Position-dependent effective mass and Galilean invariance

Jean-Marc Lévy-Leblond
Physique Théorique, Université de Nice, Parc Valrose, F-06108 Nice Cedex 2, France
(Received 25 April 1995)

Instantaneous Galilean invariance is used to derive from first principles the expression for the Hamiltonian of an electron with a position-dependent effective mass, as well as the adequate boundary conditions for the wave function in the case of abrupt heterojunctions. A very elementary model sustaining these results in the envelope-function approximation is also proposed.

PACS number(s): 03.65.−w

I. INTRODUCTION

The study of semiconductor heterostructures [1] (and, more generally, of inhomogeneous crystals) has given rise in the past decade to an extended discussion concerning the use of simple effective-mass model descriptions (related to the envelope-function approximations [2–4]) for the dynamics of electrons in such systems [5–17]. The main points in debate have concerned the following:

(a) The boundary conditions at abrupt interfaces, characterized by discontinuities in the mass function M [5,7,8,9,11,12]. While one keeps to the continuity of the wave function, the question concerns the replacement of the continuity condition on its derivative by a condition involving the (discontinuous) mass M.

(b) The form of the kinetic Hamiltonian [7–9,13–17]. How is one to generalize the usual expression

$$H_{\text{kin}}=\frac{1}{2m}P^2 \tag{1.1}$$

if the mass m becomes a spatial function $M(X)$ (P and X being, respectively, the momentum and positron operators)? Indeed, $M(X)$ no longer commutes with P, which makes H_{kin} non-Hermitian if one simply replaces m with $M(X)$ in (1.1). The point is that there are many different ways to generalize (1.1) in order to obtain a Hermitian operator; for instance, the two-parameter family

$$H_{\text{kin}}\overset{?}{=}\tfrac{1}{4}(M^{\alpha}PM^{\beta}PM^{\gamma}+M^{\gamma}PM^{\beta}PM^{\alpha}) \quad \text{with } \alpha+\beta+\gamma=-1 \tag{1.2}$$

(about which we will have more to say later).

(c) The very applicability of the concept of a position-dependent effective mass [6,10]. As the notion of mass in nonrelativistic quantum theory is closely linked to the Galilean invariance of the theory for free particles, it has been argued that the obvious lack of Galilean invariance in inhomogeneous crystalline structures would invalidate the use of a nonconstant effective mass [6].

Now, most specific studies comparing exact treatments of simple models (such as Kronig-Penney lattices), with their effective-mass approximate calculations, have led to the agreement that, in many cases, the following conclusions are valid—taking up the above questions in reverse order:

(c) The idea of a position-dependent effective mass $M(X)$ is consistent and useful.

(b) The correct kinetic Hamiltonian is

$$H_{\text{kin}}=P\frac{1}{2M(X)}P\ . \tag{1.3}$$

(a) The correct boundary conditions for the derivative of the wave function Ψ consist in requiring the continuity of $[1/M(X)](d\Psi/dx)$.

This work is devoted to sustaining and strengthening these conclusions from a more fundamental point of view; that is, by basing them on considerations of Galilean invariance [18], as adapted to the present situation. In other words, not only does the use of position-dependent effective mass give correct approximations, but it is also a conceptually consistent approach.

We will first show how the idea of instantaneous Galilean invariance offers a well-defined and useful framework for the concept of a position-dependent mass and leads to a specific family of acceptable Hamiltonians. Consideration of the ensuing Schrödinger equation then allows a derivation of the appropriate boundary conditions, and its application to the case of abrupt interfaces enables one to select the unique form (1.3) for the kinetic Hamiltonian. Finally, the Appendix offers a very elementary discrete model of a heterojunction, which enables one to discuss the approximate validity of the boundary condition. Let us point out here that the validity of the effective-mass description for abrupt heterojunctions gives novel interest to the standard textbook problems of elementary quantum mechanics with flat potentials (step, barrier, and Kronig-Penney models) generalized to the case of a position-dependent mass [19].

