\text{in}}|^2$$

$$= 4(k_F^2/v_F)\cos\alpha_n \Phi^{\text{in}} ,$$

(B8)

where Φ^{in} is the flux of electrons incident on the boundary. Semiclassically, this flux is the ratio of the total current I_n carried by the nth edge state and the chord length $2l_{\text{cycl}}\cos\alpha_n$ of the corresponding skipping orbit, so that

$$\left| \frac{\partial\Psi_n}{\partial x} \right|^2 = (2k_F^2/v_F l_{\text{cycl}})I_n .$$

(B9)

We have thus found that in this approximation a fraction

$$T_n = \varepsilon(2k_F^2/v_F l_{\text{cycl}})$$

(B10)

of the current is transmitted through the point contact. Note that this transmission probability T_n turns out to be independent of the edge-state quantum number n.

The conductance of the point contact is given by the two-terminal Landauer formula[11,21,22]

$$G = \frac{2e^2}{h} \sum_{n=1}^{n_{\max}} T_n = \frac{2e^2}{h} n_{\max}\varepsilon(2k_F^2/v_F l_{\text{cycl}}) .$$

(B11)

The conductance given by Eq. (17) in Sec. V is obtained from Eq. (B11) by approximating the number of occupied edge states $n_{\max} \approx \tfrac{1}{2}k_F l_{\text{cycl}}$. The point-contact conductance is then independent of the magnetic field. The discreteness of $n_{\max}$, neglected in this approximation, may lead to a small modulation of the conductance periodic in $1/B$ (similar to the Shubnikov–de Haas oscillations). Since this modulation has no definite B periodicity, it would not significantly affect our calculations and is ignored.

APPENDIX C: CALCULATION OF THE CAUSTICS

The caustics in electron focusing are points in the 2D EG at which the cross section of a particle flux tube from the injector vanishes. Consider a flux tube consisting of

trajectories leaving the injector at angles with the x axis between α and $\alpha+\delta\alpha$ The flux tube is bounded by two trajectories $\mathbf{r}_\alpha$ and $\mathbf{r}_{\alpha+\delta\alpha}$, where $\mathbf{r}_\alpha(\phi,p)$ is defined by

$$x = l_{\text{cycl}}(\cos\phi - \sin\alpha) ,$$
$$y = l_{\text{cycl}}[\sin\phi + (2p-1)\cos\alpha] \tag{C1}$$

Here, $-\pi < \phi < \pi$, but only that part of the trajectory with $x > 0$ plays a role An intersection of these two trajectories is a caustic By equating $\mathbf{r}_\alpha(\phi,p) = \mathbf{r}_{\alpha+\delta\alpha}(\phi+\delta\phi, p)$, for infinitesimal $\delta\alpha$ and $\delta\phi$, we find that the caustics are points on $\mathbf{r}_\alpha(\phi,p)$ with $-\tan\phi = (2p-1)^{-1}\cot\alpha \equiv u$ By eliminating ϕ we obtain the required equation for the caustics,

$$x = l_{\text{cycl}}[-\sin\alpha + (1+u^2)^{-1/2}] ,$$
$$y = l_{\text{cycl}}[(2p-1)\cos\alpha - u(1+u^2)^{-1/2}] \tag{C2}$$

Here, $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$, and the integer $p \geq 2$ for positive α but $p \geq 1$ for negative α

The caustics are plotted in Fig 16 (bottom), up to $p = 3$ Note, as a check on the calculation, that the curves follow the regions of high density of trajectories apparent in Fig 16 (top) We see two families of caustics one family for trajectories which leave the injector at a negative angle α with the x axis [curves from the point $(2l_{\text{cycl}},0)$ to the 2D EG boundary], and one for trajectories with positive α (curves with a cusp, which begin and end on the 2D EG boundary) Each trajectory passes once through a caustic between two subsequent reflections at the boundary In addition, the trajectories with negative α (but not those with positive α) pass once through a caustic prior to the first reflection This is what we need to know for the WKB calculation in Appendix B

APPENDIX D: LANDAUER-BUTTIKER FORMALISM FOR ELECTRON FOCUSING

The Landauer-Buttiker (LB) formalism[11] relates resistances to transmission probabilities into current and voltage probes Since this method has not been used previously for the electron-focusing problem, we will describe it here in some detail The usual approach[55] is to solve the Boltzmann transport equation for the charge and current density to linear order in the applied electric field, while determining the nonequilibrium electric field distribution self-consistently from the requirement of charge neutrality The LB approach, in contrast, yields the potential of the voltage probes directly, without the need to first calculate the electric field distribution itself In this way the resistances are obtained entirely from equilibrium properties of the system, as it should be for a linear-response calculation The equivalence of the LB resistance formula in terms of transmission probabilities to the Kubo formula in terms of Green's functions has been demonstrated by Stone et al[21,56]

