Gauge-Invariant Formulation of Electron Linear Transport

Fernando Sols

Departamento de Física de la Materia Condensada, C-XII, Universidad Autónoma de Madrid, E-28049 Madrid, Spain[a]
and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801
(Received 31 July 1991)

It is shown that the invariance of linear-response formulations under large gauge transformations requires the use of a generalized continuity equation in which current sources at infinity are included. These sources, which are described by delta functions located at the points of infinity, guarantee global charge conservation in models with open boundary conditions. Descriptions of steady-state transport in terms of global conductance coefficients and nonlocal conductivity tensors are shown to be equivalent. The reciprocity theorem for a four-lead resistance is proved in its full generality for interacting electron systems.

PACS numbers: 72.10.Bg, 05.60.+w

The approximation of linear response has been extensively used in the theory of electron transport. After the classic work of Kubo [1], microscopic calculations of conductivity have been routinely performed by theorists and the knowledge gained from them has been instrumental in the development of solid-state physics. In Kubo's approach to linear transport, the current density at one point is a linear and, in general, nonlocal function of the external electric field. In the steady-state regime,

$$J_k(\mathbf{r}) = \int \sigma_{kl}(\mathbf{r},\mathbf{r}')E_l(\mathbf{r}')d\mathbf{r}' , \qquad (1)$$

where $\sigma_{kl}(\mathbf{r},\mathbf{r}')$ is the conductivity tensor. Simultaneous with Kubo's work, an alternative approach to electron linear transport was developed by Landauer [2] in which the resistance is related to the electron scattering. Landauer's work only gained popularity many years later with the advent of the scaling theory of localization [3] and, more recently, with the development of the physics of mesoscopic systems [4]. His work was later refined and generalized by other authors (for two historical accounts, see Refs. [5] and [6]). Linear transport in a multilead structure is characterized by the conductance coefficients g_{mn} relating the total current I_m in a given lead m to the voltages ϕ_n in the asymptotic regions of the various leads attached to the sample,

$$I_m \equiv \int d\mathbf{s}_m J_m(\mathbf{s}_m) = \sum_n g_{mn}\phi_n , \qquad (2)$$

where J_m is the component of the current density parallel to the direction of the lead. Büttiker [7] showed that, in the independent electron approximation and at zero temperature

$$g_{mn} = \frac{2e^2}{h}\left(\sum_{i=1}^{N_m}\sum_{j=1}^{N_n} T_{mn,ij} - N_m\delta_{mn}\right) , \qquad (3)$$

where $T_{mn,ij}$ is the electron transmission probability from channel j of lead n to channel i of lead m.

The relation of the conductance to the electron scattering, as described by the Landauer-Büttiker equations (2) and (3), is usually considered the characteristic feature of the Landauer approach to linear transport. However, in this work we will be more concerned with the fact, expressed by Eq. (2), that the current is uniquely determined by the asymptotic voltages. This property is in apparent contrast with the Kubo formula (1), which gives the current as a nonlocal function of the external electric field. Thus one of the goals of this paper is to establish the validity and equivalence of both formulations. All the arguments given in this paper apply to systems of *electrons with arbitrary interactions*. The expression (3) for the conductance coefficients g_{mn} will apply to the particular case of noninteracting electrons and zero temperature. Finally, symmetry relations of the Onsager type will be derived. We omit in this presentation the discussion of questions related to Van Kampen's objection to linear response and to the exchangeability of the concepts of electrostatic potential and chemical potential. A more detailed version of this work will be given elsewhere [8].

The question of the equivalence between the Kubo and Landauer formulations of electron linear transport has already been considered by several authors within the framework of the independent electron approximation [5-11]. However, these valuable works are of limited generality [9,10] (assumption of a constant electric field) or contain inconsistencies in the treatment of bulk and boundary terms [5,8,11]. We will see that the study of this problem requires an adequate treatment of the delicate interplay between surface and bulk contributions in systems with open boundary conditions. This question is related to the more fundamental problem of the invariance of electromagnetic linear response under *large gauge transformations*, defined as those in which the function of the gauge transformation $\Lambda(\mathbf{r})$ does not vanish at infinity [12]. The invariance under *small* gauge transformations [i.e., those in which $\lim_{r\to\infty}\Lambda(\mathbf{r})=0$] was already proved by Kadanoff and Martin [13]. The extension of the proof to the case of large gauge transformations requires a generalization of the continuity equation that explicitly includes current sources at infinity [see Eq. (12)]. This generalization is not an *ad hoc* solution to the problem, but a rigorous consequence of the operator properties in systems with open boundary conditions.

