Linear geometrical magnetoresistance effect: Influence of geometry and material composition

Daniel R. Baker and Joseph P. Heremans
Department of Physics and Physical Chemistry, General Motors Research and Development Center, Warren, Michigan 48090-9055
(Received 8 December 1998)

This work reports theorems on geometrical magnetoresistance effects in simply or multiply connected surfaces. We consider planar or three-dimensional device geometries, and assume that the current/voltage relations are linear. It is shown that the resistance of any two-wire device must be an even function of the magnetic field. We further calculate the magnitude of the highest and lowest second-order geometrical magnetoresistance of a planar two-wire device, for given isotropic material parameters. The largest change in magnetoresistance occurs when the boundaries of the device contain only electrically conducting leads and no insulating components. When the Fermi surface and the mobility tensor are isotropic, as is the case for *n*-type InAs and InSb, the magnetoresistance obtained in the Corbino geometry is the largest possible, but this conclusion does not generalize to materials with anisotropic conductivity tensors. The presence of multiple ellipsoidal carrier pockets in Bi also explains why the geometrical magnetoresistance effects are much smaller in the trigonal plane of Bi than in InAs, even though these materials have similar mobilities at room temperature, and the conductivity of Bi is isotropic in that plane. [S0163-1829(99)15421-3]

I. INTRODUCTION

Semiconductor magnetoresistors have been known for decades,[1] and are commonly used in magnetic position sensing applications.[2] Most of these devices make use of electrically conducting inhomogeneities in or on top of a high-mobility semiconductor. For instance, Weiss[1] described InSb magnetoresistors created by precipitating needles of metallic NiSb inside an InSb host material. Planar thin-film InSb magnetoresistors are commercially available.[2] They consist of an InSb film grown on an insulating substrate and etched in the shape of a mesa. Metallic ''shorting bars'' are then deposited on top of the mesa at select locations.[2] The presence of inhomogeneities have also been found to increase the magnetoresistance of $Hg_{1-x}Cd_xTe$ (Ref. 3) over the intrinsic magnetoresistance in these materials,[4] which is already very large due to the very high electron mobility in these materials and the presence of several carrier types. Inhomogeneities in Bi have been studied[5] and modeled,[6] but do not necessarily result in a large geometrical magnetoresistance.[5] Inhomogeneities have even been suspected to cause the remarkable linear magnetoresistances observed in $Ag_{2+\delta}Te$ and $Ag_{2+\delta}Se$.[7] The problem of geometrical magnetoresistances due to inhomogeneities in semiconductors thus has a long history.[8,9]

The geometrical magnetoresistance arises by virtue of the fact that a magnetic field exerts a Lorenz force on moving charged particles, perpendicular to their direction of motion. Such forces change the direction of current flow, as depicted for the rectangular semiconductor device in Fig. 1(a). If either the positive or the negative charge carriers dominate in the current, the change in current flow will cause charge buildup on the edge of the device pictured in Fig. 1(a), until the electric field induced by the charge buildup negates the force from the magnetic field. Figure 2(a) shows the current flow lines, along with the corresponding equipotential lines, that result from this process. A side effect of this is that a potential difference, called the Hall potential, develops between the two edges. The Hall potential always counteracts the effects of the Lorenz force. The change in resistance is increased by shorting out the Hall potential, allowing the current between the leads to develop a transverse component due to the magnetic field. Two different technologies exist to do this. One method takes advantage of surface recombination, if the active region of the device is an intrinsic semiconductor. This allows the positive and negative carriers shown in Fig. 1(a) to cancel each other out by recombining at the device edge, creating a transverse current. This results in a class of devices called magnetodiodes,[2] which have magnetic-field-dependent carrier densities and nonlinear current-voltage characteristics. The other method consists of inserting inhomogeneities, such as ''shorting bars'' of highly conductive material in the device, to short out the Hall potential as shown in Fig. 1(b). In this paper we label ''shorting bars'' regions in the device with zero resistance to which no electrical contact is made; regions to which contact is made are ''leads.'' Hypothetically, the most effective geometry in which the Hall voltage is shorted is the Corbino geometry (Fig. 3). If the contacts are Ohmic, these devices remain purely resistive, with linear current-voltage characteristics. This paper will solely treat such linear devices. It assumes that all transport is diffusive, and ignores ballistic effects. Bergman and Strelniker[9] numerically modeled the influence of various electrically conducting and nonconducting inclusions into semiconducting material, using a full three-dimensional model. In contrast, this paper will focus mostly on planar situations, but we will develop general theorems. The goal of this paper is to answer the following questions.

(1) Given a certain semiconductor material system, what is the optimal device geometry for a planar linear magnetoresistor? In particular, can one develop geometries that give more change of resistance in a magnetic field than the Corbino geometry?

(2) Is it at all possible to build a linear two-wire device with an asymmetric resistance curve? If not, why not? What

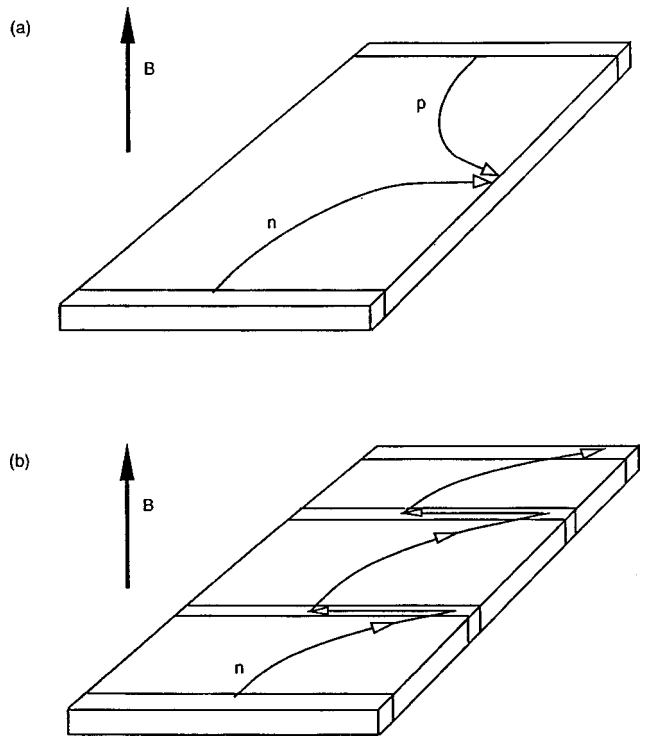

FIG. 1. Positive and negative charge-carrier currents moving in a semiconductor in the presence of a magnetic field. If these carriers recombine at the right edge of the device, as shown in (a), a current, transverse to the direction between the leads at either end of the device, will flow. When one type of charge carrier is dominant, shorting bars can be used, as shown in (b), to create these transverse currents, which, in turn, increase the resistance of the device.

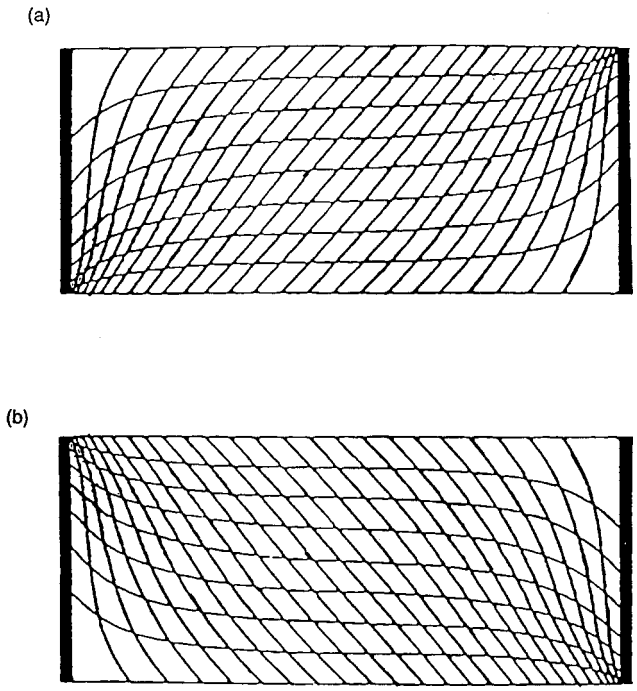

FIG. 2. Current and equipotential lines for a rectangular, two-lead device in a magnetic field pointing out of the page (a) or into the page (b). Note that the transverse component of the current flow distorts the potential on the upper and lower edges of the device, creating a potential difference (the Hall potential) between boundary points on the same vertical line. When the sign of the magnetic field changes, the flow pattern is reflected, as shown in (a) and (b), but the total current flowing through the leads on either end remains unchanged.

distinguishes the two-wire case from the three-wire case, where an asymmetric response is known to be possible?

(3) How does the choice of materials from which the magnetoresistor is made affect its response to magnetic field? For example, why does the semimetal Bi, which has electron mobilities close to those of InAs, give a much smaller geometrical magnetoresistance than InAs?

These questions may seem obvious at first glance. However, most analytical treatments of the current distribution in a slab of semiconductor in the presence of a perpendicular magnetic field used conformal mapping,[10] and are thus not obviously valid for multiply connected geometries, for three-dimensional geometries, or for devices with inhomogeneous material composition. The answer to question (2) starts with the observation that the resistance R of a rectangular magnetoresistor, made from a material such as indium antimonide (InSb), depends on the magnetic field B according to the equation $R(B)=R(-B)$. This fact follows easily from the reflection symmetry of the rectangle about an axis through its leads, and the isotropy of the conductivity tensor for InSb (see Fig. 2). This implies that the resistor will have no first-order response to small changes in magnetic field at $B=0$. It may thus seem desirable to try to disturb the symmetry of the rectangular shape in the hopes of producing an asymmetric resistance curve with a nonvanishing first derivative at $B=0$. The possibility of building in asymmetry is further motivated by the fact that it is possible to build three-lead linear devices, measuring Hall potential or transverse currents, which in fact do have first-order responses at $B=0$. A number of suggestions have been made over the years for disturbing this geometric symmetry. The simplest of these was to build a trapezoidal device, thereby eliminating the reflection symmetry of the rectangle. More complicated suggestions involve multiply connected geometries, including device designs with holes placed asymmetrically, with ''shorting bars'' or patches of material with zero resistance

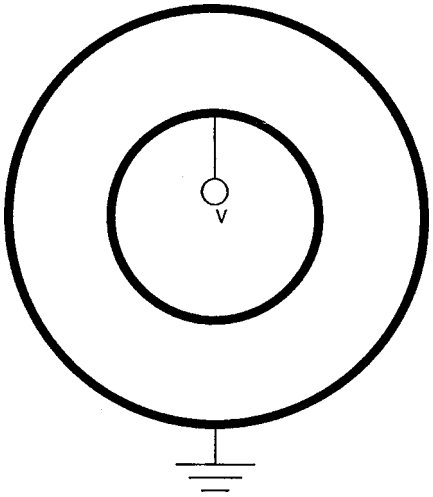

FIG. 3. A Corbino disk, an annular magnetoresistor whose entire boundary is covered by metal leads, holding each component of the boundary at constant potential.

to establish asymmetric equipotential lines in the device, or designs that use different materials, with different magnetoresistive properties, in different regions of the device. Recent experimental studies[11,12] on four-probe $Hg_{1-x}Cd_xTe$. Corbino disks have shown asymmetrical responses to magnetic field. Such responses would contradict theorem 1 of this work unless the device contains an element that departs from one of the assumptions used in the present paper, such as a nonlinear element.

