Position-dependent effective mass for inhomogeneous semiconductors

K. Young
Department of Physics, The Chinese University of Hong Kong, Hong Kong
(Received 10 November 1988)

A systematic approach is adopted to extract an effective low-energy Hamiltonian for crystals with a slowly varying inhomogeneity, resolving several controversies. It is shown that the effective mass $m(\mathbf{R})$ is, in general, position dependent, and enters the kinetic energy operator as $-\nabla[m(\mathbf{R})^{-1}]\nabla/2$. The advantage of using a basis set that exactly diagonalizes the Hamiltonian in the homogeneous limit is emphasized.

I. INTRODUCTION

The motion of charge carriers in a conduction band, for example in a semiconductor, is described by the band edge E and the effective mass m. When lattice distortion,[1-3] graded chemical composition,[4-6] or heterostructure[7,8] introduces inhomogeneities which vary slowly, on a length scale $\lambda_q \gg a$, where a is the lattice constant, it is natural to consider a *position-dependent* band edge $E(\mathbf{R})$ and effective mass $m(\mathbf{R})$, both determined by the local properties of the crystal near $\mathbf{R}$. The force $e\nabla E(\mathbf{R})$ on the charge carriers e has well-known consequences for charge transport,[9] while the varying band gap can match a range of wavelengths in photovoltaic devices.[10] However, there are several controversies concerning a position-dependent $m(\mathbf{R})$. (1) It has been asserted that $m(\mathbf{R})$ is in fact independent of $\mathbf{R}$.[11] (2) Various inequivalent forms have been proposed for the kinetic energy operator $\hat{T}$ in terms of $\mu(\mathbf{R})=m(\mathbf{R})^{-1}$:

$$\hat{T}=-(\mu\nabla^2+\nabla^2\mu)/4 \quad \text{(Ref. 4)},$$

$$\hat{T}=-\nabla\mu\nabla/2 \quad \text{(Ref. 11)},$$

$$T=-\sqrt{\mu}\nabla^2\sqrt{\mu}/2 \quad \text{(Ref. 7)},$$

or more generally as

$$\hat{T}=-(\mu^{\alpha}\nabla\mu^{\beta}\nabla\mu^{\gamma}+\text{H.c.})/4$$

in terms of the von Roos parameters[12] α, β, and γ with $\alpha+\beta+\gamma=1$. The ambiguity is of first and second order in $\nabla m(\mathbf{R})/m(\mathbf{R})\sim q$, where $q=2\pi/\lambda_q$ is a characteristic wave number associated with the inhomogeneity. The lack of uniqueness has caused concern,[12] and the different possibilities affect the boundary condition across heterojunctions.[7] (3) The consistency of a position-dependent $m(\mathbf{R})$ with Galilean invariance and the Bargmann superselection rule[13] has been discussed.[12,14] (4) Finally, the semiclassical limit has received attention, with one suggestion that a dissipative force appears.[6]

We address these issues in this paper, with particular reference to the problem of uniqueness. The apparent lack of determinacy stems from attempts to approximate a matrix element of the form

$$\int d^3r\, W(\mathbf{r}-\mathbf{R})^* L(\mathbf{r}) U(\mathbf{r}) W(\mathbf{r}-\mathbf{R}'), \tag{1}$$

in which W are (fairly localized) Wannier functions centered at $\mathbf{r}=\mathbf{R}$ and $\mathbf{r}=\mathbf{R}'$, L is the slowly varying inhomogenity profile, and $U(\mathbf{r})$ is some potential (not necessarily slowly varying). The effective-mass approximation is obtained by regarding $L(\mathbf{r})$ as nearly constant near $\mathbf{R}$ and $\mathbf{R}'$, but different approximations, e.g.,

$$L(\mathbf{r})\approx[L(\mathbf{R})+L(\mathbf{R}')]/2, \tag{2a}$$

$$L(\mathbf{r})\approx L((\mathbf{R}+\mathbf{R}')/2) \tag{2b}$$

lead to inequivalent kinetic energy operator $\hat{T}$, i.e., different von Roos parameters α, β, and γ. These ambiguities can only be resolved if (2) is replaced by an approximation valid to $O(q^2)$, where $(\nabla)^N L(\mathbf{r})=O(q^N)$. We shall show that a systematic formulation to this accuracy yields a *unique* kinetic energy operator, of the form $-\nabla\mu\nabla/2$, i.e., von Roos parameters $\alpha=\gamma=0$ and $\beta=1$. There are, however, corrections to the band edge proportional to ∇L and $\nabla^2 L$.

We first comment on the length and wave-number scales in the problem and on the appropriate basis set for expressing the calculation.

The effective-mass approximation, based on an expansion in the wave number k (to second order), is valid for $k \ll a^{-1}$ (though in practice it often works well for a large fraction of the band). The inhomogeneity length scale defines another small parameter $q=2\pi/\lambda_q \ll a^{-1}$. Since $k \lesssim a^{-1}$ and $q\sim$(size of sample)$^{-1}$, it might appear that powers of q need not be kept. This argument is implicitly invoked when the possible nonuniqueness of $\hat{T}$ is said to be unimportant, being of order $\nabla m/m\sim q$.[15] However, terms in the Hamiltonian of order ∇m or ∇L lead to differences in the wave number of the same order, and hence to phases [e.g., in the Wentzel-Kramers-Brillouin (WKB) approximation] of order

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2}\nabla L\cdot d\mathbf{r}=L(\mathbf{r}_2)-L(\mathbf{r}_1), \tag{3}$$

which are *not* negligible in graded semiconductors, where the composition may change by typically 30% across the sample, or across heterojunctions. For this reason, the first power of q *must* be kept, no matter how smooth the inhomogeneity. To be systematic, we shall keep all terms $k^M q^N$, with $M+N\leq 2$, in order to extract a low-energy Hamiltonian from the full theory. Except in Sec. V, we

do not assume the amplitude of the inhomogeneity to be small.