II. INSTANTANEOUS GALILEAN INVARIANCE AND THE HAMILTONIAN

For a single "nonrelativistic" quantum particle, the functional form of the Hamiltonian in terms of the canonical operators (X,P) is governed by Galilean invariance [18]. More precisely, a free particle is characterized by its mass m, a constant specifying the relevant unitary projective representation of the Galilean group, and its Hamiltonian reads

$$H_0=\frac{P^2}{2m}+W_0 \ , \tag{2.1}$$

where W_0 is a constant (internal energy). For a particle subjected to external forces, complete symmetry under the Galilean group is obviously broken; however, an argument based on a partial symmetry, namely, invariance under *instantaneous* Galilean transformations, dictates without ambiguity the following form for the Hamiltonian:

$$H=\frac{P^2}{2m}+W(X) \ , \tag{2.2}$$

proving that interaction has to be specified only through a potential function $W(X)$. We are going to show that the argument may be generalized most naturally to the case where the mass is no longer supposed to be a constant but may depend on the position of the particle (and possibly on time).

Let us first briefly recall the derivation for the standard case of constant mass. For simplicity, we stick to the one-dimensional case. In classical mechanics, a Galilean transformation effected at the instant t_0, with velocity v, transforms the position x and momentum p of a particle with mass m according to

$$x'(t)=x(t)-v(t-t_0) \ , \tag{2.3a}$$

$$p'(t)=p(t)-mv \ . \tag{2.3b}$$

An instantaneous Galilean transformation, effected at the very instant $t_0=t$, thus is defined by

$$x'=x \ , \tag{2.4a}$$

$$p'=p-mv \ . \tag{2.4b}$$

In the same way, we may always define, for a quantum particle, a unitary transformation $U(v)$, implementing the instantaneous Galilean transformation with velocity v and acting on the canonical pair of operators X and P according to

$$U(v)XU^{-1}(v)=X \tag{2.5a}$$

$$U(v)PU^{-1}(v)=P-mvI \ . \tag{2.5b}$$

We now *require* the standard Galilean transformation law for the velocity operator V to be valid. Namely, if H is the Hamiltonian and

$$V=i[H,X] \tag{2.6}$$

the velocity of the particle, we want the following transformation law to hold:

$$U(v)VU^{-1}(v)=V-vI \ . \tag{2.7}$$

Because (2.5) defines $U(v)$ without ambiguity, and because V depends on H through (2.6), the condition (2.7) is in effect a constraint on the Hamiltonian. To exploit it, let us introduce the infinitesimal generator K of instantaneous Galilean transformations through

$$U(v)=\exp(ivK) \ . \tag{2.8}$$

From (2.5) and (2.6), we may derive the equivalent commutation rules:

$$[K,X]=0 \ , \tag{2.9a}$$

$$[K,P]=imI \ , \tag{2.9b}$$

$$[K,V]=iI \ , \tag{2.9c}$$

and from (2.9a) and (2.9b), it now follows, according to the canonical commutation rule

$$[X,P]=iI \ , \tag{2.10}$$

that the generator K must read

$$K=mX \tag{2.11}$$

(up to a trivial additive constant). From (2.9b) and (2.9c), it is seen that

$$[K,P-mV]=0 \ , \tag{2.12}$$

implying, because of (2.10), that $P-mV$ is a function of X only (and not P):

$$P-mV=A(X) \ . \tag{2.13}$$

Similarly, one easily computes the following commutator:

$$[K,H-\tfrac{1}{2}mV^2]=m[X,H]-\frac{m}{2}[X,V^2]=imV-imV=0 \ . \tag{2.14}$$

Hence, $H-\frac{1}{2}mV^2$ is a function of X only:

$$H-\tfrac{1}{2}mV^2=W(X) \ . \tag{2.15}$$

Finally, the Hamiltonian takes the form

$$H=\frac{1}{2m}[P-A(X)]^2+W(X) \ , \tag{2.16}$$

that is, is characterized by a "vector" potential $A(X)$ and a scalar one $W(X)$. For the one-dimensional case, the "vector" potential $A(X)$ may be eliminated by a simple phase transformation, so that we recover the common form (2.2) of the Hamiltonian.