Under very general assumptions, two-wire linear devices will always have resistance curves which are invariant under sign change of the magnetic field. The property is a consequence of a general transformation principle, which holds, not only for arbitrary device geometries and conductivity tensors, but also for devices with more than two leads, such as the three-wire Hall sensor or magnetotransistor.[2] This transformation principle, stated in theorem 1 of Sec. III is in turn a consequence of a relationship from nonequilibrium thermodynamics, known as the Onsager relations,[13] which describe a property of the conductivity tensor. In Appendix A, we give a proof of theorem 1, based only on the model equations and the Onsager relations. This proof can be understood as a restatement of the Onsager relations in macroscopic terms. All that is needed to go from the tensor formulation to the macroscopic form is the conservation of charge and the fact that the electric field is the gradient of a potential function. The main contribution of this result is to remind us of some practical macroscopic consequences of the Onsager relations.

The invariance under sign change of magnetic field implies that no linear two-wire device can exhibit a first-order response to magnetic field. It is, however, still of interest to know how the second-order response can be maximized, in particular, what effects the shape of the magnetoresistor and the configuration of its leads will have on its response. For planar devices made of materials whose conductivity tensors are isotropic in the plane of the device, we will show that the strongest response is obtained by a resistor whose leads fill the entire boundary of the device, leaving no insulating boundary components. An example of such a device configuration is the Corbino disk, shown in Fig. 3, but our results imply that, for a device of homogeneous material properties, any device geometry having no insulating boundary components will exhibit the same sensitivity to magnetic field as the Corbino disk. All such planar shapes must, however, contain holes. On the other end of the scale, the worst sensitivity is obtained by devices with short leads and large insulating components on their boundaries. These statements are made more precise by explicit theoretical bounds for device response, given in terms of the intrinsic conductivity tensor of the resistive material. One can define the geometrical magnetoresistance as the difference between the best and worst response curves.

The above results also provide us with a means of answering question (3), since the theoretical bounds for device response are given in terms of the conductivity tensor of the resistive material. In Sec. IV we plot the response curves for three different materials: InSb, InAs, and Bi. It can be seen that InSb has on the order of a five times better response than InAs. This effect is due to the difference in electron mobility between InSb and InAs. It can also be seen that InAs has a larger geometrical magnetoresistance than Bi, even though the average electron mobility is roughly the same in both materials, and the conductivity tensor of Bi is isotropic in the trigonal plane. This effect is due to the presence of multiple carrier pockets in the electron Fermi surface of Bi.

Section II of this paper presents the model equations for the linear devices. Section III gives the main results just discussed, and Sec. IV contains some examples and discussion. Appendix A gives proofs of the theorems from Sec. III. Appendix B describes how to use conformal mapping techniques to calculate the response curves for some linear devices. Appendix C gives derivations of the conductivity tensors for InSb, InAs, and Bi.

II. THE MODEL EQUATIONS

The models to be described are two dimensional. The device to be modeled will be assumed to lie in the x-y plane, with a magnetic field $\mathbf{B}$ perpendicular to the plane of the device. As such, $\mathbf{B}$ is characterized by the scalar quantity B, which is its component in the z direction.

Under the assumption that the carrier densities remain independent of both position and magnetic field, the vector current $\mathbf{J}$ in the device is related to the gradient of the potential $\nabla\psi$ via a conductivity tensor σ, such that

$$\mathbf{J}=\sigma\cdot\nabla\psi. \tag{2.1}$$

The matrix notation used here assumes that both $\mathbf{J}$ and $\nabla\psi$ are column vectors, with $\nabla\psi=(\psi_x,\psi_y)^T$, where $\psi_x=\partial\psi/\partial x$, $\psi_y=\partial\psi/\partial y$, and the superscript T denotes transpose. The matrix σ will be a function of magnetic field B, such that

$$\sigma(-B)=\sigma(B)^T.$$

This is the tensor form of the Onsager relations.[13]

The current $\mathbf{J}$ can be obtained from $\nabla\psi$ by first multiplying $\nabla\psi$ by a scale factor r and then rotating it through some angle α. In general, the values of r and α depend on the direction of the vector $\nabla\psi$, and in this case the conductivity tensor is anisotropic. However, in many situations, the values of r and α are independent of the direction of $\nabla\psi$, a fact which will simplify many of our calculations. If this is so, the conductivity tensor is referred to as isotropic, and takes the form

$$\sigma=\begin{bmatrix}\sigma_{xx} & -\sigma_{xy}\\ \sigma_{xy} & \sigma_{xx}\end{bmatrix}.$$

In what follows, it will be useful to identify the vectors (x,y) in the plane with complex numbers $w=x+iy$, so that $\nabla\psi=\psi_x+i\psi_y$ and $\mathbf{J}=J_1+iJ_2$. If the conductivity tensor is isotropic, one can write $\sigma=\sigma_{xx}+i\sigma_{xy}$, so that the matrix multiplication in Eq. (2.1) can be interpreted as a multiplication of the two complex numbers σ and $\nabla\psi$. The complex number σ now has the form $re^{i\alpha}$, and the Onsager relation can now be written as

$$\sigma(-B)=\overline{\sigma(B)}.$$

Assuming that there are two leads placed on the device, in a fashion shown in Fig. 2, we would like to calculate the

resistance V/I of the device, as a function of the magnetic field B. To do this, we solve the partial differential equation.

$$\nabla \cdot (\sigma(B) \nabla \psi) = 0, \tag{2.2}$$

subject to the boundary conditions,

$$\psi = V \quad \text{on the left lead,}$$

$$\psi = 0 \quad \text{on the right lead,}$$

$$\mathbf{J} \cdot \mathbf{n} = 0 \quad \text{on the rest of the boundary,}$$

where $\mathbf{n}$ is a normal vector to the boundary at the point in question. The total current through either lead is then given as

$$I = \int \mathbf{J} \cdot \mathbf{n},$$

where the integral is taken over the entire lead.

If the conductivity tensor is isotropic, lemma 1 of Appendix A implies that Eq. (2.2) is equivalent to Laplace's equation, and the potential is a harmonic function on the interior of the device region.

The model generalizes to multilead devices having an arbitrary planar geometry. Assume that the device occupies some bounded subset D of the plane, whose boundary is given by smooth curves, possibly with a finite number of corners. This boundary is divided up into segments which are leads and segments which are insulating arcs. The leads can be thought of as infinitely conducting metal strips placed along a segment of the boundary, and holding that entire boundary segment at a constant potential. Along the insulating arcs, the normal component to the current must always vanish. As such, the boundary condition $\mathbf{J} \cdot \mathbf{n} = 0$ specifies on insulating arcs that $\nabla \psi$ makes either an angle $-\alpha$ or an angle $\pi - \alpha$ to the boundary. The device region D may have a finite number of holes in it, and leads may be placed on part or all of the boundaries of these holes.

To maximize the geometrical magnetoresistance it is useful to include shorting bars in the device as shown in Fig. 1(b), i.e., metal bars placed along the interior of the device to prevent the buildup of Hall potential in the presence of magnetic fields. Such shorting bars can be thought of as slit-shaped holes in the device region. The entire hole is filled with metal, so that the entire boundary of the hole becomes one lead, held at a constant potential.

We will also consider devices in which D is divided up into subregions, each having its own conductivity tensor. In general, the conductivity tensor in each subregion may vary continuously, but across the boundary between subregions the conductivity tensor is allowed to jump discontinuously, symbolizing a discrete change in the materials in the two regions. This boundary may be a shorting bar, in which case a constant voltage condition will be placed along the interface. If no shorting bar is used, we will require that both the potential and the normal component to the current be continuous along the interface between regions. It should be noted that, in general, when two regions with different conductivity tensors come in contact with each other, diffusion currents and charge buildup will occur along the interface, causing the model equations to become nonlinear. There are, however, exceptions to this. These include Ohmic metal-semiconductor interfaces, in particular, situations where it is desirable to model shorting bars as having a small, finite resistance, and it will also be convenient to model linear contact resistances which occur along metal leads as thin, highly resistant layers of material, surrounding the lead, having no sensitivity to magnetic field. Such models are also of use for Ohmic semiconductor-semiconductor contacts.

Boundary conditions are specified for the device by giving either current or voltage conditions along each lead. A voltage condition specifies a constant potential V along the entire lead. A current condition also requires that the potential along the lead remain constant, but its value is determined by the equation

$$I = \int \mathbf{J} \cdot \mathbf{n},$$

where I is the total current flowing through the lead, and the integral is taken over the entire lead. Note that, although, in general, current will be flowing through shorting bars in the device, the current flows in one side of the slit and out the other side, so that the total current is always zero. Thus shorting bars are to be treated as leads on which zero current conditions are imposed.

The inverse to the conductivity tensor is the resistivity tensor

$$\rho(B) = \sigma(B)^{-1},$$

so that one has

$$\nabla \psi = \rho \mathbf{J}.$$

When the conductivity tensor is isotropic, so is the resistivity tensor, and we will write

$$\rho = \rho_{xx} + i\rho_{xy},$$

where

$$\rho_{xx} = \frac{\sigma_{xx}}{|\sigma|^2} \quad \text{and} \quad \rho_{xy} = \frac{-\sigma_{xy}}{|\sigma|^2}.$$

Of the theorems stated in Sec. III, theorem 1 is valid in the more general context of three-dimensional device geometries. In such situations, the conductivity tensor is a 3×3 matrix, still satisfying the Onsager relations. Equation (2.2), and the boundary conditions given below it, are still valid.

