PHYSICAL REVIEW B VOLUME 39, NUMBER 17 15 JUNE 1989-I

Operator ordering in effective-mass theory

J. Thomsen, G. T. Einevoll, and P. C. Hemmer
Institutt for Fysikk, Norges Tekniske Høgskole (NTH), Universitetet i Trondheim,
N-7034 Trondheim, Norway

(Received 19 December 1988)

We study the problem of operator ordering in the effective-mass Hamiltonian with kinetic energy operator $\frac{1}{4}(m^{\alpha}pm^{\beta}pm^{\gamma}+m^{\gamma}pm^{\beta}pm^{\alpha})$, for a position-dependent effective mass m. Here $\alpha+\beta+\gamma=-1$. We use merely the inherent criterion that the eigenvalues of the Hamiltonian should correspond to finite and uniquely determined energies. Through exact model calculations we show first that divergences occur unless $\alpha=\gamma$, and second that $\alpha=\gamma=0$, $\beta=-1$ is the only universal set of operator ordering parameters which gives unique results.

I. INTRODUCTION

The effective-mass approximation[1] is an extremely useful method in semiconductor physics, by means of which dynamic and static properties of charge carriers can be studied free of complexities due to the lattice potential of the material. The effective-mass treatment rests on a firm basis for homogeneous semiconductors.[2] It becomes in fact asymptotically exact when the perturbations of the periodic Hamiltonian, for instance in the form of a shallow impurity potential, become sufficiently gentle.

In its simplest form the effective-mass wave function (envelope function) for the conduction band, say, obeys the Schrödinger equation with a Hamiltonian

$$H=-\frac{\hbar^2}{2m}\nabla^2+E_c+U(\mathbf{r})\ . \tag{1}$$

Here m is the effective mass, E_c the (simple) minimum of the energy dispersion $E(\mathbf{k})$, and $U(\mathbf{r})$ is the perturbing potential. m is considered scalar since complications arising from tensorial effective masses are not germane to the issue to be discussed here.

The effective-mass approach has also been used extensively as a computational method to deal with *nonuniform* crystals, in particular because of the increasing attention paid to man-made heterostructures like quantum wells and superlattices.[3] The molecular-beam epitaxy technique makes it possible to produce abrupt interfaces between materials and in these applications the effective-mass parameters become position dependent. The most straightforward generalization of Eq. (1) uses the now position-dependent conduction-band edge $E_c(\mathbf{r})$ in the potential-energy part of the Hamiltonian. Since, however, a position-dependent effective mass $m(\mathbf{r})$ and the momentum operator do not commute, the question arises of the correct form for the kinetic energy operator.[4,5]

The basic requirement is, of course, that the Hamiltonian is Hermitian. The operator ordering[6]

$$H=-\tfrac{1}{4}\hbar^2(m^{\alpha}\nabla m^{\beta}\nabla m^{\gamma}+m^{\gamma}\nabla m^{\beta}\nabla m^{\alpha})+E_c(\mathbf{r})+U(\mathbf{r})\ , \tag{2}$$

with

$$\alpha+\beta+\gamma=-1\ , \tag{3}$$

secures Hermiticity, and includes special cases that have appeared in the literature,[4] viz.,

$$\tfrac{1}{2}m^{-1}\nabla^2+\tfrac{1}{2}\nabla^2 m^{-1},\quad m^{-1/2}\nabla^2 m^{-1/2},\quad \nabla m^{-1}\nabla \tag{4}$$

for the kinetic energy (omitting the constant $-\hbar^2/2$).

The question of the actual values of α, β, and γ remains, and the present article addresses this issue in a limited sense. We do *not* discuss the relation of (2) to a more complete treatment of the underlying crystal structure, nor the question of whether an effective-mass description of the simple type (2) can be justified at all. Our objective here is different. We stay completely within the framework of effective-mass theory, i.e., we accept Eq. (2) as our starting point. The point of interest is possible *inherent limitations* on the values of α, β, and γ. It is perhaps surprising that there are any limitations on the operator ordering beyond Hermiticity. As we shall see, however, unacceptable physical consequences occur unless specific choices are made.