III. CASE OF POSITION-DEPENDENT MASS

It is a simple matter to generalize the argument if the mass is no longer supposed to be a numerical constant m and becomes instead an operator function $M(X)$. Indeed, the very fact that instantaneous Galilean transformations do not modify the position [see (2.5a)] means that they are quite indifferent to a possible position dependence of the mass. That is the crucial point that allows a consistent generalization of the preceding argument. We thus modify the transformation law for the momentum (2.5b) in

$$U(v)PU^{-1}(v)=P-M(X)v \ , \tag{3.1}$$

yielding the commutator

$$[K,P]=iM(X) \tag{3.2}$$

instead of (2.9b). From (2.9a) (unchanged), it follows that

K is only a function of X:

$$K=N(X) . \tag{3.3}$$

The canonical commutation rule now implies that

$$[K,P]=[N(X),P]=iN'(X) . \tag{3.4}$$

The function N giving the generator K in (3.3) thus has M as its derivative, generalizing (2.11). Consider now the constraining condition on H, that is, (2.9c). It reads, according to (2.6) and (3.3),

$$[N(X),[H,X]]=I . \tag{3.5}$$

From the Jacobi identity, it follows that since $[X,N(X)]=0$, this may be rewritten as

$$[X,[N(X),H]]=I . \tag{3.6}$$

Comparison with (2.10) now shows that

$$[N(X),H]=i[P-A(X)] . \tag{3.7}$$

where $A(X)$ is an arbitrary and, as above, inessential "vector" potential, which we eliminate. The final condition on the Hamiltonian H reads

$$[N(X),H]=iP . \tag{3.8}$$

It may be checked easily that a simple solution (for a Hermitian Hamiltonian) is

$$H_0=\frac{1}{2}P\frac{1}{M(X)}P . \tag{3.9}$$

Indeed,

$$\begin{aligned}[N(X),H_0]&=\frac{1}{2}\left[N(X),P\frac{1}{M(X)}P\right]\\&=\frac{1}{2}[N(X),P]\frac{1}{M(X)}P+\frac{1}{2}P\frac{1}{M(X)}[N(X),P]\\&=iP\end{aligned} \tag{3.10}$$

when use is made of (3.2) and (3.3). As a consequence, the difference $H-H_0$ (between the most general solution H and the particular one H_0) commutes with $N(X)$ and must itself depend on X only. The *general* Hamiltonian thus reads,

$$H=\frac{1}{2}P\frac{1}{M(X)}P+W_0(X) . \tag{3.11}$$

Instead of H_0 (3.9), we could have chosen another particular solution of (3.8); for instance, the rather natural one:

$$H_1=\frac{1}{4}\left[P^2\frac{1}{M(X)}+\frac{1}{M(X)}P^2\right] . \tag{3.12}$$

However, it is easily seen that

$$H_1-H_0=\frac{1}{4}\left[P\left[P,\frac{1}{M(X)}\right]\right]=Q(X) . \tag{3.13}$$

where the function Q is defined according to

$$Q(x)=-\frac{1}{4}[M^{-1}(x)]''=-\frac{1}{2}\frac{M'^2}{M^3}+\frac{1}{4}\frac{M''}{M^2} . \tag{3.14}$$

The Hamiltonian H_1 thus is of the general form (3.11), as expected. In terms of H_1, the general Hamiltonian H may be rewritten as

$$\begin{aligned}H&=\frac{1}{2}P\frac{1}{M(X)}P+W_0(X)\\&=\frac{1}{4}\left[P^2\frac{1}{M(X)}+\frac{1}{M(X)}P^2\right]+W_1(X) ,\end{aligned} \tag{3.15}$$

with the relationship

$$W_1(X)=W_0(X)+Q(X) . \tag{3.16}$$

The preceding remarks suffice to prove that one should not identify *a priori* H_0 (3.9) to the purely kinetic term of the Hamiltonian nor W_0 in (3.11) to the potential, since, indeed, H_1 (3.12) and W_1 (3.16) may qualify also, as well as an infinite number of other possibilities.