III. MAIN RESULTS

The theorems stated in this section are proven in Appendix A. Consider the first-order response to the magnetic field of the rectangular three-wire device shown in Fig. 4(a). Both of the leads on one side of the device are grounded, and the lead on the opposite side is held at a constant voltage. Note that the two leads on one side are placed asymmetrically, so as to destroy the symmetry which previously existed in the two-wire device. Appendix B shows how to calculate the current curves through each of the leads, as a function of the angle α of deflection due to the magnetic field, and the results are depicted in Fig. 4(b). It is apparent that both of the two grounded leads have asymmetric current responses. Note

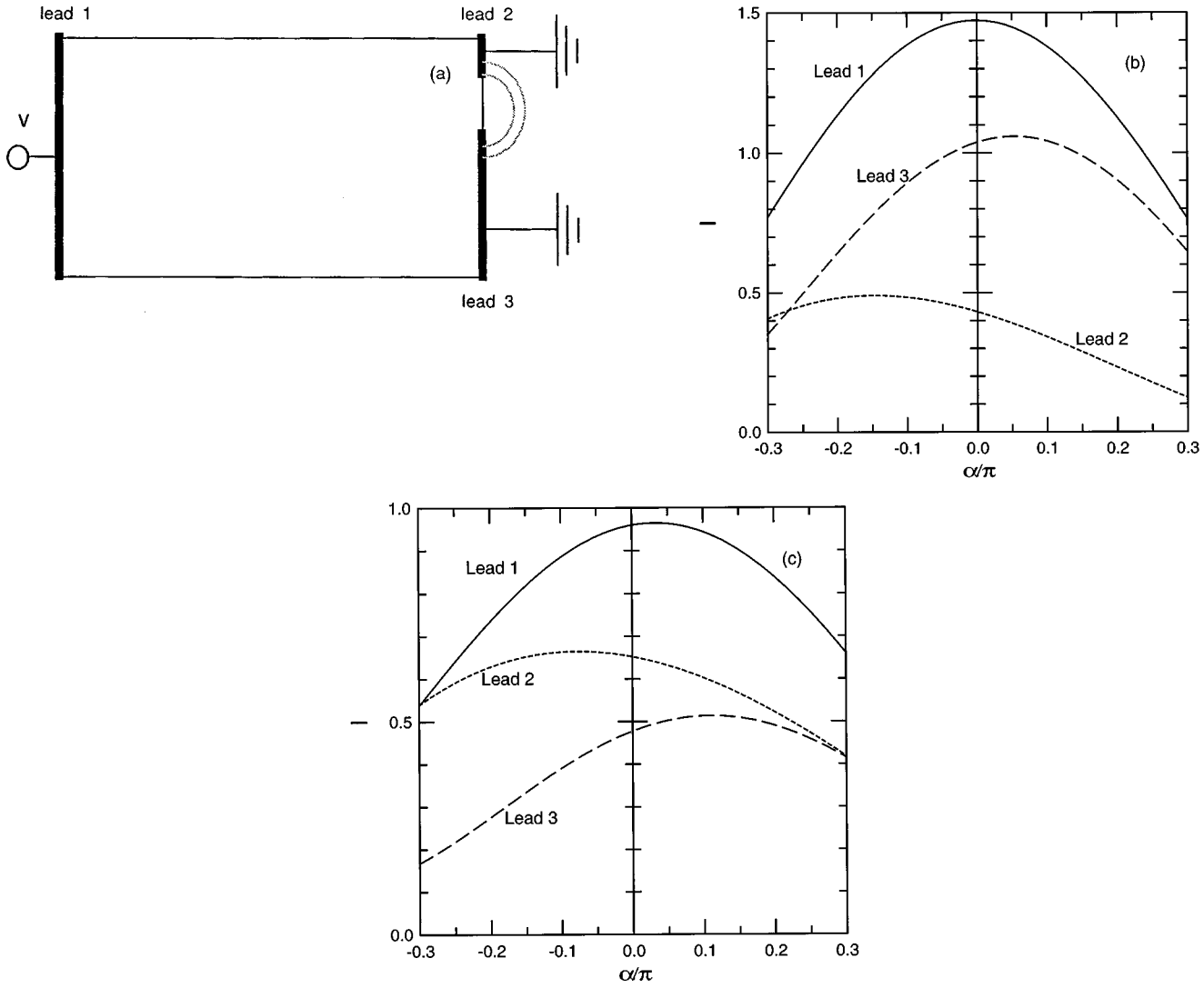

FIG. 4. (a) A three-lead magnetoresistor need not have current curves which are invariant under sign change of the magnetic field. When two of the leads are grounded, these leads can be thought of as being connected by a metal bridge, represented by the dotted lines. The resulting device can be thought of as a two-lead device with a hole in it. Part of the hole is bounded by semiconductor material, and part of it is bounded by the metal bridge. For materials such as InSb or InAs, (b) shows plots of dimensionless quantities which are proportional to the current curves through each of the three leads, as a function of α/π, where α is the angle of deflection due to magnetic field. (The proportionality constants are material dependent, and are described at the end of Appendix B.) Although each of the two grounded leads has an asymmetric current curve, the sum of the currents, which flows through lead 1, is symmetric under sign change of the magnetic field, i.e., of α. When lead 3 is set at half the voltage of lead 1, with lead 2 grounded, all three current responses become asymmetric, as shown in (c), which again shows dimensionless quantities proportional to the current.

further that, if any reflections of this device existed of the type used to prove symmetry in the two-lead case, they would be observable in the current curves for the individual leads. For example, if it were possible to reflect the device onto itself, permuting the three leads so that lead i is reflected onto lead j, this would imply that the currents $I_i(B) = I_j(-B)$. Since Fig. 4(b) shows that no such symmetries exist in the current curves, the geometric reflections cannot exist either.

It is tempting to convert this device into a two-wire device by shorting the two leads on the one side together, as indicated by the dotted line in Fig. 4(a). The resulting device can be thought of as a two-lead device with a hole in it. One part of the hole is bounded by the metal of the shorting bar, and the remainder of the boundary is bounded by semiconductor material. The previous arguments still show that no analytic reflections can demonstrate symmetry under sign reversal of the magnetic field, but the current which can be measured is now the sum of the currents in the two joined end leads, which must be equal to the current flowing through the lead on the opposite end. As one can see from Fig. 4(b), this current curve is symmetrical, so that somehow, in spite of all the built in asymmetries, $I(B)$ will be equal to $I(-B)$.

More generally, one can consider devices with arbitrary numbers of holes, regions with zero resistance built into the device in various ways to create equipotential lines, and materials with different magnetoresistive properties, all used in order to disturb the geometric symmetry of the device. Although the numerical simulation of such a linear two-wire device is potentially very complicated, it turns out that it will always have a symmetric current response, and this fact con-

tinues to be true, even if the conductivity tensors are anisotropic.

To understand why this is so, let us first assume that the device in question has an arbitrary number of leads, say $n+1$. In this context, we will regard regions with zero resistance built into the device as leads, on which we impose a zero current condition. Some of the leads may be placed on part, or all, of the boundaries of the holes. Note that the sum of the currents from the different leads must be zero, so that the current through the $(n+1)$st lead is equal to minus the sum of the currents through the first n leads. Since we can choose the zero potential arbitrarily, we will also assume that the $(n+1)$st lead is always at zero voltage. We denote by $\mathbf{V}=(V_1,\ldots,V_n)^T$ an n vector of voltages which are to be applied at each of the remaining n leads, and by $\mathbf{I}=(I_1,\ldots,I_n)^T$ the currents through these leads which result when the voltages $\mathbf{V}$ are applied. The linearity of the model equations implies that there is a matrix $\mathbf{R}(B)$, such that

$$\mathbf{V}=\mathbf{R}(B)\cdot\mathbf{I},$$

where, as indicated by the notation, the matrix $\mathbf{R}$ depends on the magnetic field B.

We remark that the determination of the matrix $\mathbf{R}$ for a given device involves solving the partial differential equation (2.2) for n linearly independent sets of current boundary conditions and determining the voltages associated to them, and, in general, this is a nontrivial undertaking. For simply connected device regions, conformal mapping techniques[14,15,10] can produce solutions in the form of integrals which must be evaluated numerically. These techniques, as discussed in Appendix B, were used to calculate the various simulation results presented here.

Theorem 1: Let D be any device region, two or three dimensional, with a finite number of holes (possibly none), containing any number of regions with zero resistance and leads, and subregions with different conductivity tensors, which need not be isotropic and which may vary as a function of position. The matrix $\mathbf{R}(B)$ for such a linear device is equal to the transpose of the matrix $\mathbf{R}(-B)$.

Remark: Unless physical symmetries are present in the device region D, as in the case of a rectangular device, there is no simple way to transform the current flow pattern associated with B to that associated with $-B$. Thus, while the above theorem is, in some sense, an integrated version of the Onsager relation, the proof given for rectangular two-lead devices does not generalize. In fact, it does not even generalize to the case of a rectangular device with an anisotropic conductivity tensor.

Assuming now that the voltages $\mathbf{V}$ are held constant as we vary the magnetic field, the theorem implies that

$$\begin{aligned}\mathbf{I}(-B)^T\cdot\mathbf{V}&=\mathbf{I}(-B)^T\cdot\mathbf{R}(B)\cdot\mathbf{I}(B)\\&=\mathbf{I}(B)^T\cdot\mathbf{R}(-B)\cdot\mathbf{I}(-B)=\mathbf{I}(B)^T\cdot\mathbf{V}.\end{aligned}\quad(3.1)$$

For a device with only two leads, one of the leads is at zero voltage, so that the vectors $\mathbf{I}$ and $\mathbf{V}$ are one dimensional. Equation (3.1) thus implies that the current through the device is invariant under the sign change of the magnetic field. If the device in question also has regions with zero resistance, as already noted, these regions can be treated as leads, on which zero current conditions are imposed. In such cases the product of voltage and current from this lead will always be zero, so that the above argument generalizes to two-wire devices containing regions with zero resistance as well.

One also gains some insight into the behavior of the three-wire device in Fig. 4(a). If both leads on one side are grounded ($\mathbf{V}_i=0$), then Eq. (3.1) only yields information about the current flowing through the lead on the other side, and it implies that this current must be invariant to sign changes in magnetic field. On the other hand, if each of the three leads has a different voltage applied to it, one arrives at a more complicated symmetry. Assuming, say, that $V_3=0$, one then finds that

$$V_1I_1(B)+V_2I_2(B)=V_1I_1(-B)+V_2I_2(-B).\quad(3.2)$$

In general, none of the currents will respond symmetrically to the magnetic field. Only the sum from Eq. (3.2) will have this property. Figure 4(c) shows the current responses of the three-wire device, when the third lead is at half the voltage of the first lead, and the second lead is grounded.

Note that the dot product $\mathbf{I}(B)^T\mathbf{V}$ is the power dissipated by the resistor, in the magnetic field B and with the voltage conditions $\mathbf{V}$. As such, Eq. (3.2) states that the power dissipation under constant voltage conditions is invariant under sign change of the magnetic field.

There are some other consequences of theorem 1 which are worth noting. As shown in Fig. 4(c), when each of the leads of a three-lead device have different fixed voltages, none of the current curves will, in general, be symmetric under sign reversal of magnetic field. One way to try to make all the leads have different voltages is to place a resistor between the second and third lead, as shown in Fig. 5(a). Such configurations can be produced by appropriate manufacturing processes, so that no extra wires must be soldered to the leads, and the result would still constitute a two-wire device. Unfortunately, the resistor will not hold the third lead at a fixed voltage, as the magnetic field varies. In fact, the configuration in Fig. 5(a) can be modeled as the device in Fig. 5(b), where two different materials are used in the device, separated by a shorting bar. The second material represents the resistor, and has no sensitivity to magnetic fields. This can be modeled by using a conductivity tensor with no B dependence. Theorem 1 now implies that the resistor configuration will also produce a response which is invariant under sign reversal of the magnetic field. In fact, the same conclusion follows for any composite of different devices, connected together in arbitrary ways with linear resistors.