In one group of works in the literature,[5,11] wave functions are represented in the Luttinger-Kohn (LK) basis[16] $\{\varphi_{n\mathbf{k}}(\mathbf{r})\}$, where $\mathbf{k}$ is a Bloch wave vector and n is superficially a band label. These, unfortunately, do not diagonalize even the homogeneous part of the Hamiltonian. In other words, while the set $\{\varphi_{n\mathbf{k}}\}$ for all $\mathbf{k}$ and all n is complete, the subspace spanned by $\{\varphi_{n\mathbf{k}}\}$ for all $\mathbf{k}$ and *one fixed* n does not coincide with the nth band, so that n in $\{\varphi_{n\mathbf{k}}\}$ is a pseudoband label. A unitary transformation $e^{i\hat{\theta}}$ is therefore needed to diagonalize the effective low-energy Hamiltonian, but in practice this can only be implemented to a certain order in $\hat{\theta}$, so that the homogeneous limit, which provides the motivation and physical insights for the position-dependent effective-mass theory, is not obtained exactly. Secondly, the classification of matrix elements in the LK basis, e.g., $\langle n\mathbf{k}|U|n'\mathbf{k}'\rangle$, as diagonal ($n=n'$) or off-diagonal ($n\neq n'$) does not correspond to the distinction between intraband and interband processes, obscuring physical interpretations.

Therefore we shall adopt the more convenient Wannier basis,[6,12,15] which does not suffer from these problems. In fact, one can write the unitary transformation as[11] $\hat{\theta}=\hat{\theta}_1+\hat{\theta}_2+\cdots$, where $\hat{\theta}_1$ serves to remove the off-diagonal part of the homogeneous Hamiltonian. Obviously, $\hat{\theta}_1$ exactly takes us from the LK basis to the Wannier or Bloch basis, so that most of the problem associated with working to finite order in $\hat{\theta}$ is automatically taken care of.

The practical significance of a correct treatment of a position-dependent effective mass may be illustrated through semiconductor materials of current interest to research in optoelectronic devices, for example $In_xGa_{1-x}As$, where the effective mass, say of the electrons, varies by over a factor of 2 (from 0.07 to 0.026 times the electron mass) as x varies from 0 to 1. In materials relevant to research in band-gap engineering, such variations can be achieved over hundreds or even tens of atomic layers, implying that ∇m is substantial and must be dealt with correctly.

The rest of this paper is organized as follows. In Sec. II we first present a simple physical argument to show why $m(\mathbf{R})$ must be position dependent; otherwise the other issues would not arise. We then formulate the problem and derive our main result. Section III presents a similar low-energy analysis for the external potential V, showing that it is modified by terms of order $a^2\nabla^2 V$. In Sec. IV we discuss Galilean invariance and semiclassical correspondence and in Sec. V we deal with claims that $m(\mathbf{R})$ is independent of $\mathbf{R}$ and show that incautious use of the LK basis can lead to erroneous results. Our results in the Wannier basis are explicitly compared with the known results in the LK basis.[5] The conclusion is given in Sec. VI.

II. FORMULATION

Consider a crystal with the Hamiltonian

$$\hat{H}=\hat{H}_0+L(\mathbf{r})U(\mathbf{r})+V(\mathbf{r}) , \tag{4}$$

where H_0 describes the reference homogeneous material with a certain crystal symmetry, for example,

$$\hat{H}_0=-\frac{1}{2m_0}\nabla^2+\sum_{\mathbf{R}} U_A(\mathbf{r}-\mathbf{R}) , \tag{5}$$

where $\hbar=1$, $\mathbf{R}$ denotes the lattice sites, and U_A is the potential due to an atom A. $L(\mathbf{r})$ is the slowly varying inhomogeneity profile and the potential $U(\mathbf{r})$ has the same crystal symmetry as $\hat{H}_0$. For example, if a fraction $\rho(\mathbf{r})$ of the A atoms are replaced by B atoms, then in the virtual-crystal approximation $L(\mathbf{r})=\rho(\mathbf{r})$ and

$$U(\mathbf{r})=\sum_{\mathbf{R}}[U_B(\mathbf{r}-\mathbf{R})-U_A(\mathbf{r}-\mathbf{R})] . \tag{6}$$

The last term in (4) is an external potential, whose time dependence, if any, will not be explicitly indicated.

In view of claims to the contrary, we first give a physical argument to show why $m(\mathbf{R})$ must depend on $L(\mathbf{R})$, i.e., $\nabla m(R)$ is in general not zero and not negligible. Consider a sample divided into three regions 1, 2, and 3, each of macroscopic dimension D. In regions 1 and 3, $L(\mathbf{R})$ assumes constant values L_1 and L_3, which may be quite different; in the intermediate region 2, L interpolates smoothly between these two values. Construct a wave packet of size $d \ll D$ in region 1; so long as $d \gg a$, this can be achieved by superposing states with momentum $|\mathbf{k}| \ll a^{-1}$, to which the concept of effective mass applies, i.e., the expansion to $O(k^2)$ will be accurate. For simplicity let the external potential be zero. The evolution of this wave packet must be indistinguishable from that in a *homogeneous* material described by $\hat{H}_1=\hat{H}_0+L_1U(\mathbf{r})$, for which there is, by the usual theory, an effective mass m_1. The same argument applies to region 3. However, by the usual *homogeneous* theory, the different limiting Hamiltonians $\hat{H}_1$ and $\hat{H}_3$ obviously give different effective masses m_1 and m_3. Except in Sec. V, we do not assume that the change in L is small, only that it is slow; thus $m_1-m_3 \propto L_1-L_3$ cannot be ignored. The above remarks, essentially a paraphrasing of the arguments leading to (3), provide the main motivation behind the concept of a position-dependent effective mass, and underlines the need to preserve an exact homogeneous limit, and thus the danger associated with the LK basis—an issue to be discussed at length.