II. EIGENVALUE PROBLEM

We will in Sec. III study the eigenvalue problem

$$H\psi=E\psi\ . \tag{5}$$

For that purpose reformulations of the operator (2) will be useful. An obvious possibility is to let the position-dependence mass dependence take the form of an effective-potential term. Introducing

$$\phi=m^{-1/2}\psi \tag{6}$$

and carrying out all differentiations in Eq. (5) we end up with

$$-\nabla^2\phi=\{2\hbar^{-2}[E-V(\mathbf{r})]m(\mathbf{r})-\beta\nabla^2 l(\mathbf{r})+[(\gamma-\alpha)^2-\beta^2](\nabla l)^2\}\phi\ . \tag{7}$$

Here

$$V(\mathbf{r})=E_c(\mathbf{r})+U(\mathbf{r})$$

and

$$l(\mathbf{r})=\tfrac{1}{2}\ln m(\mathbf{r})\ . \tag{8}$$

From the form (7) it is seen that applications to abrupt heterojunctions, i.e, to discontinuous effective-mass variations, apparently result in two types of singularities in the effective potential: derivatives of δ functions and squares of δ functions, both ominous signs. This observation is superficial, however, since we will see later on that because of cancellation effects it may not *necessarily* be disastrous to have both types of singularities present simultaneously.

A second useful reformulation employs the wave function

$$\chi=m^{\alpha/2+\gamma/2}\ . \tag{9}$$

In this case the new wave function satisfies

$$-\frac{\hbar^2}{2m(r)}[\nabla^2\chi+2\beta(\nabla l)\nabla\chi+(\gamma-\alpha)^2(\nabla l)^2\chi]+V(\mathbf{r})\chi=E\chi\ ,$$

or

$$\hat{H}\chi=E\chi\ . \tag{10}$$

This is the Euler-Lagrange equation for the variational functional

$$R(\chi)=\frac{\int d\mathbf{r}\, m^{\beta+1}\chi^*\hat{H}\chi}{\int d\mathbf{r}\, m^{\beta+1}\chi^*\chi} \tag{11}$$

to be stationary. Inserting the explicit form of $\hat{H}$, we obtain, after a partial integration,

$$R(\chi)=\frac{\frac{1}{2}\hbar^2\int d\mathbf{r}\, m^{\beta}[\,|\nabla\chi|^2-(\gamma-\alpha)^2(\nabla l)^2|\chi|^2+2\hbar^{-2}m(\mathbf{r})V(\mathbf{r})|\chi|^2]}{\int d\mathbf{r}\, m^{\beta+1}|\chi|^2}\ . \tag{12}$$

In particular, we will make use of the Rayleigh-Ritz inequality

$$E_0\le R(\chi) \tag{13}$$

for the ground-state energy.

III. NECESSITY OF $\alpha=\gamma$

As a test case we solve the simplest possible eigenvalue problem, a particle in a one-dimensional box whose interior consists of two materials with effective masses m_1 and m_2, respectively (Fig. 1). For definiteness we assume $m_2>m_1$. This could represent transverse motion in a heterostructure with no conduction-band offset between material 1 and material 2, and with a third surrounding material that the electron cannot penetrate. The physical relevance of the model is, however, not crucial for the conclusions we draw.

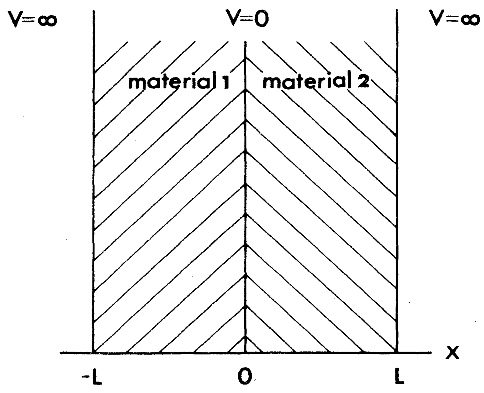

FIG. 1. An abrupt heterostructure in a box. The effective masses are m_1 and m_2, respectively.