On the other hand, the above analysis becomes invalid if the resistor used in Fig. 5(a) is nonlinear, for instance if it were replaced by a diode. Since holding all three leads at different constant voltages produces an asymmetric response, any two-wire electronic component which comes close to maintaining such conditions over the range of currents going through it for practical values of B will produce asymmetry in the current curves. Although it is conceivable to design nonlinear devices to produce this type of response, this approach is not further considered here.

The conclusion to be drawn from all of this is that it will not be possible to obtain a first-order response to magnetic field from a linear two-wire device. We now turn our attention to calculating the strength of the second-order response from such a device: in particular, how the geometry of the

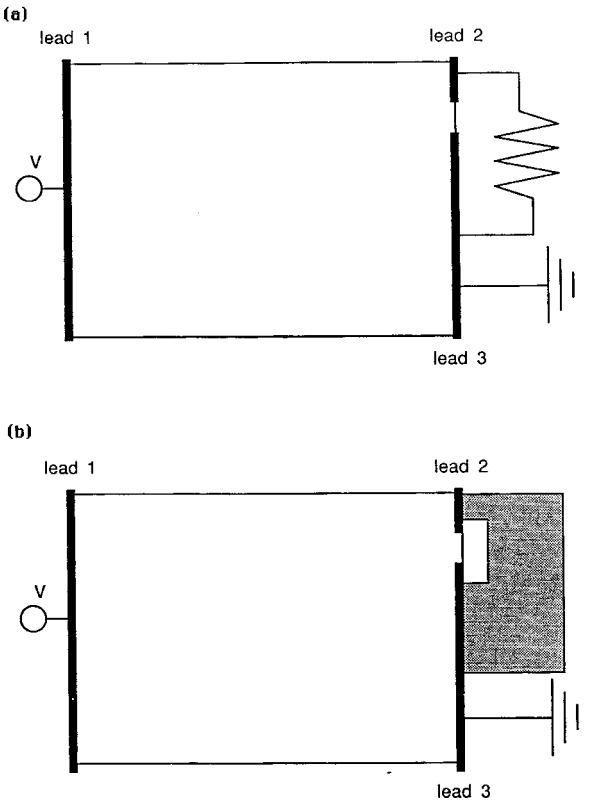

FIG. 5. A three-lead device can be used as a two-lead device by connecting two of the leads with a resistor, as in (a). This can be modeled as a magnetoresistor composed of two different substances, as shown in (b), where the shaded substance will have no sensitivity to magnetic field. As such, the analysis of theorem 1 applies, and the device will perform invariantly under sign change of magnetic field.

device influences this response. Our goal is to determine what physical device configurations produce the best, and worst, responses to magnetic field, and to derive explicit expressions for these responses, in terms of the conductivity tensor of the semiconductor material, or materials, involved. We will start by assuming that only one type of material, with an isotropic conductivity tensor, is to be used in the device, and state the results for this case. The modifications to the more general situation will then be discussed.

The bounds on resistance that we seek are equivalent to bounds on the power dissipation of the device, i.e., the scalar product $\mathbf{I}^T\mathbf{V}$. The results can best be proven in this form, because the power dissipation can be expressed as an integral over the device region, whereas the resistance cannot. It should, however, be remarked that the power dissipation is assumed to be small enough so that the device remains at uniform temperature. Specifically, the temperature gradients are small enough so that they do not contribute to current flow in the device through a Peltier effect, or create spatial variations in the conductivity tensor. Note that, if the conductivity or resistivity tensor is isotropic, then the Onsager relation implies that $\sigma(0)$ and $\rho(0)$ have no imaginary parts in their representation, i.e., the potential gradient will point in the same direction as the current.

Theorem 2: Let D be any planar device region with a finite number of holes (possibly none), containing any number of regions with zero resistance and leads, and assume that the conductivity tensor is isotropic and constant throughout the entire device region. Define boundary conditions by specifying either a fixed voltage or zero current conditions on the leads and regions with zero resistance. Then

$$\mathbf{V}^T\mathbf{I}(B) \geqslant \frac{\sigma_{xx}(B)}{\sigma(0)}\mathbf{V}^T\mathbf{I}(0).$$

If the boundary of the device contains only leads and no insulating components, then the inequality becomes an equality.

Theorem 3: For the same device region as in theorem 2, define boundary conditions by specifying current conditions on each lead and region with zero resistance of the device. Then

$$\mathbf{V}(B)^T\mathbf{I} \geqslant \frac{\rho_{xx}(B)}{\rho(0)}\mathbf{V}(0)^T\mathbf{I}.$$

Corollary: For a planar, linear two-wire device, having an arbitrary geometry with a finite number of holes (possibly none) and containing an arbitrary number of regions with zero resistance, assume that the conductivity tensor is isotropic and constant throughout the device region. Define

$$\sigma_1(B) = \sigma_{xx}(B)/\sigma(0)$$

and

$$\rho_1(B) = \rho_{xx}(B)/\rho(0).$$

Then the resistance of the device must satisfy the inequalities

$$\rho_1(B)R(0) \leqslant R(B) \leqslant R(0)/\sigma_1(B).$$

If the boundary of the device contains only leads and no insulating parts, the upper bound on $R(B)$ becomes an equality.

Proof. Applying theorem 2 to the resistor, only two leads will have nonzero total currents flowing through them, and one of these leads can be assumed to be grounded. The products $\mathbf{V}^T\mathbf{I}(B)$ thus become products of scalars, and

$$R(B) = V/I(B) = V^2/(VI(B)) \leqslant V^2/(\sigma_1(B)VI(0))$$
$$= R(0)/\sigma_1(B).$$

Applying theorem 3 to the resistor, again, the total currents through all but two of the leads can be assumed to be zero, the products $\mathbf{V}(B)^T\mathbf{I}$ have only one term in the sum, and one has

$$R(B) = V(B)I/I^2 \geqslant \rho_1(B)V(0)I/I^2 = \rho_1(B)R(0).$$

The corollary is thus proven.

The significance of this corollary is that it determines the maximum and minimum possible geometrical magnetoresistance in terms of the intrinsic properties of conductivity and resistivity, and no calculations involving the geometry of the device are involved. Examples of this are given in Sec. IV.

One can show that the lower bound on resistance is attained, for example, in the limiting case of a rectangular two-lead device, as the lengths of the insulating sides of the device go to infinity. More generally, an examination of the proofs of theorems 2 and 3, in Appendix A, shows that the

resistance of any device shows more sensitivity to magnetic field as the leads on the device are made larger and the insulating arcs along the device boundary become smaller.

For theorem 2, these results can be generalized to the case of devices whose conductivity tensors vary as a function of position, but remain isotropic, and they are proven in Appendix A in this form. The factors $\sigma_1(B)$ must be generalized to take on their minimum value over any region in the device, and the inequality of theorem 2 cannot become an equality, if there are different values for $\sigma_1(B)$ in different regions of the device. Even in the case of anisotropic conductivity tensors, theorem 2 generalizes to yield an upper bound on the resistance of two-wire devices. The bound is, however, no longer sharp, i.e., it may not actually be attained by any device. Furthermore, it seems likely that devices such as the Corbino disk will no longer provide the best response to magnetic field in such situations. The analysis in theorem 3 does not seem to generalize at all to anisotropic conductivity tensors. A discussion of the details of both theorems, pertaining to the anisotropic case, is given at the end of Appendix A.

In situations with varying conductivity tensors, the bounds on power dissipation are of some interest in their own right, because they help us understand how these different conductivity tensors influence the behavior of the device. For example, normally a magnetoresistor will also have a contact resistance along the leads. As already noted in Sec. II, this contact resistance can be modeled as a thin subregion along the leads, having a conductivity tensor which is independent of the magnetic field. However, the addition of this extra subregion can substantially change the functions $\sigma_1(B)$ and $\rho_1(B)$, and thus the sharpness of the above estimates on magnetoresistivity. Sharper estimates can be obtained by examining the proofs of these theorems. In these proofs, the total power dissipation is evaluated as the sum of the power dissipation over each subregion. For the case of contact resistance, this power dissipation should be very small in comparison to the total power dissipated by the magnetoresistor. If this is the case, the conductivity tensor for this subregion can be ignored in the definition of $\rho_1(B)$, with little effect on the integrals and the above estimates.

IV. SOME EXAMPLES AND DISCUSSION

The explicit dependence of the conductivity tensor σ depends on the detailed shape of the Fermi surface of the semiconductor. We consider the case of two classes of solids: narrow-gap III-V compounds, and bismuth. The choice of these materials is motivated by the fact that these are the solids that have the highest mobilities at high electron densities where the materials are degenerate and the contacts are ohmic.[16] Numerical values for these conductivity tensors are given in Appendix C. InSb and InAs have isotropic conductivity tensors. We have used room-temperature electron mobilities for these materials that are characteristic of thin films,[17] not of bulk material. The tensor for Bi is isotropic, when restricted to the plane perpendicular to the trigonal axis (see Appendix C), even though it is anisotropic in all of three-dimensional space. The mobility values used for Bi are bulk values.[18]

Figure 6 shows a graph of what the bounds on magnetoresistance, derived in Sec. III look like for these materials. The results indicate that InSb will produce a five times better geometrical magnetoresistance than InAs, which is consistent with the difference in mobilities between the two materials. However, InAs produces a much larger response than Bi, even though the electron mobility used for InAs is comparable to the average in-plane electron mobility used for Bi. The presence of multiple carrier pockets for electrons in the Brillouin zone of Bi lowers its average geometrical magnetoresistance.

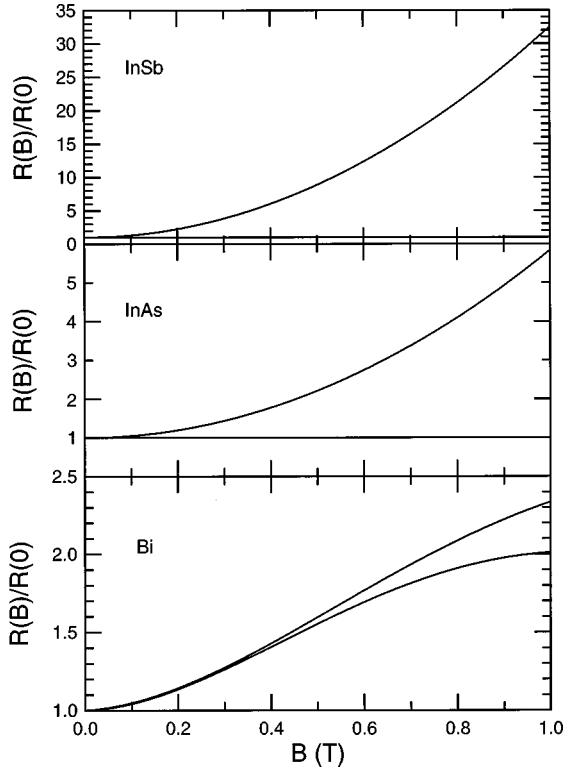

FIG. 6. Relative resistance for the best and worst geometries of devices made from InSb, InAs, and Bi. Note that, for Bi, there is almost no difference in resistance at low-field strengths between the best and worst geometries.