Returning to (4), we work in the Bloch basis which diagonalized $\hat{H}_0$:

$$\hat{H}_0\psi_{n\mathbf{k}}(\mathbf{r})=E_n(\mathbf{k})\psi_{n\mathbf{k}}(\mathbf{r}) , \tag{7}$$

where n is a band index and for simplicity we assume that the bands have nondegenerate minima at $\mathbf{k}=\mathbf{0}$. The usual effective mass $\vec{\mathbf{m}}_n=(\vec{\mu})^{-1}$ in band n is given through the expression

$$E_n(\mathbf{k})=E_n+\tfrac{1}{2}\mu_n^{ij}k_ik_j+O(k^3) , \tag{8}$$

in which i and j are Cartesian indices. As usual, we write

$$\psi_{n\mathbf{k}}(\mathbf{r})=e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r}) , \tag{9}$$

where, for future reference, we expand the periodic functions $u_{n\mathbf{k}}$ in powers of k:

$$u_{n\mathbf{k}}(\mathbf{r})=f_n(\mathbf{r})+ik_i g_n^i(\mathbf{r})-\tfrac{1}{2}k_i k_j h_n^{ij}(\mathbf{r})+\cdots . \quad (10)$$

Since $\psi_{n\mathbf{k}}^*$ and $\psi_{n-\mathbf{k}}$ differ at most by a phase, we can choose f_n, g_n^i, and h_n^{ij} all to be real functions. We shall also need the Wannier functions at site $\mathbf{R}$,

$$\psi_{n\mathbf{k}}(\mathbf{r})=N^{-1/2}\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}W_n(\mathbf{r}-\mathbf{R}) . \quad (11)$$

We now represent an arbitrary wave function $\psi(\mathbf{r},t)$ as

$$\psi(\mathbf{r},t)=\sum_{n,\mathbf{R}} F_n(\mathbf{R},t)W_n(\mathbf{r}-\mathbf{R}) , \quad (12)$$

and the object is to find an effective Hamiltonian to describe the evolution of the envelope functions F_n. Using the completeness of the Wannier functions, we have

$$\begin{aligned} i\frac{\partial F_n}{\partial t}(\mathbf{R},t)&=\left\langle n\mathbf{R}\left| i\frac{\partial}{\partial t}\right|\psi\right\rangle \\ &=\langle n\mathbf{R}|\hat{H}|\psi\rangle \\ &=\sum_{n',R'}\langle n\mathbf{R}|\hat{H}|n'\mathbf{R}'\rangle F_{n'}(\mathbf{R}',t) , \end{aligned} \quad (13)$$

where in obvious notation $|n\mathbf{R}\rangle$ denotes the Wannier state $W_n(\mathbf{r}-\mathbf{R})$. The problem then reduces to evaluating the matrix elements of $\hat{H}$ between Wannier states. In this section we ignore the external potential V and consider only $\hat{H}_0$ and $L(\mathbf{r})U(\mathbf{r})$.

Going through standard manipulations, we find

$$\langle n\mathbf{R}|\hat{H}_0|n'\mathbf{R}'\rangle=N^{-1}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')}E_n(\mathbf{k})\delta_{nn'} , \quad (14)$$

and using (8) for $E_n(\mathbf{k})$ and going to a continuum approximation [i.e., assuming $F_n(\mathbf{R},t)$ is sufficiently slowly varying in $\mathbf{R}$], we get

$$\sum_{n',\mathbf{R}'}\langle n\mathbf{R}|\hat{H}_0|n'\mathbf{R}'\rangle F_{n'}(\mathbf{R}'t)=(E_n-\tfrac{1}{2}\mu_n^{ij}\nabla_i\nabla_j)F_n(\mathbf{R},t) . \quad (15)$$

[More generally, the factor in parentheses should be replaced by $E_n(-i\nabla)$, but for virtually all applications, one stops at second order.] We may summarize (15) schematically as

$$\hat{H}_0\rightarrow(E_n-\tfrac{1}{2}\mu_n^{ij}\nabla_i\nabla_j)\delta_{nn'} . \quad (15')$$

The effect of inhomogeneity is reflected in the matrix element

$$\begin{aligned} &\langle n\mathbf{R}|LU|n'\mathbf{R}'\rangle \\ &\quad=\int d^3r\, W_n(\mathbf{r}-\mathbf{R})^*L(\mathbf{r})U(\mathbf{r})W_{n'}(\mathbf{r}-\mathbf{R}') , \end{aligned} \quad (16)$$

and various approximations are based on replacing $L(\mathbf{r})$ by a constant; see discussion following (1). To resolve the ambiguities requires an approximation accurate to ∇L and $\nabla^2 L$. A Taylor expansion of $L(\mathbf{r})$ about, say, $\mathbf{R}_0=(\mathbf{R}+\mathbf{R}')/2$ is inappropriate because the operators $(\mathbf{r}-\mathbf{R}_0)_i$, $(\mathbf{r}-\mathbf{R}_0)_i(\mathbf{r}-\mathbf{R}_0)_j$, etc. map functions out of the Hilbert space of functions satisfying periodic boundary conditions. It is therefore best to imagine that $L(\mathbf{r})$ also satisfies periodic boundary conditions[17] and expand it as

$$L(\mathbf{r})=\sum_{\mathbf{q}}\tilde{L}(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{r}} , \quad (17)$$

the slow variation of $L(\mathbf{r})$ being reflected in the concentration of $\tilde{L}(\mathbf{q})$ in regions of small $\mathbf{q}$. Then some arithmetic gives

$$\begin{aligned} &\langle n\mathbf{R}|LU|n'\mathbf{R}'\rangle \\ &\quad=N^{-1}\sum_{\mathbf{p},\mathbf{q}}\tilde{L}(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{R}'}e^{i\mathbf{p}\cdot(\mathbf{R}-\mathbf{R}')}S_{nn'}(\mathbf{p},\mathbf{p}+\mathbf{q}) , \end{aligned} \quad (18)$$

where

$$S_{nn'}(\mathbf{p},\mathbf{p}')=N\int_c d^3r\, u_{n\mathbf{p}}(\mathbf{r})^*U(\mathbf{r})u_{n'\mathbf{p}'}(\mathbf{r}) , \quad (19)$$

the integral being over a unit cell c. On account of the exponential phase factors in (18), $\mathbf{p}$ and $\mathbf{q}$ will become ∇ operators, acting either on $L(\mathbf{R})$ or $F_n(\mathbf{R})$, hence giving the small quantities $\nabla L=O(q)$ and $\nabla F_n=O(k)$; thus in order to extract a low-energy effective theory, we shall expand $S_{nn'}$ to second order in its arguments. By using (10) in (19), we find

$$\begin{aligned} S_{nn'}(\mathbf{p},\mathbf{p}')=&U_{nn'}^{0,0}+i(p_i'U_{nn'}^{0,i}-p_iU_{nn'}^{i,0}) \\ &-\tfrac{1}{2}(p_i'p_j'U_{nn'}^{0,ij}+p_ip_jU_{nn'}^{ij,0}) \\ &+p_ip_j'U_{nn'}^{i,j}+\cdots , \end{aligned} \quad (20)$$