To state the problem, we begin by considering a physical system with arbitrary many-body interactions that is subject to a time-dependent perturbation

$$V(t)=\int \rho(\mathbf{r})\phi(\mathbf{r},t)d\mathbf{r}\,, \tag{4}$$

where $\rho(\mathbf{r})$ is the electron density operator. In first-order perturbation theory the expectation value of the induced current is

$$J_k(\mathbf{r},t)=\frac{1}{i\hbar}\int_{-\infty}^{t}dt'\int d\mathbf{r}'\langle[j_k(\mathbf{r},t),\rho(\mathbf{r}',t')]\rangle\phi(\mathbf{r}',t')=-\beta\int_{-\infty}^{t}dt'\int d\mathbf{r}'\langle j_k(\mathbf{r},t);\dot\rho(\mathbf{r}',t')\rangle\phi(\mathbf{r}',t') \tag{5}$$

($\beta=1/kT$), where $j_k(\mathbf{r},t)$ is the electron current operator in the interaction picture and the canonical correlation

$$\langle a;b\rangle=\beta^{-1}\int_0^{\beta}\langle a(-i\hbar\lambda)b(0)\rangle d\lambda \tag{6}$$

has been introduced [14]. By invoking the continuity equation, $\dot\rho(\mathbf{r}',t')$ in Eq. (5) can be replaced by $-\nabla_l' j_l(\mathbf{r}',t')$ and, after integration by parts, one obtains

$$J_k(\mathbf{r},t)=\beta\int_{-\infty}^{t}dt'\int d\mathbf{r}'\langle j_k(\mathbf{r},t);j_l(\mathbf{r}',t')\rangle[-\nabla_l'\phi(\mathbf{r}',t')]+\beta\int_{-\infty}^{t}dt'\int d\mathbf{s}'\langle j_k(\mathbf{r},t);j_n(\mathbf{s}',t')\rangle\phi(\mathbf{s}',t')\,, \tag{7}$$

where the second term contains an integration over the surface of infinity, i.e., a surface that is far enough from the sample for the physical fields to vanish [15], and $j_n(\mathbf{s}',t')$ is the component of the current normal to that surface. For a mesoscopic structure in which leads are attached to the sample, the surface of infinity is formed by a finite number of disconnected pieces and the second term of Eq. (7) becomes

$$\beta\int_{-\infty}^{t}dt'\sum_n\phi_n(t')\int ds_n'\langle j_k(\mathbf{r},t);j_n(\mathbf{s}_n',t')\rangle\,, \tag{8}$$

where $\phi_n(t)$ is the asymptotic voltage at lead n. These surface terms were essentially neglected in Kubo's original work [1].

The remarkable fact about Eq. (7) is that, while its derivation was motivated by an attempt to prove the equivalence between a bulk and a boundary formulation of linear electron transport, it seems to contain both types of contribution. While the first term of Eq. (7) can be easily identified with the Kubo formula, the second term seems to yield a Landauer-type contribution (actually, with an opposite sign) which depends on the voltages in the leads. If Eq. (7) were correct, it would imply a major revision of the commonly held views about linear transport. Thus it deserves a very serious scrutiny before it can be definitely accepted. A natural check of this equation consists in trying to rederive it in a different gauge. If, for example, one takes a pure vector potential $\mathbf{A}(\mathbf{r},t)$, as is most often done in derivations of the Kubo formula of conductivity, to describe the electric field, $\mathbf{E}(\mathbf{r},t)=-(1/c)\dot{\mathbf{A}}(\mathbf{r},t)$, one obtains the rather surprising result that the predicted current is different from (7). Thus, the somewhat remarkable consequences of Eq. (7) are automatically held in suspense and we find ourselves in front of a more general and fundamental problem, namely, the gauge invariance of electromagnetic linear response. It is easy to see that the change from a purely scalar potential to a purely vector potential requires a *large* gauge transformation if the electric field is zero everywhere at $t=-\infty$ (the steady-state regime is achieved by introducing the perturbation adiabatically), yields a net potential drop, and is zero at all times in the asymptotic leads.