Theorem 2 states that the largest change in magnetoresistance occurs when the leads of a device cover its entire boundary, so that no insulating components are present. This is the case for a Corbino disc, and it is worth noting that any planar region whose leads can be placed in this way must contain holes. This is because the region must have at least two different boundary components held at different voltages, in order not to short out the leads, and the second boundary component must bound a hole. Devices of this type have an inherently low resistance and the processing of such devices may be complicated by the need to make contact to the lead in the hole. The bounds given in the corollary still provide useful guidelines for assessing the responses of practical[2] but suboptimal designs, which can, in turn, be simulated using the techniques described in Appendix B.

ACKNOWLEDGMENTS

The authors would like to acknowledge useful conversations with Dr. D. L. Partin and Thaddeus Schroeder.

APPENDIX A: PROOFS OF THE THEOREMS

Theorem 1: Let D be any device region, two or three dimensional, with a finite number of holes (possibly none), containing any number of regions with zero resistance and leads, and subregions with different conductivity tensors, which need not be isotropic and which may vary as a function of position. The matrix $\mathbf{R}(B)$ for such a linear device is equal to the transpose of the matrix $\mathbf{R}(-B)$.

Proof: We will first prove the case where only one material is used in the device with one conductivity tensor, $\sigma(B)$, which may vary as a function of position. The Onsager relation

$$\sigma(-B)=\sigma(B)^T$$

implies that the dot products of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ obey a transformation law

$$\mathbf{v}_1\cdot(\sigma(B)\mathbf{v}_2)=(\sigma(-B)\mathbf{v}_1)\cdot\mathbf{v}_2.$$

Consider now two functions $\psi_i(x,y)$, $i=1$ and 2, which are to be thought of as potential functions on the interior of the device region. Associated to these potential functions are vector currents $\mathbf{J}_i$, satisfying the identities

$$\mathbf{J}_1=\sigma(-B)\nabla\psi_1 \quad \text{and} \quad \mathbf{J}_2=\sigma(B)\nabla\psi_2.$$

In particular, we assume that

$$\nabla\cdot\mathbf{J}_i=0,$$

so that

$$\nabla\cdot(\psi_1\mathbf{J}_2)=\nabla\psi_1\cdot\mathbf{J}_2=\nabla\psi_1\cdot(\sigma(B)\nabla\psi_2)$$
$$=(\sigma(-B)\nabla\psi_1)\cdot\nabla\psi_2=\nabla\cdot(\psi_2\mathbf{J}_1).$$

Letting D be the device region, with boundary ∂D, the divergence theorem now implies that

$$\int_{\partial D}\psi_1\mathbf{J}_2\cdot\mathbf{n}=\int_D\nabla\cdot(\psi_1\mathbf{J}_2)=\int_D\nabla\cdot(\psi_2\mathbf{J}_1)=\int_{\partial D}\psi_2\mathbf{J}_1\cdot\mathbf{n}. \tag{A1}$$

Let us now pick two sets of current conditions for the device, $\mathbf{I}_1$ and $\mathbf{I}_2$. Let ψ_1 be the potential function which determines the conditions $\mathbf{I}_1$ in the presence of a magnetic field $-B$, with corresponding voltage conditions $\mathbf{V}_1$. The potential function ψ_2 is likewise chosen to determine the conditions $\mathbf{I}_2$ in the presence of a field B, with voltage conditions $\mathbf{V}_2$. One then has the equations

$$\mathbf{V}_1=\mathbf{R}(-B)\cdot\mathbf{I}_1 \quad \text{and} \quad \mathbf{V}_2=\mathbf{R}(B)\cdot\mathbf{I}_2. \tag{A2}$$

On the boundary components of the device which are not leads, the currents $\mathbf{J}_i\cdot\mathbf{n}$ will vanish, and the integrals from Eq. (A1) above will vanish on these boundary components as well. On the leads, the potential functions ψ_i will have the constant values given by $\mathbf{V}_i$. Equation (A1) now implies that

$$\mathbf{I}_2^T\cdot\mathbf{V}_1=\mathbf{I}_1^T\cdot\mathbf{V}_2. \tag{A3}$$

Substituting Eq. (A2) into the above equation yields

$$\mathbf{I}_2^T\cdot\mathbf{R}(-B)\cdot\mathbf{I}_1=\mathbf{I}_1^T\cdot\mathbf{R}(B)\cdot\mathbf{I}_2.$$

Since this equation holds for arbitrary $\mathbf{I}_1$, and $\mathbf{I}_2$, it follows that

$$\mathbf{R}(-B)=\mathbf{R}(B)^T.$$

If the device is composed of several regions with discontinuities in the conductivity tensors across the boundaries between regions, equation (A1) will still hold in each region, but there is no reason to assume constant potential along the interface between two regions. Still, both the potential and the normal component of the currents will be continuous across such an interface, and as integrals (A1) are performed on each region, and then summed together, the interface terms must cancel each other. Thus Eq. (A3) will still be valid for such devices, and the conclusion of the theorem still follows.

We turn now to the proof of theorems 2 and 3. Recall from Sec. II that we are thinking of vectors in the plane as complex numbers, so that isotropic conductivity and resistivity tensors can be viewed as complex numbers which are multiplied by $\nabla\psi$ and $\mathbf{J}$. For vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, and a complex number z_0, the following identity will be used:

$$\mathbf{v}_1\cdot(z_0\mathbf{v}_2)=(\overline{z_0}\mathbf{v}_1)\cdot\mathbf{v}_2,$$

where $\overline{z_0}$ is the conjugate of z_0. The following lemma will be useful in the proofs of the theorems.

Lemma 1: Suppose $\nabla^2\psi=0$, and let $z_0=a+ib$ be any complex number. Then

$$\nabla\cdot(z_0\nabla\psi)=0.$$

Proof. Using vector cross products, we can write $z_0\nabla\psi$ as

$$a\nabla\psi+b\mathbf{k}\times\nabla\psi,$$

where $\mathbf{k}$ is the unit vector in the z direction, perpendicular to the x-y plane. The lemma now follows from the identities

$$\nabla\times\nabla\psi=0,$$

$$\nabla\cdot(\mathbf{v}_1\times\mathbf{v}_2)=\mathbf{v}_2\cdot(\nabla\times\mathbf{v}_1)-\mathbf{v}_1\cdot(\nabla\times\mathbf{v}_2).$$

The version of theorem 2 stated below allows different regions with varying conductivity tensors, but it is assumed that each conductivity tensor is isotropic. Theorem 2 can be generalized to the case of anisotropic conductivity tensors, and a discussion of this generalization is given at the end of the Appendix. In order to facilitate the treatment of this more general case, lemma 2, which is used in the proof of theorem 2, does not assume isotropic conductivity tensors. At this time we do not know how to generalize theorem 3 to the case of anisotropic conductivity tensors or of tensors that vary as a function of position. A discussion of the problems involved is also given at the end of this appendix.

In what follows, the function $\psi(B)$ will be the solution to

$$\nabla\cdot(\sigma(B)\nabla\psi(B))=0, \tag{A4}$$

satisfying given boundary conditions, in the presence of the magnetic field B. Noting that Eq. (A4) is a special case of the so-called quasiharmonic equation, the following lemma gives the standard interpretation of this equation as a minimization problem, as stated, e.g., in Chapter 10.5 of Ref. 19.

The interpretation is of particular interest when $B=0$, and lemma 2 is stated only for this case.

Lemma 2: Let D be any planar device region, containing regions with zero resistance and leads, and subregions with different conductivity tensors, which need not be isotropic and may vary as a function of position. Define boundary conditions by specifying either a fixed voltage, or zero total current on each lead and shorting bar of the device. Then $\psi(0)$ minimizes

$$\int_D (\sigma(0)\nabla\psi)\cdot\nabla\psi$$

over all functions ψ, continuously differentiable on the interior of D, continuous on the boundary of D, and satisfying the same boundary conditions as $\psi(0)$ on the leads of D. No conditions at all are imposed on the functions ψ along insulating arcs on the boundary. The matrix $\sigma(0)$ may vary as a function of position in each subregion, and may have discontinuities across the boundaries between regions.

Proof: Suppose first that the device is composed of only one region, so that the matrix $\sigma(0)$ has no discontinuities. Let ϕ satisfy the voltage conditions on D, and let $g=\varphi-\psi(0)$, so that $\varphi=\psi(0)+g$. Defining the operator

$$E(\phi_1,\phi_2)=\int_D (\sigma(0)\nabla\phi_1)\cdot(\nabla\phi_2),$$

it follows that

$$\int_D (\sigma(0)\nabla\phi)\cdot\nabla\phi = E(\psi(0),\psi(0))+E(g,g) + 2E[g,\psi(0)].$$

Since the symmetric matrix $\sigma(0)$ must have positive eigenvalues, it is not difficult to show that, for any function ψ,

$$E(\psi,\psi)\geqslant 0$$

and, to complete the proof, it suffices to show that the last term on the right side of the above expression vanishes. From Eq. (A4) and the Onsager relation, it follows that

$$\nabla\cdot(g\sigma(0)\nabla\psi(0))=\sigma(0)\nabla g\cdot\nabla\psi(0),$$

and the divergence theorem implies that

$$\int_D \sigma(0)\nabla g\cdot\nabla\psi(0)=\int_{\partial D} g\sigma(0)\nabla\psi(0)\cdot\mathbf{n}, \qquad \text{(A5)}$$

where ∂D is the boundary of D, and $\mathbf{n}$ is the outward normal vector field along the boundary. By assumption, $g=0$ on each lead where voltage is specified, and, on leads with the zero current condition, g is still constant, so that

$$\int g\sigma(0)\nabla\psi(0)\cdot\mathbf{n}=g\int \mathbf{J}(0)\cdot\mathbf{n}=0,$$

where the integral is taken over the entire lead. Furthermore, on insulating parts of the boundary, $\sigma(0)\nabla\psi(0)\cdot\mathbf{n}=0$. It then follows that the integral in Eq. (A5) must vanish, and the lemma is proven in the case where $\sigma(0)$ has no discontinuities.

In cases where different regions have different conductivity tensors, the divergence theorem will contain boundary terms along the interfaces between regions of the form

$$\int g\sigma(0)\nabla\psi(0)\cdot\mathbf{n}=\int g\mathbf{J}(0)\cdot\mathbf{n}.$$

Both ψ and ϕ, and hence g, will be continuous across the interface, and so will the current term $\mathbf{J}(0)\cdot\mathbf{n}$, so that these interface terms will all cancel, and the lemma still holds.