where the notation for the superscripts is $0\rightarrow f$, $i\rightarrow g^i$, and $ij\rightarrow h^{ij}$, or more explicitly

$$\begin{aligned} &U_{nn'}^{0,0}=\langle f_n|U|f_{n'}\rangle=N\int_c d^3r\, f_n(\mathbf{r})U(\mathbf{r})f_{n'}(\mathbf{r}) , \\ &U_{nn'}^{0,i}=\langle f_n|U|g_{n'}^i\rangle , \quad U_{nn'}^{i,0}=\langle g_n^i|U|f_{n'}\rangle , \\ &U_{nn'}^{0,ij}=\langle f_n|U|h_{n'}^{ij}\rangle , \quad U_{nn'}^{ij,0}=\langle h_n^{ij}|U|f_{n'}\rangle , \\ &U_{nn'}^{i,j}=\langle g_n^i|U|g_{n'}^j\rangle . \end{aligned} \quad (21)$$

In the rest of this section we assume that U has negligible interband matrix elements,[6,11] so that all matrix elements in (21) are proportional to $\delta_{nn'}$, and we further suppress the band index n in intermediate steps. We shall see in Sec. V that interband matrix elements only affect the results to $O(L^2)$ and are therefore negligible for small L, irrespective of the nature of U. The analogous assumption would not be sensible in the LK basis. For the diagonal matrix elements, (19) simplifies slightly to

$$\begin{aligned} S(\mathbf{p},\mathbf{p}')=&U^{0,0}+i(p_i'-p_i)U^{0,i} \\ &-\tfrac{1}{2}(p_i'p_j'+p_ip_j)U^{0,ij}+p_ip_j'U^{i,j} . \end{aligned} \quad (22)$$

We insert (22) into (18) and perform the sum over $F_{n'}(\mathbf{R}',t)$ as in (13) to get

$$\sum_{n',\mathbf{R}'}\langle n\mathbf{R}|LU|n'\mathbf{R}'\rangle F_{n'}(\mathbf{R}',t)=A_0+A_1+A_{2a}+A_{2b} , \quad (23)$$

where, corresponding to the four terms in (22),

$$A_0 = U^{0,0} L(\mathbf{R}) F_n(\mathbf{R},t) , \tag{24a}$$

$$A_1 = -U^{0,i}[\nabla_i L(\mathbf{R})] F_n(\mathbf{R},t) , \tag{24b}$$

$$A_{2a} = U^{0,ij}\{\nabla_i [L(\mathbf{R})\nabla_j F_n(\mathbf{R},t)] + \tfrac{1}{2}[\nabla_i \nabla_j L(\mathbf{R})] F_n(\mathbf{R},t)\} , \tag{24c}$$

$$A_{2b} = -U^{i,j}\nabla_i [L(\mathbf{R})\nabla_j F_n(\mathbf{R},t)] , \tag{24d}$$

which can be summarized schematically as

$$LU \rightarrow [U^{0,0} L - U^{0,i}(\nabla_i L) + \tfrac{1}{2} U^{0,ij}(\nabla_i \nabla_j L)] + [(U^{0,ij} - U^{i,j})\nabla_i L \nabla_j] . \tag{24'}$$

The calculation leading to (24) is sketched in Appendix A. In (24'), the term in the first set of square brackets is a c number (i.e., diagonal in the $\mathbf{R}$ variable), affecting the band edge but not the effective mass. In the second term, the ∇ operator will operate on the $\mathbf{R}$ variable in F_n as well.

Collecting (15') and (24'), we see that $\hat{H}_0 + LU$ can be represented as the kinetic energy operator $\hat{T}$ (now restoring band indices),

$$\hat{T} = [E_n(\mathbf{R}) - \tfrac{1}{2}\nabla_i \mu_n^{ij}(\mathbf{R})\nabla_j]\delta_{nn'} , \tag{25}$$

where

$$E_n(\mathbf{R}) = E_n + [U_{nn}^{0,0} L(\mathbf{R}) - U_{nn}^{0,i}\nabla_i L(\mathbf{R}) + \tfrac{1}{2} U_{nn}^{0,ij}\nabla_i \nabla_j L(\mathbf{R})] , \tag{26}$$

$$\mu_n^{ij}(\mathbf{R}) = \mu_n^{ij} - 2(U_{nn}^{0,ij} - U_{nn}^{i,j}) L(\mathbf{R}) . \tag{27}$$

We therefore arrive at the major conclusions of this section. (1) Since $U^{0,ij} - U^{i,j}$ is in general nonzero, the effective mass has a nontrivial $\mathbf{R}$ dependence. It is not surprising that the corrections involve $U^{0,ij}$ and U^{ij}, referring to the periodic functions g^i and h^{ij}, which summarize the physics of the band up to $O(k^2)$. In contrast, in the LK basis, which refers only to f but not explicitly to g^i and h^{ij}, the corresponding physics would be hidden in the canonical transformation $\hat{\theta}$. (2) The ordering of the operator is uniquely in the form $-\nabla\mu\nabla/2$, or in terms of the von Roos parameters $\alpha=\gamma=0$, $\beta=1$. Although this choice has been suggested[12] as one of the possible forms which is accurate to $O(q^0)$, we have, for the first time, *derived* it by a low-energy reduction accurate to $O(q^2)$; insofar as any ambiguity arises only in $O(q)$ and $O(q^2)$, only such a derivation can be meaningful. (3) However, there is a correction to the band bottom, the meaning of which becomes clearer in the next section. (4) The effective Hamiltonian in the Wannier basis is necessarily linear in L.

There is no difficulty in principle in carrying the formulation to higher orders in $\mathbf{k}$ and $\mathbf{q}$.