In order to tame the singularities we consider the discontinuous effective-mass variation above as the limit of the continuous mass-dependence (Fig. 2),

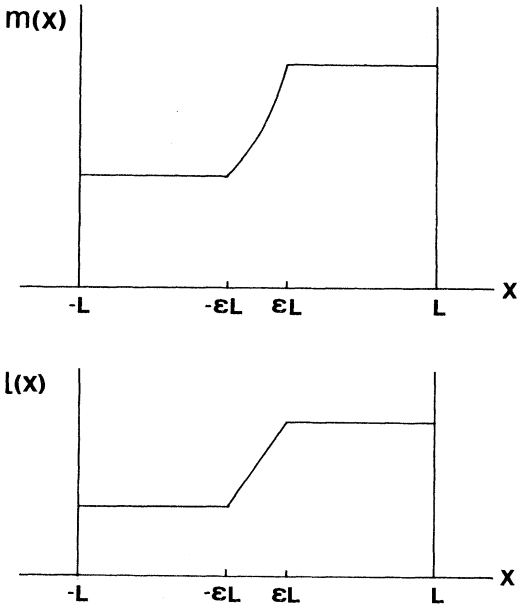

FIG. 2. A continuous effective-mass variation.

$$m(x)=\begin{cases} m_1 \text{ for } -L<x<-\epsilon L \ , & \text{(14a)} \\ (m_1m_2)^{1/2}(m_2/m_1)^{x/2\epsilon L} \text{ for } |x|<\epsilon L \ , & \text{(14b)} \\ m_2 \text{ for } \epsilon L<x<L \ , & \text{(14c)} \end{cases}$$

when $\epsilon\rightarrow 0$. This choice secures for finite ϵ a piecewise *linear* variation of the function $l(x)$, Eq. (8), that enters the effective potential.

With the notation

$$2\hbar^{-2}E=\hat{E} \ , \qquad (15)$$

Eq. (7) now takes the form

$$\phi''=\begin{cases} -\hat{E}m_1\phi \text{ for } -L<x<-\epsilon L \ , & \text{(16a)} \\ [-\hat{E}m(x)+\Delta]\phi \text{ for } -\epsilon L<x<\epsilon L \ , & \text{(16b)} \\ -\hat{E}m_2\phi \text{ for } \epsilon L<x<L \ . & \text{(16c)} \end{cases}$$

Here Δ is a *constant* potential,

$$\Delta=(\nu/4\epsilon L)^2\ln^2(m_2/m_1) \ , \qquad (17)$$

with the abbreviation

$$\nu^2\equiv\beta^2-(\gamma-\alpha)^2=(1+2\alpha)(1+2\gamma) \ . \qquad (18)$$

The boundary and matching conditions are

$$\phi(-L)=\phi(L)=0 \ , \qquad (19)$$

continuity of ϕ at $x=\pm\epsilon L$, plus definite discontinuities $\Delta\phi'$ in the slope ϕ' at $x=\pm\epsilon L$. The discontinuities in the derivative occur because the $\nabla^2 l$ term in (7) produces δ functions at $x=\pm\epsilon L$. Integrating the one-dimensional version of Eq. (7) across $x=\pm\epsilon L$ we obtain

$$[\Delta\phi']_{x=\pm\epsilon L}=\mp(\beta/4\epsilon L)\ln(m_2/m_1)\phi(\pm\epsilon L) \ , \qquad (20)$$

For $-L<x<-\epsilon L$ the relevant solution of (16a) is

$$\phi(x)=A\sin[(m_1\hat{E})^{1/2}(x+L)] \ , \qquad (21)$$

while for $\epsilon L<x<L$ we have

$$\phi(x)=B\sin[(m_2\hat{E})^{1/2}(x-L)] \ . \qquad (22)$$

In the narrow intermediate range the solution is less elementary. By introduction of the new variable

$$z=4\epsilon L(m_1m_2)^{1/4}\hat{E}^{1/2}(m_2/m_1)^{x/4\epsilon L}/\ln(m_2/m_1) \ . \qquad (23)$$