Theorem 2: Let D be any planar device region with a finite number of holes (possibly none), containing any number of regions with zero resistance and leads, and subregions with different isotropic conductivity tensors, which may vary as a function of position, and define boundary conditions by specifying either a fixed voltage or zero current conditions on the leads and regions with zero resistance. Let

$$\sigma_1(B)=\min\left\{\frac{\sigma_{xx}(B)}{\sigma(0)}\right\},$$

where the minimization is over the conductivity tensors in the different regions of the device. Then

$$\mathbf{V}^T\mathbf{I}(B)\geqslant\sigma_1(B)\mathbf{V}^T\mathbf{I}(0).$$

If only one constant conductivity tensor appears in the device, and the boundary of the device contains only leads and no insulating components, then the inequality becomes an equality. This conclusion holds when there are regions with zero resistance inside the device.

Proof:

$$\mathbf{V}^T\mathbf{I}(B)=\int_{\partial D}\psi(B)\mathbf{J}(B)\cdot\mathbf{n}=\int_D \nabla\psi(B)\cdot\mathbf{J}(B) = \int_D r\cos(\alpha)\|\nabla\psi(B)\|^2 \qquad \text{(A6)}$$

where $\sigma(B)=re^{i\alpha}$. Noting that $r\cos(\alpha)=\sigma_{xx}(B)$, lemma 2 now implies that

$$\mathbf{V}^T\mathbf{I}(B)=\int_D \sigma_{xx}(B)\|\nabla\psi(B)\|^2 \geqslant\sigma_1(B)\int_D(\sigma(0)\nabla\psi(B))\cdot\nabla\psi(B) \geqslant\sigma_1(B)\int_D\sigma(0)\|\nabla\psi(0)\|^2=\sigma_1(B)\mathbf{V}^T\mathbf{I}(\mathbf{0}).$$

This demonstrates the inequality.

Suppose now that the same conductivity tensor applies throughout the device. We will show that, when the boundary contains no insulating arcs, the inequality becomes an equality:

$$\mathbf{V}^T\mathbf{I}(B)=\int_{\partial D}\psi(0)\mathbf{J}(B)\cdot\mathbf{n}=\int_D\nabla\psi(0)\cdot\mathbf{J}(B) =\int_D\nabla\psi(B)\cdot(\sigma(-B)\nabla\psi(0))$$

$$= \int_{\partial D} \psi(B)(\sigma(-B)\nabla\psi(0))\cdot\mathbf{n} \quad \text{using lemma 1}$$

$$= \int_{D_1} \psi(B)(\sigma(-B)\nabla\psi(0))\cdot\mathbf{n} + \int_{D_2} \psi(B)(\sigma(-B)\nabla\psi(0))\cdot\mathbf{n},$$

where ∂D has been broken up into two different components, the leads D_1 and the insulating parts D_2. On the leads of the device $\nabla\psi(0)$ points in the same direction as $\mathbf{n}$, so that one has

$$\sigma(-B)\nabla\psi(0)\cdot\mathbf{n} = r\cos(\alpha)\nabla\psi(0)\cdot\mathbf{n} = \sigma(B)\nabla\psi(0)\cdot\mathbf{n}$$
$$= \sigma_{xx}(B)\nabla\psi(0)\cdot\mathbf{n} = \sigma_1(B)\sigma(0)\nabla\psi(0)\cdot\mathbf{n}$$
$$= \sigma_1(B)\mathbf{J}(0)\cdot\mathbf{n},$$

implying that

$$\int_{D_1} \psi(B)(\sigma(-B)\nabla\psi(0))\cdot\mathbf{n} = \sigma_1(B)\mathbf{V}^T\mathbf{I}(0).$$

It now follows that, if there are no insulating boundary components, the integral over D_2 vanishes and the inequality of the theorem becomes an equality. This completes the proof of theorem 2.

Lemma 3: Let D be any planar device region, containing regions with zero resistance and leads, and assume a constant isotropic conductivity tensor throughout D. Define boundary conditions by specifying fixed total currents $\mathbf{I}$ through each of the leads of the device. Then

$$\int_D \rho(0)\|\mathbf{J}(B)\|^2 \geq \int_D \rho(0)\|\mathbf{J}(0)\|^2,$$

where the boundary conditions on the leads are held constant and only the magnetic field is allowed to vary.

Proof. Because the resistivity is isotropic, the Onsager relation implies that $\rho(0)$ reduces to scalar multiplication. The factor $\rho(0)$ appearing in the integrals is constant and can be ignored. To prove the lemma, we must define a function

$$g^*(x,y) = \int_{(x_0,y_0)}^{(x,y)} (\mathbf{J}(B)-\mathbf{J}(0))\cdot\mathbf{n},$$

where the integral is taken over any path in D between the fixed base point (x_0,y_0) and the point (x,y). To make sense of this definition, one must show that the integral is independent of which path in D we choose. This will follow if we show that the integral over any closed path, i.e., with the same starting and end points, must be zero. A rigorous method for doing calculations like this, using a homology basis, can be found in Ref. 20, Chap. 4. The main ideas are given as follows.

In general, the closed paths in D fall into two different classes, as depicted in Fig. 7. Either (i) the closed path is the boundary of some subregion C in D (shaded in Fig. 7), or (ii) some subregion C in D is bounded by the union of the given closed path with the boundaries of some of the holes in the region D. Note that the boundary of a hole in D is also a closed path, and belongs to the type (ii) paths, where the subregion C is infinitely thin and is mapped entirely onto the boundary.

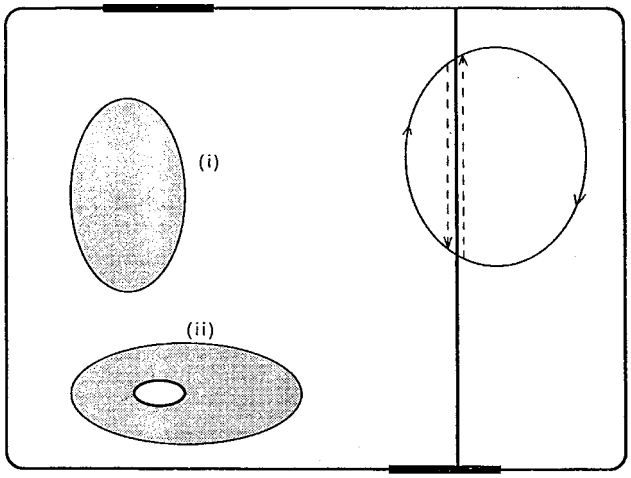

FIG. 7. The different classes of closed loops in a device region. Either the loop bounds a region, shaded as in (i), or the boundary of the region is given as the union of the closed loop with some holes in the region, as shown in (ii).

In case (i), since $\nabla\cdot\mathbf{J}(B) = \nabla\cdot\mathbf{J}(0) = 0$, the divergence theorem implies that

$$\int_{\partial C} (\mathbf{J}(B)-\mathbf{J}(0))\cdot\mathbf{n} = \int_C \nabla\cdot(\mathbf{J}(B)-\mathbf{J}(0)) = 0,$$

which is the desired conclusion.

In case (ii) the divergence theorem again shows that

$$\int_{\partial C} (\mathbf{J}(B)-\mathbf{J}(0))\cdot\mathbf{n} = 0,$$

so the integral over the given closed curve must vanish once we show that the integral over the boundary of every hole in the region D vanishes. The boundary of any such hole consists of some set of leads C_1 joined by some set of insulating arcs, C_2. Along the insulating arcs, both $\mathbf{J}(B)\cdot\mathbf{n}$ and $\mathbf{J}(0)\cdot\mathbf{n}$ are zero, so the integral over C_2 vanishes. Since

$$\int_{C_1} \mathbf{J}(B)\cdot\mathbf{n}$$

is just the total current flowing through the leads in C_1, and this total current is, by assumption, the same for $\mathbf{J}(B)$ and $\mathbf{J}(0)$, the integral over C_1 vanishes also, and we have shown that the function g^* is well defined.

We note that

$$g^*(x,y) = \int_{(x_0,y_0)}^{(x,y)} \nabla g^*\cdot dx = \int_{(x_0,y_0)}^{(x,y)} (\mathbf{J}(B)-\mathbf{J}(0))\cdot\mathbf{n},$$

where $\mathbf{n} = -i\,dx$, i.e., the outward normal is just the unit tangent vector to the curve, rotated clockwise by 90°. From this it follows that

$$\nabla g^* = i(\mathbf{J}(B) - \mathbf{J}(0)).$$

Returning to the proof of the lemma, we write

$$\mathbf{J}(B) = \mathbf{J}(0) + (\mathbf{J}(B) - \mathbf{J}(0)),$$

so that

$$\int_D \|\mathbf{J}(B)\|^2 = \int_D [\|\mathbf{J}(0)\|^2 + \|\mathbf{J}(B) - \mathbf{J}(0)\|^2 + 2(\mathbf{J}(B) - \mathbf{J}(0)) \cdot \mathbf{J}(0)],$$

so we must show that

$$\int_D (\mathbf{J}(B) - \mathbf{J}(0)) \cdot \mathbf{J}(0) = \int_D [i(\mathbf{J}(B) - \mathbf{J}(0))] \cdot [i\mathbf{J}(0)] = 0. \tag{A7}$$

But, again, since $\nabla \cdot (i\mathbf{J}(0)) = 0$, one finds that

$$\nabla \cdot [g^*(i\mathbf{J}(0))] = \nabla g^* \cdot (i\mathbf{J}(0)),$$

and the divergence theorem shows that

$$\int_D [i(\mathbf{J}(B) - \mathbf{J}(0))] \cdot [i\mathbf{J}(0)] = \int_{\partial D} g^*(i\mathbf{J}(0)) \cdot \mathbf{n} = \int_{\partial D} g^* \mathbf{J}(0) \cdot (-i\mathbf{n}) = \int_{\partial D} g^* \mathbf{J}(0) \cdot (-dx).$$

Let C be one of the connected components of ∂D, i.e., C is either the boundary of one of the holes in D or C is the outside boundary of D. Again, we can write C as the union of C_1 and C_2, where C_1 are the leads and C_2 are the insulating segments which join them. The product $\mathbf{J}(0) \cdot dx$ will be zero on C_1, whereas on C_2

$$\mathbf{J}(0) \cdot dx = \sigma(0) \nabla \psi(0) \cdot dx = \sigma(0) d\psi(0),$$

so that

$$\int_C g^* \mathbf{J}(0) \cdot (-dx) = -\int_C g^* \sigma(0) d\psi(0).$$

Since $d\psi(0) = 0$ on C_1, if we show that g^* is constant on C_2, we can pull it and $\sigma(0)$ out of the integral, and since we will be integrating $d\psi(0)$ around a closed loop, the result will be zero, implying that the integral in Eq. (A7) vanishes.

To show that g^* is constant on C_2, note that $\mathbf{J}(B) \cdot \mathbf{n}$ and $\mathbf{J}(0) \cdot \mathbf{n}$ vanish on C_2, so that the integrand in the definition of g^* vanishes on C_2. Thus g^* must be constant on each connected component of C_2. All that remains is to show that g^* will take the same value on two different connected components of C_2 which are separated by a lead. But the difference in the values of g^* on the end points of this lead is just the difference in currents from $\mathbf{J}(B)$ and $\mathbf{J}(0)$ flowing through the lead, and this difference is zero by assumption. Thus g^* is constant on C_2 and the proof is complete.