III. EXTERNAL POTENTIAL

In a single-electron theory, $V(\mathbf{r})$ in (4) includes the potential due to other carriers and in equilibrium must adjust itself so as to cancel the effect of the varying band edge and effective mass; e.g., see (33). As a consequence, $V(\mathbf{r})$ will vary on the same length scale as $L(\mathbf{r})$, and to be consistent it will be necessary to keep to $O(q^2)$ in $V(\mathbf{r})$ as well. The task is trivial, since we need only replace $L(\mathbf{r}) \rightarrow V(\mathbf{r})$ and $U(\mathbf{r}) \rightarrow \hat{I}$, where $\hat{I}$ is the identity operator in the derivation of the preceding section. Thus in analogy to (26), we find that $V(\mathbf{r})$ in the nth band should be replaced by

$$V_n^{\text{eff}}(\mathbf{R}) = I_{nn}^{0,0} V(\mathbf{R}) - I_{nn}^{0,i}\nabla_i V(\mathbf{R}) + \tfrac{1}{2} I_{nn}^{0,ij}\nabla_i \nabla_j V(\mathbf{R}) , \tag{28}$$

where the notation for the matrix elements of $\hat{I}$ follows (21). Similarly, one generates an additional kinetic term analogous to (27):

$$(I_{nn}^{0,ij} - I_{nn}^{i,j})\nabla_i L(\mathbf{R})\nabla_j . \tag{29}$$

It can be shown (Appendix B) that (29) vanishes identically, which is expected since an external potential should not generate a kinetic energy term. Incidentally this provides indirect support for the particular form of the coefficients of $L(\mathbf{R})$ in (27). Moreover, symmetry of the crystal will eliminate the second term in (28) and it can be shown that (Appendix B)

$$I_{nn}^{0,0} = 1 ,$$

$$I_{nn}^{i,j} = \langle f_n | (r - \bar{r})_i (r - \bar{r})_j | f_n \rangle \equiv \langle \delta r_i \delta r_j \rangle_n , \tag{30}$$

where $\bar{r}_i = \langle f_n | r_i | f_n \rangle$. Thus,

$$V_n^{\text{eff}}(\mathbf{R}) = V(\mathbf{R}) + \tfrac{1}{2}\langle \delta r_i \delta r_j \rangle_n \nabla_i \nabla_j V(\mathbf{R}) = N \int_c d^3 r\, f_n(\mathbf{r})^* V(\mathbf{R}) f_n(\mathbf{R}) , \tag{31}$$

i.e., the average potential seen by a particle in the bottom of the band, as is entirely sensible.

IV. GALILEAN INVARIANCE AND CLASSICAL CORRESPONDENCE

The claim[12] that a position-dependent $m(\mathbf{R})$ violates Bargmann's theorem[13] based on Galilean invariance has been disputed and analyzed at length,[14] so it suffices to emphasize just one trivial, but nevertheless, crucial point. It is indeed true that there is a conflict between Galilean invariance and a position-dependent $m(\mathbf{R})$, but this conflict is not to be resolved in favor of Galilean invariance. On the contrary, the position dependence of $m(\mathbf{R})$, mathematically derived from an accepted Hamiltonian (4), is not in doubt. The conflict simply means that there is no Galilean invariance, which is only to be expected if, for example, the material composition has a spatial gradient. Even for a homogeneous crystal, Galilean invariance is only an approximate concept for wave functions varying slowly on the lattice scale a, and is a *result* derived from the effective-mass approximation rather than an *a priori* condition.

We expect classical correspondence in the regime $q \ll k \ll a^{-1}$, i.e., when the parameters in the effective Hamiltonian, say $L(\mathbf{R})$, are nearly constant over one wavelength. To simplify the discussion, we restrict to

one band, drop the band index, assume isotropy, and write $V_n^{\rm eff}(\mathbf{R})$ simply as $V(\mathbf{R})$. Then the effective Hamiltonian is, from (25),

$$\hat{H}^{\rm eff}=E(\mathbf{R})-\tfrac{1}{2}\nabla\mu(\mathbf{R})\nabla+V(\mathbf{R}) \tag{32a}$$

with the classical analog

$$H^{\rm eff}=P^2/[2m(\mathbf{R})]+[E(\mathbf{R})+V(\mathbf{R})]\ , \tag{32b}$$

where $\mathbf{P}$, corresponding to $-i\nabla$, is the conjugate momentum. In the classical limit, the ordering of the $\nabla\mu\nabla$ in (32) is, of course, unimportant. The equations of motion are

$$d\mathbf{R}/dt=\partial H^{\rm eff}/\partial\mathbf{P}=\mathbf{P}/m(\mathbf{R})\ , \tag{33a}$$

$$\begin{aligned} d\mathbf{P}/dt&=-\partial H^{\rm eff}/\partial\mathbf{R}\\ &=-\nabla[E(\mathbf{R})+V(\mathbf{R})]+[P^2/2m(\mathbf{R})^2]\nabla m(\mathbf{R})\ . \end{aligned} \tag{33b}$$

The last term in (33b), quadratic in the velocity, has been interpreted as a dissipative force.[6] This cannot be correct, since the Hamiltonian ensures preservation of phase volume. We can write (33b) in a more transparent form. In the case where $\partial V(\mathbf{R})/\partial t=0$, $H^{\rm eff}$ is a constant of the motion, say with value E. Then (33b) becomes

$$\begin{aligned} d\mathbf{P}/dt&=\nabla[E-E(\mathbf{R})-V(\mathbf{R})]\\ &\quad+[E-E(\mathbf{R})-V(\mathbf{R})][\nabla m(\mathbf{R})/m(\mathbf{R})]\\ &=\frac{1}{m(\mathbf{R})}\nabla\{m(\mathbf{R})[E-E(\mathbf{R})-V(\mathbf{R})]\}\ . \end{aligned} \tag{33b'}$$

Note the necessity of ordering $m(\mathbf{R})$ and ∇ carefully in order to obtain such a better form.

We can also seek a WKB solution to the quantum equation $\hat{H}^{\rm eff}F(\mathbf{R})=EF(\mathbf{R})$ in the one-dimensional case, through the standard ansatz $F(R)=\exp iS(R)$; we find

$$F(R)=[m(R)/P(R)]^{1/2}\exp\left[i\int^R dR'P(R')\right]\ , \tag{34}$$

where

$$P(R)=\{2m(R)[E-E(R)-V(R)]\}^{1/2} \tag{35}$$

is the classical momentum and the prefactor in (34) indicates that the time spent in any ΔR is inversely proportional to the velocity $P(R)/m(R)$. The presence of the *non*constant factor $m(R)$ in the prefactor is crucial for consistent interpretation.