It is possible, nevertheless, to restrict the form of purely kinetic Hamiltonian to a rather specific class. We start from the idea that the kinetic Hamiltonian must depend only on the mass function M. Accordingly, it will take the general form (3.11)

$$H_{\text{kin}}=\frac{1}{2}P\frac{1}{M(X)}P+W_{\text{kin}}(X) , \tag{3.17}$$

with the condition that the term W_{kin} be a functional of M, possibly involving its derivatives. Dimensional arguments now require this term to be homogeneous of degree (-1) in M and of degree (-2) in X. Analyticity conditions precluding nonintegral powers of the derivatives of M and, finally, the condition that for a constant function $M(X)=m$ one recovers the usual expression, implying that the derivatives of M must appear with positive (integral) powers, lead to two possible terms only in W_{kin}:

$$W_{\text{kin}}=\lambda\frac{M'^2}{M^3}+\mu\frac{M''}{M^2} . \tag{3.18}$$

This result may be used to prove that the most general kinetic Hamiltonian (under the conditions we have stated) is precisely of the form (1.2). Indeed, using the standard commutation relations, one may check the following relationship:

$$\begin{aligned}H_{\text{kin}}&=\frac{1}{4}(M^{\alpha}PM^{\beta}PM^{\gamma}+M^{\gamma}PM^{\beta}PM^{\alpha})\\&=P\frac{1}{2M}P+\frac{1}{2}(\alpha+\gamma+\alpha\gamma)\frac{M'^2}{M^3}-\frac{1}{4}(\alpha+\gamma)\frac{M''}{M^2}\end{aligned}$$

$$(\text{for } \alpha+\beta+\gamma=-1) . \tag{3.19}$$

It is sufficient to relate the coefficients (α,β,γ) in (3.19) and (λ,μ) in (3.18) through

$$\tfrac{1}{2}(\alpha+\gamma+\alpha\gamma)=\lambda, \quad -\tfrac{1}{4}(\alpha+\gamma)=\mu \tag{3.20}$$

to identify (3.17) and (3.18) with (1.2).

IV. SCHRÖDINGER EQUATION AND THE CONTINUITY CONDITIONS

We will now show that, among the general class of kinetic Hamiltonians just discussed, it is indeed H_0 (3.9) that must be chosen as *the* kinetic Hamiltonian. The time-dependent Schrödinger equation derived from the general Hamiltonian (3.11) rules the wave function $\Psi(x,t)$ according to

$$i\frac{\partial\Psi}{\partial t}=-\frac{1}{2}\frac{\partial}{\partial x}\left[\frac{1}{M(x)}\frac{\partial\Psi}{\partial x}\right]+W(x)\Psi\ . \tag{4.1}$$

We have already stressed in Sec. I the practical interest of the cases where the mass function $M(x)$ shows discontinuities. It is thus necessary to inquire about the continuity conditions on the wave function. As a matter of fact, it is no longer the derivative $\partial\Psi/\partial x$ itself but rather the combination $[1/M(x)](\partial\Psi/\partial x]$ that must be made continuous. Indeed, this may be proved through a simple extension of the usual textbook argument. Let us start from the Schrödinger equation (4.1) and suppose the mass $M(x)$ and the potential $W(x)$, *as defined by the choice* (3.11), to be bounded on some interval around the point x_0, while Ψ is twice differentiable in x (and once in t). Integrating between $x_0-\varepsilon$ and $x_0+\varepsilon$, we obtain

$$\frac{1}{M(x)}\frac{\partial\Psi}{\partial x}\Bigg|_{x_0-\varepsilon}^{x_0+\varepsilon}=2\int_{x_0-\varepsilon}^{x_0+\varepsilon}\left[W(x)\Psi-i\frac{\partial\Psi}{\partial t}\right]dx\ . \tag{4.2}$$

If we let ε go to zero, the right-hand side vanishes, proving the continuity of $[1/M(x)](\partial\Psi/\partial x)$, from which follows that of Ψ itself. We thus have the following.