Theorem 3: For the same device region as in theorem 2, assume a constant, isotropic conductivity tensor, and define boundary conditions by specifying current conditions on each lead and region with zero resistance of the device. Let

$$\rho_1(B) = \frac{\rho_{xx}(B)}{\rho(0)}.$$

Then

$$\mathbf{V}(B)^T \mathbf{I} \geq \rho_1(B) \mathbf{V}(0)^T \mathbf{I}.$$

Proof: Writing $\rho(B) = re^{i\alpha}$ and $\rho(0) = r_0$, one has

$$\nabla \psi(B) \cdot \mathbf{J}(B) = (\rho(B)\mathbf{J}(B)) \cdot \mathbf{J}(B) = r\cos(\alpha)\|\mathbf{J}(B)\|^2 = \rho_{xx}(B)\|\mathbf{J}(B)\|^2 = \rho_1(B)\rho(0)\|\mathbf{J}(B)\|^2.$$

From lemma 3 it follows that

$$\mathbf{V}(B)^T \mathbf{I} = \rho_1(B) \int_D \rho(0)\|\mathbf{J}(B)\|^2 \geq \rho_1(B) \int_D \rho(0)\|\mathbf{J}(0)\|^2 = \rho_1(B)\mathbf{V}(0)^T \mathbf{I}.$$

This completes the proof of theorem 3.

As already mentioned, it is possible to generalize theorem 2 to situations where the conductivity tensor is not necessarily isotropic. The proof already given does not use the isotropic property until Eq. (A6). At this point, if the conductivity tensor is anisotropic, the factor $r\cos(\alpha)$ will vary according to the direction of the potential gradient. It will, however, always be larger than

$$\min_{\|\mathbf{v}\| = 1, \sigma(B)} \{\sigma(B)\mathbf{v} \cdot \mathbf{v}\},$$

where the minimization is over all unit tangent vectors $\mathbf{v}$ and over all the conductivity tensors in the different regions of the device. By defining

$$\sigma_1(B) = \min_{\|\mathbf{v}\| = 1, \sigma(B)} \left\{ \frac{\sigma(B)\mathbf{v} \cdot \mathbf{v}}{\sigma(0)\mathbf{v} \cdot \mathbf{v}} \right\},$$

the inequality

$$\mathbf{V}^T \mathbf{I}(B) \geq \sigma_1(B) \mathbf{V}^T \mathbf{I}(0)$$

will still hold. It is, however, no longer clear that this inequality ever becomes an equality, and, in particular, it seems likely that device configurations such as the Corbino disk, having no insulating boundary components, may no longer have the best possible response curve. In particular, if a very thin rectangular device with long leads is designed so that current flows mostly in the direction $\mathbf{v}$, in which the minimum in the definition of $\sigma_1(B)$ is attained, the resistance at field strength B should be higher than that of the annular Corbino disk, in which substantial amounts of current will be flowing in directions other than $\mathbf{v}$.

Such generalizations of theorem 3 do not, at this time, seem possible. The difficulty occurs already in lemma 3, whose proof requires that $\nabla \cdot (\rho(0)\mathbf{J}(B)) = 0$. If $\rho(0)$ is anisotropic, or even if $\rho(0)$ varies as a function of position, this may not be the case, and we do not know if lemma 3 remains true in this generality.

APPENDIX B: RESISTANCE MATRIX FOR A THREE-LEAD DEVICE

In this appendix we will outline how to calculate the resistance matrix $\mathbf{R}(B)$ for the three-wire device discussed in Sec. III assuming an isotropic conductivity tensor. The methods are based on conformal mapping techniques, as described in Refs. 10 and 20. To calculate $\mathbf{R}(B)$, it suffices to calculate the currents $\mathbf{I}(B)$ associated to any set of three linearly independent voltage conditions $\mathbf{V}_j$, $j=1$, 2, and 3. We will take $\mathbf{V}_j$ to specify a voltage of 1 on lead j and a voltage of zero on the other two leads. Since the calculations for all three voltage conditions are basically the same, we will just do the calculation for $\mathbf{V}_1$.

The solution when $B=0$ is obtained by constructing a conformal mapping $w(z)$ from the device region pictured in Fig. 4(a), thought of as lying in the z plane, to the region pictured in Fig. 8(a), lying in the w plane. Both the real and the imaginary parts of the function $w(z)$ will be harmonic functions, and we write $w=\psi(z)+i\,\eta(z)$. In Ref. 10 it was shown that $\psi(z)$ is, in fact, the desired potential function when $B=0$. Suppose now that $\sigma(B)=re^{i\alpha}$, for some r and α and that $w(z,B)$ is a conformal mapping from the device region in Fig. 4(a) to the region depicted in Fig. 8(b). Then, writing $w(z,B)=\psi(z,B)+i\,\eta(z,B)$, $\psi(z,B)$ is the potential function when the magnetic field is given by B. It was further shown in Ref. 21 that, if z_0 and z_1 are the end points of any lead in Fig. 4(a), say lead j, then

$$|\eta(z_1,B)-\eta(z_0,B)|=\left|\int \nabla\psi(z,B)\cdot\mathbf{n}\right|,$$

where the integral is taken over lead j. Since $\mathbf{J}=re^{i\alpha}\cdot\nabla\psi$, it follows that the total current through the lead is given by

$$\mathbf{I}_j=d\int \mathbf{J}\cdot\mathbf{n}=\pm d|\eta(z_1,B)-\eta(z_0,B)|r\cos(\alpha),$$

where d is the thickness of the planar device.

The function $w(z,B)$ can be expressed as the composition of two functions, $g(f(z))$, where $f(z)$ is a conformal mapping from the region in Fig. 4(a) to the upper half-plane $\mathbf{H}$, as shown in Fig. 8(c), and $g(\zeta)$ is a conformal mapping from $\mathbf{H}$ to the region in Fig. 8(b). As we shall see, it is not necessary to know the exact form of $f(z)$ in order to calculate the current-voltage relations we are seeking. A specific form for $g(\zeta)$ can be given in terms of definite integrals via the Schwarz-Christoffel formula (see Ref. 20)

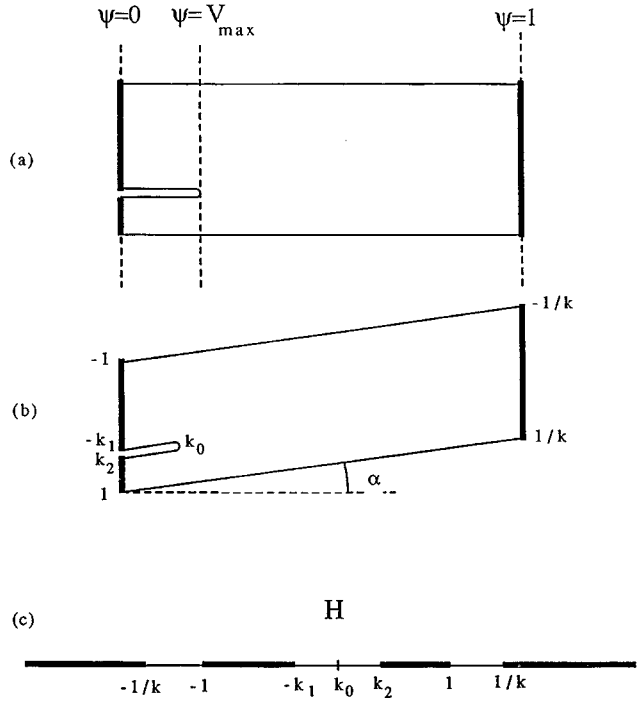

FIG. 8. The three-lead device shown in Fig. 4(a) can be mapped conformably onto the regions shown in (a) and (b). In (a), the real part of the mapping gives the potential when no magnetic field is present. In (b) the real part of the mapping gives the potential when the magnetic field deflects the current through an angle a. Both mappings are obtained by factoring through the upper half-plane H, shown in (c). This reduces the problem of finding the potential to that of finding the parameters $1/k$, $-k_1$, k_0, and k_2.

$$g(\zeta)=\psi(\zeta,\alpha)+i\,\eta(\zeta,\alpha)=Z_0\int_0^{\zeta}\frac{(\zeta-k_0)d\zeta}{[(\zeta+1)(\zeta-1/k)(\zeta-k_2)]^{(1/2+\alpha)}[(\zeta-1)(\zeta+1/k)(\zeta+k_1)]^{(1/2-\alpha)}}.$$

The real line in the ζ plane is mapped to the boundary of the region in Fig. 8(b), and the corners of this region are the images of the points $-1/k$, -1, $-k_1$, k_0, k_2, 1, and $1/k$, as shown in Fig. 8(c). The values of Z_0, k, k_0, k_1, and k_2 must still be determined.

The currents through the three leads in the device are given by

$$I_2(\alpha)=d|\eta(-k_1,\alpha)-\eta(-1,\alpha)|r\cos(\alpha),$$

$$I_3(\alpha)=d|\eta(1,\alpha)-\eta(k_2,\alpha)|r\cos(\alpha),$$

$$I_1(\alpha)=-I_2(\alpha)-I_3(\alpha). \tag{B1}$$

All that remains is to determine the as yet unknown parameters of Z_0, k, k_0, k_1, and k_2. Note that the function $f(z)$ is independent of B, but the function $g(\zeta)$ is not. The values $\pm 1/k$, k_1, and k_2 are the images of the end points of leads under the map $f(z)$. As such, their values will be the same for any value of B. The values Z_0 and k_0, in contrast, will vary with B. In particular, k_0 determines the maximum value of the potential function on the boundary between the leads 2 and 3, and this depends on B. The values for all these parameters can be determined from a set of three physical parameters for the device performance at $B=0$ with the prescribed voltage conditions, namely, the currents I_2 and I_3, and the maximum value $V_{\max}$ of the potential between leads

TABLE I. Room-temperature electron and hole densities and mobilities for InSb, InAs, and Bi, as used for calculating the results in Fig. 6. For Bi, the quantity in the row μ_n is μ_{n1}, and that in the row μ_p it is μ_{p1}.

	Units	InSb	InAs	Bi
n	cm^{-3}	5.21×10^{16} (Ref. 17)	3.29×10^{16} (Ref. 17)	2.45×10^{18} (Ref. 18)
μ_n	$\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$	56 100 (Ref. 17)	22 000 (Ref. 17)	32 000 (Ref. 18)
p	cm^{-3}	2.54×10^{15} (Ref. 17)	n.a.	2.45×10^{18} (Ref. 18)
μ_p	$\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$	8000 (Ref. 22)	n.a.	6000 (Ref. 18)

2 and 3. Noting that the Onsager relations imply that $\sigma(0)$ has no imaginary part, one then has the equations

$$I_2(0)=d\,|\,\eta(-k_1,0)-\eta(-1,0)\,|\sigma(0),$$

$$I_3(0)=d\,|\,\eta(1,0)-\eta(-k_2,0)\,|\sigma(0),$$

and

$$V_{\max}=|\,\psi(k_0,0)-\psi(-k_1,0)\,|.$$

In addition, there are the voltage conditions

$$\psi(-1/k,0)-\psi(-1,0)=V_1=1,$$

$$\psi(k_2,0)-\psi(-k_1,0)=V_2-V_3=0.$$

These five equations can be used to determine the values for of Z_0, k, k_0, k_1, and k_2 at $B=0$. To determine the values Z_0 and k_0 at some other value of B, the potential equations

$$\psi(-1/k,\alpha)-\psi(-1,\alpha)=V_1=1,$$

$$\psi(k_2,\alpha)-\psi(-k_1,\alpha)=V_2-V_3=0$$

are again used.