Even in the classical limit, terms of *first* order in ∇m must still be retained, because its cumulative effect over a large distance $m(\mathbf{R}_1)-m(\mathbf{R}_2)\sim\nabla m\cdot(\mathbf{R}_1-\mathbf{R}_2)$ is not negligible. The suggestion that the ordering ambiguity is irrelevant for smooth variations[15] should therefore be viewed with caution.

V. OFF-DIAGONAL TERMS AND COMPARISON

Our conclusion in Sec. II is based on the neglect of off-diagonal terms. In view of our criticism below of the incorrect treatment of off-diagonal terms in the LK basis,[11] it is at least necessary to prove that the off-diagonal terms do not affect our conclusion that $m(\mathbf{R})$ is, in general, $\mathbf{R}$ dependent. In this section, the external potential V will be ignored, and it suffices to consider small $L(\mathbf{r})$.

Let us write the effective Hamiltonian as

$$\hat{H}^{\rm eff}=\lambda_0\hat{h}_0+\lambda_1\hat{h}_1+\lambda_2\hat{h}_2\ , \tag{36}$$

where $\lambda_0,\lambda_1,\lambda_2$ are formal parameters, and $\hat{h}_0$ are terms obtained from $\hat{H}_0$, as in (15'), $\hat{h}_1$ are diagonal terms obtained from LU, as in (24'), and $\hat{h}_2$ are off-diagonal terms obtained from LU, so far neglected. The nature of these terms, i.e., whether diagonal or off-diagonal and the dependence of L, is summarized in Table I.

Since L is regarded as small, the λ_2 term, neglected so far, can be dealt with perturbatively. However, because this term is off-diagonal, it affects the eigenvalues, i.e., the dispersion $E_n(\mathbf{k})$, only in second order. Thus to $O(L^1)$, there is no correction to the conclusion of Sec. II. For this reason the single band formalism is often adequate. In any case, this proves that at least generically $m(\mathbf{R})$ is truly position dependent, which is hardly surprising in view of the discussion at the beginning of Sec. II. Note, for comparison below, that this happy state of affairs would not be obtained if there is an off-diagonal term of $O(L^0)$.

Now we are in a position to examine critically a claim[11] based on a calculation in the LK basis that $m(\mathbf{R})$ has no position dependence. Apart from the general argument given at the beginning of Sec. II, it is necessary to point out the mathematical steps in that claim[11] which are questionable. In analogy to (36), we can write the Hamiltonian in this case as

$$\hat{H}^{\rm eff}=\lambda_0\hat{h}'_0+\lambda_1\hat{h}'_1+\lambda_2\hat{h}'_2+\lambda_3\hat{h}'_3\ . \tag{37}$$

(In the notation of Ref. 11, $\lambda_0=1$; $\hat{h}'_0\to H^0$, $\hat{h}'_1\to H^1$, $\hat{h}'_2\to H^2$, and $\hat{h}'_3\to V^1$.) The nature of the various terms are summarized in Table II; note, however, that now diagonal and off-diagonal refer to the LK basis. The λ_0 and λ_1 terms, independent of L, come from the homogeneous Hamiltonian.

Now the only small parameter is L, and the limiting homogeneous Hamiltonian, regarded as generic, contains no small parameters. To assume otherwise is to violate the all important requirement that the $L\to0$ limit must be recovered exactly, as emphasized earlier. In particular, λ_1 *cannot* be regarded as *small*. So in diagonalizing (37), all terms of order $\lambda_1^N\lambda_2$ and $\lambda_1^N\lambda_3$, *any* N, must be kept. In Ref. 11, only terms proportional to λ_1^2, $\lambda_1\lambda_2$, and

TABLE I. The nature of terms associated with the three coefficients in the present approach: diagonal (D) vs off-diagonal (OD) in the Wannier basis, and dependence on L.

	D or OD	L^n
λ_0	D	L^0
λ_1	D	L^1
λ_2	OD	L^1

TABLE II. The nature of terms associated with the four coefficients in Ref. 11: diagonal (D) vs off-diagonal (OD) in the LK basis, and dependence on L.

	D or OD	L^n
λ_0	D	L^0
λ_1	OD	L^0
λ_2	OD	L^1
λ_3	D	L^1

$\lambda_1\lambda_3$ are kept; in particular, $\lambda_1^2\lambda_2$ and $\lambda_1^2\lambda_3$ are neglected. It is precisely such neglected terms which will generate the position-dependent mass.

The systematic way to proceed is first to remove all off-diagonal terms of order λ_1^N, and secondly to remove the remaining off-diagonal terms to first order. The first step, apparently tedious, merely transforms to a basis in which the homogeneous Hamiltonian $\lambda_0\hat{h}'_0+\lambda_1\hat{h}'_1$ is exactly diagonalized—and this is nothing but the Bloch-Wannier basis. The proposed procedure is then equivalent to our formalism in Sec. II.

By examining Table II, it would appear that diagonal terms of $O(L^1)$ should appear among the contributions retained in Ref. 11, namely λ_3 and $\lambda_1\lambda_2$ (note that two off-diagonal terms multiplied contain a diagonal term). However, the momentum vector $\mathbf{k}$ appears only in the λ_1 term and only linearly, so the λ_3 and $\lambda_1\lambda_2$ terms, respectively $O(k^0)$ and $O(k^1)$, do not contribute to the effective mass. This shows why the $\lambda_1^2\lambda_2$ term, going like k^2L^1, must be retained in order to discuss the $\mathbf{R}$ dependence of $m(\mathbf{R})$.

In one sense our result and the claim of Ref. 11 are not in conflict. The claim in Ref. 11 would be entirely correct if phrased as follows: for those *special* cases in which $\lambda_1\hat{h}'_1$ can be regarded as small (in the sense that its square is negligible), $m(\mathbf{R})$ then has negligible position dependence. Our result is, for *general* cases, $m(\mathbf{R})$ has a position dependence. It may in fact be the case that many solids fall into the "special" category, but this would have to be justified on a case-by-case basis and not assumed as a matter of formalism.