To derive the LB formula for the electron-focusing experiment, we consider the geometry in Fig 17 of a three-terminal conductor with point contacts in two of the probes In order to have well-defined initial and final states for the scattering problem, the probes are connected via perfect leads (or electron waveguides) to reservoirs which have a constant electrochemical potential We denote by μ_i and μ_c the chemical potentials of the two reservoirs connected, respectively, to the injector and collector point contact, and by μ_d the chemical potential of the third reservoir (the current drain) Following Buttiker,[11] we can relate the currents I_α ($\alpha = i,c,d$) in the three leads to these chemical potentials via the transmission probabilities $T_{\alpha\to\beta}$ (from reservoir α to reservoir β)
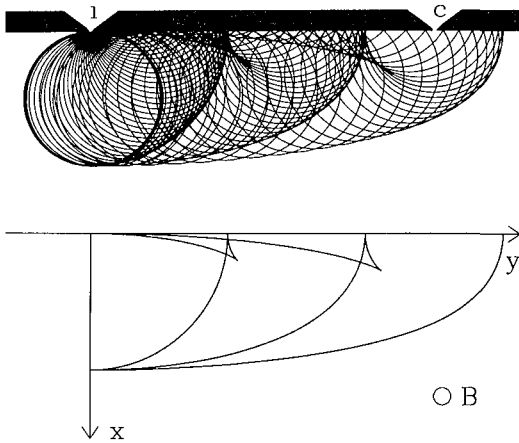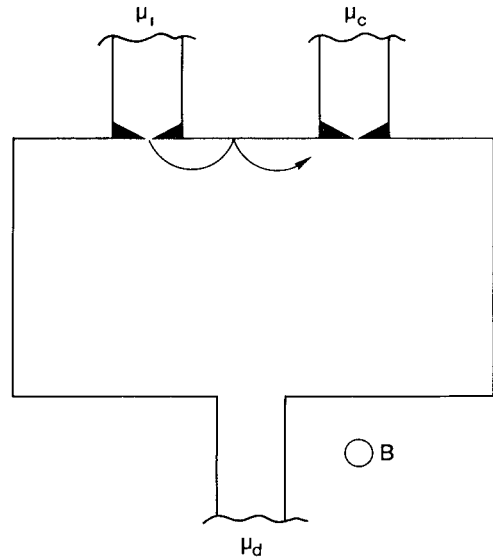
FIG 16 Top skipping orbits in a 2D EG The gate defining the injector (i) and collector (c) point contacts and the 2D EG boundary is shown schematically in black For clarity, the trajectories are drawn only up to the third specular reflection Bottom calculated location of the caustic curves

FIG 17 Three-terminal conductor in the electron-focusing geometry Three reservoirs at chemical potentials μ_i, μ_c, and μ_d are connected by leads to a wide 2D EG Two of the leads contain a narrow constriction (shown in black) The current flows from reservoir i to reservoir d, while reservoir c draws no net current

and reflection probabilities R_α (from reservoir α back to the same reservoir). These equations are of the form[11]

$$\frac{h}{2e}I_\alpha = (N_\alpha - R_\alpha)\mu_\alpha - \sum_{\beta\,(\neq\alpha)} T_{\beta\to\alpha}\mu_\beta \; . \tag{D1}$$

Here, N_α is the number of occupied (spin-degenerate) transverse wave guide modes in the lead α, also referred to as the number of channels. We now impose the condition that the collector draws no net current, which implies $I_c = 0$ and $I_d = -I_i$, and choose our zero of energy such that $\mu_d = 0$. We then find from Eq. (D1) the equations

$$\mu_c = \frac{T_{i\to c}}{N_c - R_c}\mu_i \; , \tag{D2a}$$

$$\frac{h}{2e}I_i = (N_i - R_i)\mu_i - T_{c\to i}\mu_c \; . \tag{D2b}$$

In the electron-focusing experiment we can neglect the transmission probability $T_{c\to i}$ from collector to injector, since electrons leaving the collector point contact are deflected away from the injector towards the current drain by the magnetic field. We then obtain for the ratio of collector voltage $V_c \equiv \mu_c/e$ (measured relative to the potential of the current drain) to injected current I_i the result

$$\frac{V_c}{I_i} = \frac{h}{2e^2}\frac{T_{i\to c}}{(N_c - R_c)(N_i - R_i)} \; . \tag{D3}$$

The quantities $(2e^2/h)(N_i - R_i)$ and $(2e^2/h)(N_c - R_c)$ can be interpreted as the two-terminal conductances G_i and G_c of the injector and collector point contact, respectively. [More precisely, $G_i \equiv eI_i/(\mu_i - \mu_d)$, with the constraint that collector and drain contact are at a common ground potential $\mu_c \equiv \mu_d$, and similarly for G_c.] With this identification, Eq. (D3) takes a particularly simple form

$$\frac{V_c}{I_i} = \frac{2e^2}{h}T_{i\to c}\frac{1}{G_i G_c} \; . \tag{D4}$$

The injector-collector reciprocity in electron focusing, discussed in Sec. IV C, is evident from Eq. (D4), since G_i and G_c are symmetric in B and[11] $T_{i\to c}(B) = T_{c\to i}(-B)$.