The integrals in the definition of $g(\zeta)$ were numerically evaluated. Using Eq. (B1), the currents can then be evaluated. The scale factor $r\cos(\alpha)$ is material dependent, but, for materials such as InSb or InAs, where the current is carried primarily by one type of charge carrier (see Appendix C), r can be expressed as a function of α. Equation (C1) from Appendix C implies that

$$\|\mathbf{J}\|^2=\sigma(0)\mathbf{J}\cdot\varepsilon=\sigma(0)\mathbf{J}\cdot\boldsymbol{\nabla}\psi=\|\mathbf{J}\|\|\boldsymbol{\nabla}\psi\|\sigma(0)\cos(\alpha)$$

where ε is the electric field, so that

$$r=\sigma(0)\cos(\alpha).$$

Since the currents have been determined for unit voltage at lead 1, the quantities in equations (B1) carry the units of current per volt, determined as a function of the angle α. The scale factors d (the thickness of the device) and $\sigma(0)$ (the real-valued conductivity when no magnetic field is present) multiply together to have precisely these units. Dividing Eq. (B1) by these two scale factors, one arrives at the dimensionless quantity which is plotted in Figs. 4(b) and 4(c). The values $k=0.5$, $k_1=-0.1$, and $k_2=0.9$ were used to do the calculations.

APPENDIX C: CONDUCTIVITY TENSORS FOR InSb, InAs, AND Bi

The narrow-gap III-V compounds that are most useful in galvanomagnetic devices are InSb and InAs, because of their high electron mobilities. Since we consider n-type doped material, the electron properties matter most. The electron Fermi surfaces are spheres centered around the Γ point (the center) of the Brillouin zone, and dispersion relation is characterized by one isotropic effective mass m^*. It is reasonable to assume that the electron relaxation time and hence the electron mobility (μ) are also isotropic. The conductivity in zero field is

$$\sigma(0)=nq\mu,$$

where n is the density of the carrier and q is the electron charge. In the presence of a magnetic field **B** perpendicular to the x-y plane and of magnitude B, the current density **J** due to an electric field ε, with no z component, is

$$\mathbf{J}=\sigma(0)\left[\varepsilon+\frac{\mathbf{J}\times\mathbf{B}}{qn}\right]. \tag{C1}$$

If we compare this equation to that defining the magnetoconductivity tensor, namely,

$$\mathbf{J}=\sigma(B)\varepsilon,$$

where

$$\sigma(B)=\begin{bmatrix}\sigma_{xx} & -\sigma_{xy}\\ \sigma_{xy} & \sigma_{xx}\end{bmatrix},$$

then

$$\sigma_{xx}=\frac{qn\mu}{1+\mu^2B^2},$$

$$\sigma_{xy}=\pm\mu B\sigma_{xx},$$

where the sign choice in σ_{xy} is negative when the charge carriers are electrons, and positive when the charge carriers are holes.

In the following example, we consider the InSb and InAs thin-film samples described by Kataoka,[17] rather than bulk InSb and InAs which have much higher mobilities, because thin-film material is more likely to be used in planar galvanomagnetic devices. Their electron density n and mobility μ at 300 K have been estimated from Fig. 17 of Ref. 17 using the low-field one-carrier approximations $n=1/R_Hq$ and $\mu=\sigma(0)/nq$. Values for n and μ are given in Table I.

For InSb at 300 K the contributions of the light and heavy holes must be estimated. The holes also occupy Fermi surfaces centered at the Γ point of the Brillouin zone. Since they are minority carriers in the material we consider, and since their densities and mobilities are low, we treat the hole transport properties as isotropic. The relevant densities of state can be calculated from the effective masses.[21] Knowing the electron density, one can estimate the location of the Fermi level, the hole Fermi energy, and the densities of light and heavy holes, p_1 and p_h, respectively. The density of light holes is negligible; the density of heavy holes is given in Table I. The estimated hole mobility, also given in Table I, is taken from Ref. 22.

The Fermi surfaces of the electrons and holes in bismuth are more complicated[23] because of the low symmetry of the material. We use the trigonal axis system notation, with index 1 along the binary axis, index 2 along the bisectrix, and index 3 along the trigonal axis. Holes have ellipsoidal Fermi surfaces centered at the *T* point of the Brillouin zone (the intercept of the zone with the trigonal axis). These are ellipsoids of revolution around the trigonal axis. The hole density is labeled *p*. Electrons fill three ellipsoids centered at the *L* points, rotated 120° from each other. In the trigonal axis system, we can describe one of these ellipsoids as follows: the shortest axis of the ellipsoids is along the binary (1) direction. The longest axis of the electron Fermi surface is tilted by angle $\varphi_e = 7°$ away from the bisectrix (2) toward the trigonal (3) axis. The total electron density, summed over all three ellipsoids, is labeled *n*. The mobility tensors of electrons μ_n and holes μ_p have the same symmetry as the effective mass tensors:

$$\mu_{\mathbf{n}} = \begin{bmatrix} \mu_{n1} & 0 & 0 \\ 0 & \mu_{n2} & \mu_{n4} \\ 0 & \mu_{n4} & \mu_{n3} \end{bmatrix}$$

and

$$\mu_{\mathbf{p}} = \begin{bmatrix} \mu_{p1} & 0 & 0 \\ 0 & \mu_{p1} & 0 \\ 0 & 0 & \mu_{p3} \end{bmatrix}.$$

Values for the tensor components are given, between 4.2 and 17 K, in Ref. 24, and from 77 to 300 K values for all components but μ_{p3} are reported in Ref. 18.

The calculation of the conductivity tensor elements is done in the same way as for the III-V compounds. We limit our example to the case where the magnetic field is aligned along the trigonal axis. The magnetoconductivity tensor elements for the single hole ellipsoid, which is isotropic in the 1–2 plane, then become[6]

$$\sigma_{xx}^{p} = \frac{qp\mu_{p1}}{1+\mu_{p1}^{2}B^{2}},$$

$$\sigma_{xy}^{p} = \pm \mu_{p1} B \sigma_{xx}^{p}.$$

The anisotropic electron ellipses have to be treated each separately. As there is threefold symmetry along the trigonal aids, there are three possible choices of coordinate sets, rotated by 120°. The conductivity tensor elements for all three electron pockets then can be added to give[6]

$$\sigma_{xx}^{n} = \frac{qn(\mu_{n1}+\mu_{n2})}{2(1+\mu_{n1}\mu_{n2}B^{2})},$$

$$\sigma_{xy}^{n} = \frac{-qn\mu_{n1}\mu_{n2}B}{1+\mu_{n1}\mu_{n2}B^{2}}.$$

Assuming a ratio $\mu_{n1}/\mu_{n2} = 50$,[24]

$$\sigma_{xx}^{n} = \frac{0.51qn\mu_{n1}}{1+0.02\mu_{n1}^{2}B^{2}},$$

$$\sigma_{xy}^{n} = \frac{-0.02qn\mu_{n1}^{2}B}{1+0.02\mu_{n1}B^{2}}.$$

The values for *n*, μ_{n1}, *p*, and μ_{p1} that are used in the conductivity tensors above, and in Sec. IV and Fig. 6 were taken from Ref. 18 and are summarized in Table I. The total conductivity tensor is now given as

$$\sigma = \sigma^{n} + \sigma^{p}.$$

[1] H. Weiss, *Structure and Applications of Galvanomagnetic Devices* (Pergamon, New York, 1969).

[2] J. Heremans, J. Phys. D **26**, 1149 (1993).

[3] T. Thio and S. Solin, Appl. Phys. Lett. **72**, 3497 (1998).

[4] T. Thio, S. A. Solin, J. W. Bennet, D. R. Hines, M. Kawano, N. Oda, and M. Sano, Phys. Rev. B **57**, 12 239 (1998).

[5] T. G. Digges and R. N. Tauber, Metall. Trans. **4**, 1169 (1973).

[6] J. M. Noothoven van Goor, Philips Res. Rep., Suppl. **4**, 1 (1971).

[7] R. Xu, A. Husmann, T. F. Rosenbaum, M.-L. Saboungi, J. E. Enderby, and P. B. Littlewood, Nature (London) **390**, 57 (1997).

[8] C. Herring, J. Appl. Phys. **31**, 1939 (1960).

[9] D. J. Bergman and Y. M. Strelniker, Phys. Rev. B **49**, 16 256 (1994); Y. M. Strelniker and D. J. Bergman, *ibid.* **50**, 14 001 (1994); D. J. Bergman and Y. M. Strelniker, *ibid.* **51**, 13 845 (1995); Y. M. Strelniker and D. J. Bergman, Phys. Rev. B **53**, 1 (1996).

[10] R. F. Wick, J. Appl. Phys. **25**, 741 (1954).

[11] T. Thio, S. A. Solin, J. W. Bennett, D. R. Hines, M. Kawano, N. Oda, and M. Sano, J. Cryst. Growth **184/185**, 1293 (1998).

[12] S. A. Solin, T. Thio, J. W. Bennett, D. R. Hines, M. Kawano, N. Oda, and M. Sano, Appl. Phys. Lett. **69**, 4105 (1996).

[13] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).

[14] J. Haeusler, Z. Naturforsch. A **17a**, 506 (1962).

[15] H. J. Lippmann and F. Kuhrt, Z. Naturforsch. A **13a**, 462 (1958).

[16] J. Heremans, D. T. Morelli, D. L. Partin, C. H. Olk, C. M. Thrush, and T. A. Perry, Phys. Rev. B **38**, 10 280 (1988).

[17] S. Kataoka, Circulars of the Electricotechnical Laboratory #182, Tokyo (1974) (unpublished).
[18] J.-P. Michenaud and J.-P. Issi, J. Phys. C **5**, 3061 (1972).
[19] D. H. Norrie and G. de Vries, *The Finite Element Method* (Academic, New York, 1973).
[20] L. V. Ahlfors, *Complex Analysis* (McGraw-Hill, New York, 1966).
[21] K. Seeger, *Semiconductor Physics* (Springer-Verlag, Berlin, 1985).
[22] *Transport Phenomena*, edited by R. K. Willardson and Albert C. Beer, Semiconductors and Semimetals Vol. 10 (Academic, New York, 1975).
[23] V. S. Edelman, Adv. Phys. **25**, 555 (1976).
[24] R. Hartman, Phys. Rev. **181**, 1070 (1969).