Most works in the literature which address the question of effective mass are not accurate to $O(q^2)$; the one exception is Ref. 5, which obtained essentially the correct answer in the LK basis. In order to compare the results for the effective mass, we have to express the matrix elements $U_{nn}^{0,ij}$ and $U_{nn}^{i,j}$ involving f_n, g_n^i, and h_n^{ij} in terms of the LK basis functions $\varphi_{n\mathbf{k}}=e^{i\mathbf{k}\cdot\mathbf{r}}f_n$ by eliminating g_n^i and h_n^{ij} in favor of f_n. This is done in Appendix B and we find that

$$|g_n^i\rangle=\frac{i}{m_0}\sum_{n'\,(\neq n)}|f_{n'}\rangle\frac{p_{n'n}^i}{E_{n'}-E_n}\ , \tag{38}$$

where

$$p_{n'n}^i=\langle f_{n'}|\hat{p}^{\,i}|f_n\rangle,\quad \hat{p}^{\,i}=-i\nabla_i\ . \tag{39}$$

The second-order function h_n^{ij} is best separated into a part proportional to f_n and a part orthogonal to f_n:

$$|h_n^{ij}\rangle=\beta_n^{ij}|f_n\rangle+|\bar{h}_n^{\,ij}\rangle\ , \tag{40}$$

where

$$\beta_n^{ij}=\frac{1}{m_0^2}\sum_{n'\,(\neq n)}\frac{p_{nn'}^i p_{n'n}^j}{(E_{n'}-E_n)^2} \tag{41}$$

and

$$|\bar{h}_n^{\,ij}\rangle=-\frac{1}{m_0^2}\sum_{\substack{n'\,(\neq n),\\ n''\,(\neq n)}}|f_{n'}\rangle\frac{p_{n'n''}^i p_{n''n}^j+(i\leftrightarrow j)}{(E_{n'}-E_n)(E_{n''}-E_n)}\ . \tag{42}$$

From these we find that

$$\begin{aligned}\mu_n^{ij}(\mathbf{R})-\mu_n^{ij}&=-2L(\mathbf{R})(U_{nn}^{0,ij}-U_{nn}^{i,j})\\&=-2L(\mathbf{R})(A+B-C)\ ,\end{aligned} \tag{43}$$

where A, B, and C are given in Appendix B and come from the matrix elements $\langle f_n|U|\beta_n^{ij}f_n\rangle$, $\langle f_n|U|\bar{h}_n^{\,ij}\rangle$, and $\langle g_n^i|U|g_n^j\rangle$, respectively. This result is nearly the same as that given in Ref. 5, which is

$$\mu_n^{ij}(\mathbf{R})-\mu_n^{ij}=-2L(\mathbf{R})(\tfrac{1}{2}A+\tfrac{1}{2}B-C)\ . \tag{43'}$$

It is not difficult to see that (43′) cannot be correct. Consider the special case $U=\hat{I}$ and $L(\mathbf{R})=$const, representing an additive constant in the microscopic Hamiltonian. Obviously the effective mass should not be modified in this case. From the expressions given in Appendix B, it is readily verified that in this limit $A=C$ and $B=0$, so that (43) is consistent but (43′) is not.

Apart from this minor disagreement in the numerical coefficient in (43), the result in Ref. 5 is correct. It is unfortunate that the greater complication of the LK basis has caused Ref. 5 to be misunderstood and criticized.[11] The much simpler expressions in the Wannier basis [e.g., (27)] should help to clarify the issues.

VI. CONCLUSION

We have systematically extracted the effective low-energy Hamiltonian for a crystal with a slowly varying inhomogeneity, with use of a calculation accurate to second order in the ∇ operator. A number of controversies are then resolved. (1) The effective mass is in general dependent on the position. (2) The ordering of operators in the kinetic energy is $-\nabla\mu\nabla/2$. (3) Galilean invariance and the Bargmann theorem do not apply. (4) The classical limit does not show dissipation.

From a technical point of view, the LK basis is shown to be inconvenient for the physical reason that it does not incorporate the homogeneous limit exactly (unless one compensates by a complicated canonical transformation). In the Bloch-Wannier basis, the neglect of interband transitions is justified if the inhomogeneity is weak ($|L|\ll 1$).

Note added in proof. The same conclusion on the ordering of operators in the kinetic energy has recently been obtained from another point of view.[18]

ACKNOWLEDGMENTS

I thank P. T. Leung for a careful reading of the manuscript and S. P. Wong for discussions.

APPENDIX A

Dropping the band index and changing the dummy variables $\mathbf{R}'\to\mathbf{R}-\mathbf{R}'$, $\mathbf{p}\to\mathbf{p}-\mathbf{q}$, we have from (18)

$$\sum_{\mathbf{R}'}\langle\mathbf{R}|LU|\mathbf{R}'\rangle F(\mathbf{R}') = N^{-1}\sum_{\mathbf{R},\mathbf{p},\mathbf{q}}\tilde{L}(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{R}}e^{i\mathbf{p}\cdot\mathbf{R}'}S(\mathbf{p}-\mathbf{q},\mathbf{p})F(\mathbf{R}-\mathbf{R}') . \tag{A1}$$

It then follows that $\mathbf{q}\to i\nabla_L$, where the notation means that ∇ operates only on the $\mathbf{R}$ variable in L, but not on F. Likewise $\mathbf{p}\to -i\nabla'$ operating on $e^{i\mathbf{p}\cdot\mathbf{R}'}$, which turns into $i\nabla'_F$ by an integration by parts, and hence into $-i\nabla_F$. Thus we have

$$S(-i\nabla_F-i\nabla_L,-i\nabla_F)\left[\left[\sum_{\mathbf{q}}\tilde{L}(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{R}}\right]\left[N^{-1}\sum_{\mathbf{R}',\mathbf{p}'}e^{i\mathbf{p}\cdot\mathbf{R}'}F(\mathbf{R}-\mathbf{R}')\right]\right]=S(-i\nabla_F-i\nabla_L,-i\nabla_F)[L(\mathbf{R})F(\mathbf{R})] , \tag{A2}$$

and (24) follows by trivial algebra. Incidentally (A2) is exact, and can be expanded to higher order if desired.