Equation (D4) is derived for a three-terminal measurement (with the sign of the magnetic field chosen such that electrons are focused from the injector towards the collector). In a four-terminal configuration (with the same sign of the magnetic field), the chemical potential of the collector is measured relative to the chemical potential μ_v of an additional voltage probe, rather than relative to the current drain. If this voltage probe is along the 2D EG boundary between the current drain and the injector, then $\mu_v = \mu_d$ (under conditions of vanishing longitudinal resistance in the 2D EG), so that Eq. (D3) still applies. This is the case of a generalized Hall-resistance measurement, discussed in Sec. IV B. If, on the other hand, the voltage probe is between the drain and the collector, then $\mu_v - \mu_d = eI_iR_H$ is the normal Hall-potential difference in the 2D EG (with $R_H = h/2e^2N_L$ determined by the number N_L of occupied Landau levels in the 2D EG). We then find, instead of Eq. (D4),

$$\frac{V_c}{I_i} = \frac{2e^2}{h}\left[\frac{T_{i\to c}}{G_i G_c} - \frac{1}{N_L}\right] \; . \tag{D5}$$

Equation (D5) corresponds to the case of a generalized longitudinal-resistance measurement, also discussed in Sec. IV B.

---

*Present address: Philips Laboratories, Briarcliff Manor, NY 10510.

†Also at the Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands.

‡Present address: Research Institute for Materials, University of Nijmegen, 6525 ED Nijmegen, The Netherlands.

[1]Yu. V. Sharvin, Zh. Eksp. Teor. Fiz. **48**, 984 (1965) [Sov. Phys.—JETP **21**, 655 (1965)]; Yu. V. Sharvin and L. M. Fisher, Pis'ma Zh. Exp. Teor. Fiz. **1**, 54 (1965) [JETP Lett. **1**, 152 (1965)]; Yu. V. Sharvin and N. I. Bogatina, Zh. Eksp. Teor. Fiz. **56**, 772 (1969) [Sov. Phys.—JETP **29**, 419 (1969)].

[2]V. S. Tsoi, Pis'ma Zh. Exp. Teor. Fiz. **19**, 114 (1974) [JETP Lett. **19**, 70 (1974)]; Zh. Eksp. Teor. Fiz. **68**, 1849 (1975) [Sov. Phys.—JETP **41**, 927 (1975)].

[3]V. S. Tsoi and I. I. Razgonov, Zh. Eksp. Teor. Fiz. **74**, 1137 (1978) [Sov. Phys.—JETP **47**, 597 (1978)]; V. S. Tsoi and N. P. Tsoi, *ibid.* **73**, 289 (1977) [**46**, 150 (1977)]; V. S. Tsoi and Yu. A. Kolesnichenko, *ibid.* **78**, 2041 (1980) [**51**, 1027 (1980)]; V. S. Tsoi, J. Bass, P. A. M. Benistant, H. van Kempen, E. L. M. Payens, and P. Wyder, J. Phys. F **9**, L221 (1979).

[4]P. C. van Son, H. van Kempen, and P. Wyder, Phys. Rev. Lett. **58**, 1567 (1987).

[5]P. A. M. Benistant, G. F. A. van de Walle, H. van Kempen, and P. Wyder, Phys. Rev. B **33**, 690 (1986).

[6]P. A. M. Benistant, A. P. van Gelder, H. van Kempen, and P. Wyder, Phys. Rev. B **32**, 3351 (1985); P. C. van Son, Ph.D. thesis, University of Nijmegen, 1987.

[7]A. F. Andreev, Zh. Eksp. Teor. Fiz. **46**, 1823 (1964) [Sov. Phys.—JETP **19**, 1228 (1964)].

[8]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).

[9]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).

[10]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).

[11]M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986); IBM J. Res. Dev. **32**, 317 (1988).

[12]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. (to be published).

[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]An alternative method for fabricating point contacts is described by I. K. Yanson, Zh. Eksp. Teor. Fiz. **66**, 1035 (1974) [Sov. Phys.—JETP **39**, 506 (1974)].

[16]Point-contact spectroscopy is reviewed by A. G. M. Jansen,