To study the question of invariance under a large gauge transformation, let us describe the electric field $\mathbf{E}(\mathbf{r},t)$ with an arbitrary gauge [16], $\mathbf{E}(\mathbf{r},t)=-\nabla\phi(\mathbf{r},t)-(1/c)\dot{\mathbf{A}}(\mathbf{r},t)$. In this gauge, the expectation value of the current is written as

$$J_k(\mathbf{r},t)=\frac{1}{i\hbar}\int_{-\infty}^{t}dt'\int d\mathbf{r}'\langle[j_k(\mathbf{r},t),\rho(\mathbf{r}',t')]\rangle\phi(\mathbf{r}',t')$$
$$-\frac{1}{i\hbar c}\int_{-\infty}^{t}dt'\int d\mathbf{r}'\langle[j_k(\mathbf{r},t),j_l(\mathbf{r}',t')]\rangle A_l(\mathbf{r}',t')-\frac{e}{mc}\langle\rho(\mathbf{r})\rangle A_k(\mathbf{r},t)\,. \tag{9}$$

If a gauge transformation is performed, $\phi\to\phi-\dot\Lambda/c$, $\mathbf{A}\to\mathbf{A}+\nabla\Lambda$, one should expect the variation of the current $\Delta J_k(\mathbf{r},t)$ to be zero on physical grounds. However, one finds after some algebra,

$$\Delta J_k(\mathbf{r},t)=\frac{1}{i\hbar c}\int_{-\infty}^{t}dt'\int d\mathbf{r}'\left\langle\left[j_k(\mathbf{r},t),\left\{\nabla_l' j_l(\mathbf{r}',t')+\frac{\partial\rho(\mathbf{r}',t')}{\partial t'}\right\}\right]\right\rangle\Lambda(\mathbf{r}',t')$$
$$-\frac{1}{i\hbar c}\int_{-\infty}^{t}dt'\int d\mathbf{s}'\langle[j_k(\mathbf{r},t),j_n(\mathbf{s}',t')]\rangle\Lambda(\mathbf{s}',t')\,, \tag{10}$$

where the operator identity

$$[j_k(\mathbf{r}),\rho(\mathbf{r}')]=(\hbar c/mi)\rho(\mathbf{r})\nabla_k\delta(\mathbf{r}-\mathbf{r}') \tag{11}$$

has been used. In principle, the first term in Eq. (10) vanishes due to charge conservation. In the case of a small gauge transformation, this would complete the proof of gauge invariance, since then $\Lambda(\mathbf{s}',t')=0$ [13]. The case where $\Lambda(\mathbf{s},t)$ takes a uniform value at the surface of infinity can also be dealt with trivially by invoking charge conservation and Eq. (11). However, in the general case of a large gauge transformation in which $\Lambda(\mathbf{s},t)$ is not uniform, there is no clear way of removing the second term of Eq. (10) and thus of proving gauge invariance. Since the predictions of the theory have to be gauge invariant, there must be a flaw either in the derivation of Eq. (10) or in its interpretation. The algebra leading to this equation is quite direct; thus we investigate below the second possibility and show that a satisfactory solution to the problem can be found.

First we note that gauge invariance would be proven by Eq. (10) if we were able to replace the standard continuity equation by a modified version of it:

$$\nabla\cdot\mathbf{j}(\mathbf{r},t)+\frac{\partial\rho(\mathbf{r},t)}{\partial t}=j_n(\mathbf{s})\delta(r_n-s_n)\,, \qquad (12)$$

where r_n (s_n) is the component of $\mathbf{r}$ ($\mathbf{s}$) along the direction locally normal to the surface of infinity, defined here by the set of points $\mathbf{s}$. Essentially, this is a delta function located at the surface of infinity and, as a distribution, it can be defined in more intrinsic terms than suggested by the specific representation given in Eq. (12) [8]. This distribution assigns to a given function its asymptotic value at the various points of infinity. Physically, the right-hand side (rhs) of Eq. (12) represents a current source at infinity. It is natural to ask whether one can derive this generalized continuity equation (GCE) (12) from rigorous mathematical considerations. The surprising answer is yes. One can compute the time derivative of the density operator from the relation

$$\dot\rho=(1/i\hbar)[\rho,H_0]\,, \qquad (13)$$

where

$$H_0=\frac{1}{2m}\int d\mathbf{r}\,\psi^\dagger(\mathbf{r})\left(-i\hbar\nabla-\frac{e}{c}\mathbf{A}_0(\mathbf{r})\right)^2\psi(\mathbf{r}) \qquad (14)$$