Eq. (16b) takes the form

$$z^2\frac{d^2\phi}{dz^2}+z\frac{d\phi}{dz}+(z^2-\nu^2)\phi=0 \ , \qquad (24)$$

Bessel's equation. Hence[7]

$$\phi(x)=aJ_\nu(z)+bY_\nu(z) \qquad (25)$$

in this range.

It remains to splice the wave function at $x=\pm\epsilon L$. Continuity of ϕ together with the slope-discontinuity requirement (20) yields, at $x=-\epsilon L$,

$$\frac{aJ'_\nu(z_-)+bY'_\nu(z_-)}{aJ_\nu(z_-)+bY_\nu(z_-)}cz_- =(m_1\hat{E})^{1/2}\cot[L(m_1\hat{E})^{1/2}(1-\epsilon)]+\beta c \ , \qquad (26)$$

and, at $x=\epsilon L$,

$$-\frac{aJ'_\nu(z_+)+bY'_\nu(z_+)}{aJ_\nu(z_+)+bY_\nu(z_+)}cz_+ =(m_2\hat{E})^{1/2}\cot[L(m_2\hat{E})^{1/2}(1-\epsilon)]-\beta c \ . \qquad (27)$$

Here

$$c=(1/4\epsilon L)\ln(m_2/m_1) \qquad (28)$$

and

$$z_\pm=z(\pm\epsilon L)=(m_1m_2)^{1/4}\hat{E}^{1/2}c^{-1}(m_2/m_1)^{\pm 1/4} \ . \qquad (29)$$

The ratio b/a is determined both by (26) and by (27). Equating these two expressions for b/a, we obtain the final equation to determine the energy eigenvalues $\hat{E}$.

First, we solve this eigenvalue equation numerically for the ground state $\hat{E}_0$ as function of ϵ, for some values of α,β,γ (Fig. 3). The crucial feature that emerges is that the eigenvalue becomes negative for ϵ less than a critical width parameter ϵ_0, except when $\alpha=\gamma$.

The width ϵ_0 at which the ground-state energy goes negative can be found exactly. This is useful as a check on the numerics, but we refrain from giving the explicit formula. We show, however, in Fig. 4 the variation of this critical width with $|\alpha-\gamma|$ (for fixed β). The influence of the mass ratio m_2/m_1 is also apparent in the figure.

In the abrupt limit $\epsilon\rightarrow 0$ the ground-state energy does not merely become negative when $\alpha\neq\gamma$, it also *diverges*. This can be shown rigorously by means of the Rayleigh-Ritz inequality (13).

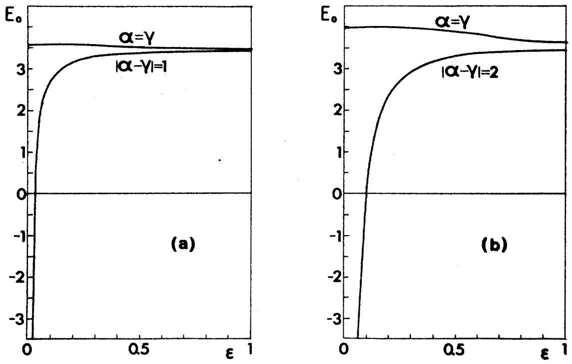

FIG. 3. The ground-state energy as function of the relative width ϵ of the healing layer for several values of α,β,γ. We have taken $m_2=2m_1$. Energies are measured in units of $\hbar^2/2m_2L^2$. (a) $\beta=-1$; (b) $\beta=-2$.