Continuity conditions for variable mass: for bounded but possibly discontinuous mass and potential functions $M(x)$ and $W(x)$, the wave function $\Psi(x,t)$ and its derivative *divided by the mass* $[1/M(x)](\partial\Psi/\partial x)$ are continuous in x.

It may now be seen that any other choice than (3.9) for the kinetic Hamiltonian would lead to inconsistencies in the simple case where the mass is discontinuous and there is no potential. Let us, for instance, choose the kinetic Hamiltonian H_1 (3.12) and suppose $M(x)$ to have a (finite) discontinuity at the point x_0. Then, according to (3.13) and (3.14), the Schrödinger equation for H_1 is equivalent to that for H_0 with a fake potential term $Q(x)$ exhibiting a derivative of the Dirac function in x_0: $Q(x)\propto\delta'(x-x_0)$. The above reasoning breaks down, as the right-hand side in (4.2) will contain a singular unbounded term in x_0. The same can be said for any other choice of the kinetic Hamiltonian except H_0, since, according to (3.18), the right-hand side will contain terms in δ' and, worse, in $\langle\langle\delta^2\rangle\rangle$. Finally, it may be said that the kinetic Hamiltonian H_0 is singled out by the requirement of a consistent continuity condition for the cases with discontinuous mass.

ACKNOWLEDGMENTS

It is a pleasure to thank G. Bastard for stimulating my interest in the present problem and giving me the necessary background information, and E. Beltrametti for his kind invitation to the University of Genoa, where I could test some of these ideas.

APPENDIX: A SIMPLE EXAMPLE OF THE ENVELOPE-FUNCTION APPROXIMATION

The most elementary model of quantum propagation in a lattice is a discrete one in which localized states, forming a basis, are attributed to each site $x_n=na$ (n integer). The Hamiltonian is supposed to have, besides its diagonal matrix elements E_0, only nearest-neighbor nondiagonal matrix elements $-A$. The eigenvalue equation for the localization amplitudes $u_n=\langle x_n|u\rangle$ of a stationary state u then reads

$$-Au_{n-1}+E_0u_n-Au_{n+1}=Eu_n\ . \tag{A1}$$

The standard solution of this recursion relation is

$$u_n=ce^{in\alpha}+de^{-in\alpha}\ , \tag{A2}$$

yielding a permitted band

$$E=E_0-2A\cos\alpha \tag{A3}$$

for the energy. From the amplitudes (A2), one may define the envelope-function

$$\varphi(x)=ce^{ipx}+de^{-ipx}\ , \tag{A4}$$

with

$$p=\frac{\alpha}{a} \tag{A5}$$

being the quasimomentum, so that

$$u_n=\varphi(x_n)\ . \tag{A6}$$

At the bottom of the band, one may write

$$E\approx E_0-2A+\frac{A}{a^2}\quad(\alpha=p/a\ll1)\ , \tag{A7}$$

which is the dispersion relation for a free quasiparticle with effective mass

$$m=\frac{a^2}{2A}\ . \tag{A8}$$

Now, this rudimentary model is easily extended to heterostructures by giving different values to the matrix elements of the Hamiltonian in different regions. Consider an abrupt heterojunction with the matrix elements

$$\begin{aligned}H_{n,n}&=\begin{cases}E_0, & n<0\\ \frac{1}{2}(E_0+E'_0), & n=0\\ E'_0, & n>0\end{cases}\\ H_{n,n+1}&=\begin{cases}-A, & n\le-1\\ -A', & n\ge0\ ,\end{cases}\end{aligned} \tag{A9}$$

so that the eigenvalue equation reads like (A1) for $n<0$, with $(E_0,-A)$ being replaced by $(E'_0,-A')$ for $n>0$, and, for $n=0$,

$$Au_{-1}+\tfrac{1}{2}(E_0+E'_0)u_0-A'u_1=Eu_0\ . \tag{A10}$$

The general solution is of the type

$$u_n = \begin{cases} ce^{in\alpha} + de^{-in\alpha}, & n \leq 0 \\ c'e^{in\alpha'} + d'e^{-in\alpha'}, & n \geq 0 \end{cases}, \tag{A11}$$