is the noninteracting part of the unperturbed Hamiltonian (the interacting part trivially commutes with ρ), which includes the possibility of a background magnetic field $\mathbf{B}=\nabla\times\mathbf{A}_0$. As is well known, application of the equal-time commutation rules leads to the standard continuity equation. However, if, in the integration by parts, one keeps surface terms that are usually neglected, what one obtains is exactly (12). Thus, the GCE is a direct consequence of the operator properties. It is important to note, however, that a modification in the usual definition of the derivative of the delta function has to be introduced when this is acting on functions that do not decay at infinity, but tend to a finite constant. Surface terms in the form of a delta function located at the points of infinity have to be included if mathematical inconsistencies are to be avoided, and the gauge invariance studied here is just one particular example [8]. This type of surface term is usually removed by assuming that the test functions have a compact support or decay at infinity. However, many functions or operators in physics, like the current distributions considered here, do not satisfy any of these conditions.

The validity of Eq. (12) can also be supported from physical considerations. The crucial feature is that *in models with open boundary conditions, local conservation of charge does not guarantee global conservation of charge.* This idea can be illustrated with a simple example. Consider an infinite electron gas in which an attractive impurity is added. Beyond a sufficiently large distance from the impurity, nothing has changed, due to total screening. However, in the region close to the impurity, an extra electron has been added. Thus, by introducing the impurity, the system has passed from having "infinite" electrons to having "infinite-plus-one" electrons. However, charge cannot be created if no term in the Hamiltonian allows for such a process. The source term in the GCE precisely guarantees global charge conservation. It tells us that whatever charge is added to the system, it is explicitly taken from the points of infinity. If one integrates the GCE over a compactified space that includes the surface of infinity, the first term in the left-hand side of (12) cancels with the source term and, as a result, global charge conservation is preserved.

Before we proceed further, let us discuss in what situations the source term of the GCE is *not* needed. The rhs of Eq. (12) is irrelevant when, as is usually done in field-theoretical contexts, matter and gauge fields are assumed to vanish far at infinity (this case would also include closed systems), or when periodic boundary conditions are adopted, as has traditionally been done in solid-state physics, since then the same physical current j_n appears on both sides of the boundary, but with opposite sign, and the two contributions cancel. The GCE is also not needed when the external electric field involves no net potential drop. However, mesoscopic multilead structures constitute a clear example of physical systems in which none of the boundary conditions mentioned above can be applied. Periodic boundary conditions are not desirable since they would introduce unwanted multiple scattering. On the other hand, current distributions at infinity cannot be assumed to vanish in the presence of a potential bias, and, furthermore, these are precisely the quantities we want to compute. Finally, we note that the GCE is an identity in the sense of distributions. Thus, the source term can be omitted when one is only interested in relations between local physical quantities.

It might be argued that, after all, infinite open systems are fictitious entities and that all physical systems are ultimately finite and thus the GCE is unnecessary. However, physical models with open boundary conditions are convenient idealizations which are most often adopted in condensed-matter physics (and, more generally, in stationary scattering studies). The GCE is simply a requirement of mathematical consistency that is essential in some types of calculation. For example, let us introduce the GCE in Eq. (5). The source term of the GCE cancels with the surface term of Eq. (7) and the expression for the current becomes

$$J_k(\mathbf{r},t)=\beta\int_{-\infty}^{t} dt'\int d\mathbf{r}'\langle j_k(\mathbf{r},t);j_l(\mathbf{r}',t')\rangle E_l(\mathbf{r}',t')\,, \quad (15)$$

which is a manifestly gauge-invariant result. Thus the Kubo formula is correct even when the boundary terms are carefully taken into account.

The question remains of whether Eq. (15) can adopt the form (2) in the steady-state regime with an expression for g_{mn} that would eventually lead to Eq. (3) in the case of independent electrons. There is a powerful argument to prove the formal equivalence between Eqs. (2) and (15). If they were not equivalent, then the total current in one lead would not be a unique function of the voltages in the leads but would depend on the details of the electric field, as seems to be suggested by Eq. (15). Then it would be possible to find two different external electric fields yielding the same asymptotic voltages that, however, would give rise to different currents in the leads. But in linear response it would be possible to obtain a third solution by subtracting the previous two. Such a solution would involve a nonzero static electric field with zero voltage in all leads yielding, however, a nonzero current in the leads, something which is clearly unacceptable on physical and mathematical grounds [8]. Thus, in the static limit, Eq. (15) has to reduce to an equation like (2). Note that this argument applies to the current density at any point of space if $\mathbf{B}=0$, but only to the total current in the asymptotic leads if $\mathbf{B}\neq 0$ [8]. To find a specific expression for g_{mn}, we can exploit the independence of I_m on the details of the potential profile and choose a $\phi(\mathbf{r})$ made of step functions in the leads of height $\phi_n-\phi_0$, where ϕ_0 is a reference potential that drops from the calculation due to charge balance. The electric field is then a sum of delta functions and the spatial integral in (15) can be trivially performed. The result is