APPENDIX B

The orthonormality of the Bloch functions gives

$$\begin{aligned}\delta_{nn'}&=\int d^3r(f_n+ik_ig_n^i-\tfrac{1}{2}k_ik_jh_n^{ij}+\cdots)^*(f_{n'}+ik_jg_{n'}^j-\tfrac{1}{2}k_ik_jh_{n'}^{ij}+\cdots)\\&=I_{nn'}^{0,0}+ik_i(I_{nn'}^{0,i}-I_{nn'}^{i,0})-\tfrac{1}{2}k_ik_j(I_{nn'}^{0,ij}+I_{nn'}^{ij,0}-2I_{nn'}^{i,j})+\cdots ,\end{aligned} \tag{B1}$$

from which we can conclude that for $n=n'$

$$I_{nn}^{0,0}=1, \text{ i.e., } \langle f_n|f_n\rangle=1 , \tag{B2}$$

$$I_{nn}^{0,ij}=I_{nn}^{i,j}, \text{ i.e., } \langle f_n|h_n^{ij}\rangle=\langle g_n^i|g_n^j\rangle , \tag{B3}$$

showing that (29) vanishes identically.

To express matrix elements involving g_n^i and h_n^{ij} in terms of f_n, we first choose a phase convention such that $I_{nn}^{0,i}=0$, i.e.,

$$\langle f_n|g_n^i\rangle=0 . \tag{B4}$$

Now from the Schrödinger equation

$$\hat{H}_0e^{i\mathbf{k}\cdot\mathbf{r}}(f_n+ik_ig_n^i-\tfrac{1}{2}k_ik_jh_n^{ij}+\cdots)=(E_n+\tfrac{1}{2}\mu_n^{ij}k_ik_j)e^{i\mathbf{k}\cdot\mathbf{r}}(f_n+ik_ig_n^i-\tfrac{1}{2}k_ik_jh_n^{ij}+\cdots) , \tag{B5}$$

we obtain

$$(\hat{H}_0-E_n)g_n^i=(i/m_0)\hat{p}^{\,i}f_n \tag{B6}$$

$$(\hat{H}_0-E_n)h_n^{ij}=(i/m_0)(\hat{p}^{\,i}g_n^j+\hat{p}^{\,j}g_n^i)+(\delta^{ij}/m_0-\mu_n^{ij})f_n . \tag{B7}$$

By projecting (B6) on $f_{n'}$ $(n'\neq n)$, the part of g_n^i orthogonal to f_n is determined; by convention there is no component parallel to f_n. Hence we obtain (38). From this

$$I_{nn}^{i,j}=\frac{1}{m_0^2}\sum_{n'(\neq n)}\frac{p_{nn'}^ip_{n'n}^j}{(E_{n'}-E_n)^2} \tag{B8}$$

and, using $\hat{p}^{\,i}/m_0=i[\hat{H}_0,r_i]$, (B8) can be written as (30).

Now h_n^{ij} can be decomposed into components parallel and perpendicular to f_n as in (40). The parallel component is determined from (B3),

$$\beta_n^{ij}=\langle f_n|h_n^{ij}\rangle=\langle g_n^i|g_n^j\rangle=I_{nn}^{i,j} , \tag{B9}$$

while the orthogonal component is determined by projecting (B7) onto $f_{n'}$ $(n'\neq n)$, giving (42).

The relevant matrix elements of U can then be expressed via

$$A\equiv\langle f_n|U|\beta_n^{ij}f_n\rangle=\frac{1}{m_0^2}\sum_{n'(\neq n)}\frac{p_{nn'}^ip_{n'n}^j}{(E_{n'}-E_n)^2}U_{nn}^{0,0} , \tag{B10}$$

$$B\equiv\langle f_n|U|\bar{h}_n^{\,ij}\rangle=-\frac{1}{m_0^2}\sum_{\substack{n'(\neq n),\\ n''(\neq n)}}\frac{U_{nn}^{0,0}p_{n'n''}^ip_{n''n}^j+(i\leftrightarrow j)}{(E_{n'}-E_n)(E_{n''}-E_n)} , \tag{B11}$$

$$C\equiv\langle g_n^i|U|g_n^j\rangle=\frac{1}{m_0^2}\sum_{\substack{n'(\neq n),\\ n''(\neq n)}}\frac{p_{nn''}^iU_{n''n}^{0,0}p_{n'n}^j}{(E_{n'}-E_n)(E_{n''}-E_n)} . \tag{B12}$$

[1]W. A. Harrison, Phys. Rev. **110**, 14 (1958).
[2]K. Kawamura, Z. Phys. B **29**, 101 (1978).
[3]T. Teichler, Phys. Lett. **87A**, 113 (1981).
[4]T. Gora and F. Williams, Phys. Rev. **177**, 1179 (1969).
[5]L. Leibler, Phys. Rev. B **12**, 4443 (1975).
[6]C. M. van Vliet and A. H. Marshak, Phys. Rev. B **26**, 6734 (1982).
[7]Q. G. Zhu and H. Kroeman, Phys. Rev. B **27**, 3519 (1983).
[8]R. A. Morrow and K. R. Brownstein, Phys. Rev. B **30**, 678 (1984).
[9]P. R. Emtage, J. Appl. Phys. **33**, 1950 (1962); L. J. van Ruyven and F. E. Williams, Am. J. Phys. **35**, 705 (1967).
[10]R. Hill and F. Williams, J. Appl. Phys. **39**, 3344 (1968); Indradev, L. J. van Ruyven, and F. Williams, *ibid.* **39**, 3344 (1968); R. B. Laur and F. Williams, *ibid.* **42**, 2904 (1971); S. K. Chattopadhyaya and V. K. Mathur, Phys. Rev. B **3**, 3390 (1971).
[11]O. von Roos and H. Mavromatis, Phys. Rev. B **31**, 2294 (1985).
[12]O. von Roos, Phys. Rev. B **27**, 7547 (1983). Our parameters differ from those commonly used by a minus sign.
[13]V. Bargmann, Ann. Math. **59**, 1 (1954).
[14]K. Kawamura and R. A. Brown, Phys. Rev. B **37**, 3932 (1988).
[15]R. A. Morrow, Phys. Rev. B **29**, 5962 (1984).
[16]J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[17]The assumption of periodic boundary conditions does not create any difficulty, since the sample can be imagined to be as large as one wishes. Even for a linearly graded solid, e.g., $L(x,y,z)=x/D$, $0\leq x\leq D$, where D is the sample size, we can adjoin an "image sample" in the range $D\leq x\leq 2D$ with $L(x,y,z)=2-x/D$, thus recovering periodicity over the whole "sample" $0\leq x\leq 2D$.
[18]G. T. Einevoll and P. C. Hemmer, J. Phys. C **21**, L1193 (1988).