with the necessary condition

$$c + d = c' + d' . \tag{A12}$$

Using in (A10) the expressions (A11) for $n = \pm 1$, and taking into account (A12), we immediately derive a second compatibility condition:

$$A(c - d)\sin\alpha = A'(c' - d')\sin\alpha' . \tag{A13}$$

Consider now the envelope function

$$\varphi(x) = \begin{cases} ce^{ipx} + de^{-ipx}, & x < 0 \\ c'e^{ip'x} + d'e^{-ip'x'}, & x > 0 \end{cases}, \tag{A14}$$

with

$$p = \frac{\alpha}{a}, \quad p' = \frac{\alpha'}{a} . \tag{A15}$$

The continuity of φ at the junction is ensured by (A12). The left and right derivatives now read

$$\left.\frac{\partial\varphi}{\partial x}\right|_{x=0+} = p'(c' - d') ,$$
$$\left.\frac{\partial\varphi}{\partial x}\right|_{x=0-} = p(c - d) . \tag{A16}$$

Using (A13) and (A15), we see that

$$A'\frac{\sin\alpha'}{\alpha'}\left.\frac{\partial\varphi}{\partial x}\right|_{0+} = A\frac{\sin\alpha}{\alpha}\left.\frac{\partial\varphi}{\partial x}\right|_{0-} . \tag{A17}$$

Finally, by remembering that one is dealing with the case $\alpha, \alpha' \ll 1$ and that effective masses may be defined according to (A8) (and similarly $m' = a^2/2A'$), we reach the condition

$$\frac{1}{m'}\left.\frac{\partial\varphi}{\partial x}\right|_{0+} = \frac{1}{m}\left.\frac{\partial\varphi}{\partial x}\right|_{0-} , \tag{A18}$$

that is, a continuity condition for *the derivative of the wave function divided by the mass.*

[1] For a recent review, see D. L. Smith and C. Mailhiot, Rev. Mod. Phys. **62**, 173 (1990).
[2] G. Bastard, Phys. Rev. B **24**, 5693 (1981).
[3] M. Altarelli, in *Semiconductor Superlattices and Heterojunctions,* Les Houches Winterschool, 1985, edited by G. Allan, G. Bastard, N. Boccara, M. Lannoo, and M. Voss (Springer, Berlin, 1986), p. 12.
[4] G. Bastard, Surf. Sci. **170**, 426 (1986).
[5] Q. G. Zhu and H. Kroemer, Phys. Rev. B **27**, 3519 (1983).
[6] O. von Roos, Phys. Rev. B **27**, 7547 (1983).
[7] R. A. Morrow and K. R. Brownstein, Phys. Rev. B **30**, 678 (1984).
[8] R. A. Morrow, Phys. Rev. B **35**, 8074 (1987).
[9] R. A. Morrow, Phys. Rev. B **36**, 4836 (1987).
[10] K. Kawamura and R. A. Brown, Phys. Rev. B **37**, 3932 (1988).
[11] W. Trzeciakowski, Phys. Rev. B **38**, 4322 (1988).
[12] I. Galbraith and G. Duggan, Phys. Rev. B **38**, 10057 (1988).
[13] J. Thomsen, G. T. Einevoll, and P. C. Hemmer, Phys. Rev. B **39**, 12 783 (1989).
[14] K. Young, Phys. Rev. B **39**, 13 434 (1989).
[15] Y. Fu and K. A. Chao, Phys. Rev. B **40**, 8349 (1989).
[16] G. T. Einevoll, P. C. Hemmer, and J. Thomsen, Phys. Rev. B **42**, 3485 (1990).
[17] G. T. Einevoll, P. C. Hemmer, and J. Thomsen, Phys. Rev. B **42**, 3497 (1990).
[18] J.-M. Lévy-Leblond, Riv. Nuovo Cimento **4**, 99 (1974).
[19] J.-M. Lévy-Leblond, Eur. J. Phys. **13**, 215 (1992).