$$g_{mn}=i\hbar\sum_{\alpha,\beta}\frac{P_\beta-P_\alpha}{\varepsilon_\beta-\varepsilon_\alpha}\frac{[\hat{I}_m]_{\beta\alpha}[\hat{I}_n]_{\alpha\beta}}{\varepsilon_\beta-\varepsilon_\alpha+i0^+}\,, \quad (16)$$

where the matrix elements $[\hat{I}_n]_{\alpha\beta}$ of the total current operators are taken between many-body states $|\alpha\rangle$ with energy ε_α whose probability distribution P_α is that of a grand canonical ensemble for the incoming channels of the many-body scattering states [8]. The independent electron limit is formally equivalent to (16) with $|\alpha\rangle$ representing a one-electron state and P_α the Fermi-Dirac distribution. Such an expression has been shown in Ref. [11] to be equivalent to Eq. (3).

From (16) it is easy to derive the Onsager relations

$$g_{mn}(\mathbf{B})=g_{nm}(-\mathbf{B})\,, \quad (17)$$

by expressing $g_{nm}(-\mathbf{B})$ as a sum over the time-reversed versions of the states appearing in (16), and this result can be shown to lead to the reciprocity theorem [8] (see Appendix B of Ref. [5]),

$$\mathcal{R}_{mn,kl}(\mathbf{B})=\mathcal{R}_{kl,mn}(-\mathbf{B})\,, \quad (18)$$

where $\mathcal{R}_{mn,kl}\equiv(\phi_k-\phi_l)/I_m$ is the four-lead resistance ($I_n=-I_m$). Büttiker [7] has derived Eqs. (17) and (18) from the symmetry properties of the scattering probabilities contained in Eq. (2), which applies to noninteracting electrons and zero temperature. Here, the symmetry relations (17) and (18) have been proved for finite temperature systems with arbitrary interactions.

I wish to thank Eduardo Fradkin for valuable discussions. This work has been supported by the National Science Foundation (Washington, DC) through Grant No. DMR-86-12860, the U.S. Army Research Office (Durham, NC) through Grant No. DAAL-03-89-K-0037, and by the U.S. Office of Naval Research (Washington, DC) through Grant No. N00014-89-J-1470. I also wish to acknowledge support from Spain's Ministerio de Educación y Ciencia.

[a]Permanent address.

[1] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
[2] R. Landauer, IBM J. Res. Dev. **1**, 233 (1957).
[3] P. W. Anderson, D. J. Thouless, E. Abrahams, and D. S. Fisher, Phys. Rev. B **22**, 3519 (1980).
[4] For a review, see Y. Imry, in *Directions in Condensed Matter,* edited by G. Grinstein and E. Mazenko (World Scientific, Singapore, 1986); C. W. J. Beenakker and H. van Houten, in *Solid State Physics,* edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1991), Vol. 44.
[5] A. D. Stone and A. Szafer, IBM J. Res. Dev. **32**, 384 (1988).
[6] R. Landauer, J. Phys. Condens. Matter **1**, 8099 (1989).
[7] M. Büttiker, IBM J. Res. Dev. **32**, 317 (1988).
[8] F. Sols (to be published).
[9] D. S. Fisher and P. A. Lee, Phys. Rev. B **23**, 6851 (1981).
[10] J. Kucera and P. Streda, J. Phys. C **21**, 4357 (1988).
[11] H. U. Baranger and A. D. Stone, Phys. Rev. B **40**, 8169 (1989).
[12] R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).
[13] L. P. Kadanoff and P. C. Martin, Phys. Rev. **124**, 670 (1961).
[14] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II* (Springer, Berlin, 1985).
[15] Physical fields can be assumed to identically vanish in multilead structures where the surface of infinity is formed by a finite number of disconnected pieces. If the voltage is nonuniform in a simply connected surface of infinity, slowly decaying tangential fields must exist which, however, do not invalidate the arguments given in this paper.
[16] We assume the external electromagnetic fields to be *c* numbers without internal dynamics and neglect magnetic induction effects.