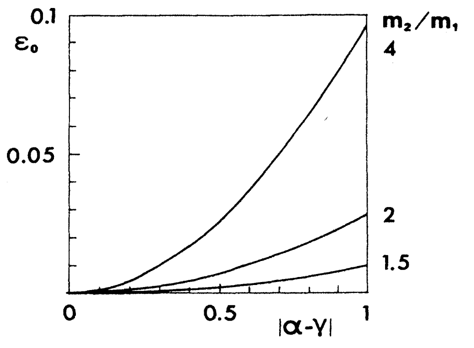

FIG. 4. The width ϵ_0 for which the ground-state energy vanishes, as a function of $|\alpha-\gamma|$, and for three different mass ratios m_2/m_1. The value $\beta=-1$ is used.

Since $m_1 \le m \le m_2$ we can form from (12) the weaker inequality

$$2E_0\hbar^{-2} \le \frac{1}{m_1}\left[\frac{m_2}{m_1}\right]^{|\beta|}\frac{\int_{-L}^{+L}(\chi')^2dx}{\int_{-L}^{+L}\chi^2dx} - \frac{(\gamma-\alpha)^2}{(4\epsilon L)^2}\frac{1}{m_2}\left[\frac{m_1}{m_2}\right]^{|\beta|}\ln^2(m_1/m_2) \times\frac{\int_{-\epsilon L}^{+\epsilon L}\chi^2dx}{\int_{-L}^{+L}\chi^2dx}\,, \tag{30}$$

for a χ that vanishes at $x=\pm L$. A convenient trial function is

$$\chi(x)=\begin{cases}0 \text{ for } |x|>\epsilon\lambda L\,,\\ \epsilon^2\lambda^2L^2-x^2 \text{ for } |x|<\epsilon\lambda L\,,\end{cases} \tag{31}$$

with a variational parameter λ in the range

$$1<\lambda<\epsilon^{-1}\,. \tag{32}$$

Insertion into (30) yields

$$E_0 \le \frac{5\hbar^2}{256m_1L^2\epsilon^2}[64\lambda^{-2}(m_2/m_1)^{|\beta|} -(\gamma-\alpha)^2\lambda^{-1}(m_1/m_2)^{|\beta|+1} \times\ln^2(m_2/m_1)]\,, \tag{33}$$

where we have used that

$$3-2\lambda^{-2}+\tfrac{3}{5}\lambda^{-4}>1\,.$$

For $\gamma=\alpha$ only the first term remains, and taking λ equal to its maximum value ϵ^{-1} we have

$$E_0 \le \frac{5\hbar^2}{4m_1L^2}(m_2/m_1)^{|\beta|}\,. \tag{34}$$

Since by Eq. (12) E_0 is positive in this case, this shows that for $\alpha=\gamma$ the ground-state energy is finite when $\epsilon\rightarrow 0$.

For $\gamma\neq\alpha$, however, we can always choose λ so large (and independent of ϵ) that the second term in (33), of $O(\lambda^{-1})$, dominates the first term, of $O(\lambda^{-2})$. We may presuppose that ϵ is chosen in advance so small that this λ is within the range (32). Hence the upper bound can be made negative and proportional to ϵ^{-2}. We conclude that for a heterostructure (i.e., $m_1\neq m_2$), and for $\alpha\neq\gamma$, the ground-state energy diverges towards minus infinity in the abrupt limit, a useless and unphysical result.

Morrow and Brownstein[4] consider the same class (2) of effective-mass Hamiltonians, and obtain also the result that α must equal γ. They conclude, however, that the junction acts as an impenetrable barrier, forcing the wave function to vanish at the boundary. The reality is, as we have just seen, rather different. The wave function is concentrated at the junction which acts as an extremely singular attractive well.

In conclusion, application of the effective-mass equation to abrupt heterostructures requires

$$\alpha=\gamma\,. \tag{35}$$

This still leaves the value of $\alpha=\gamma$, or, equivalently, of β, undetermined. In the next section we shall argue that overall consistency can only be had if α and γ vanish.

IV. ARGUMENT FOR $\alpha=\gamma=0$

With $\alpha=\gamma$ the stationary Schrödinger equation takes the form

$$-\tfrac{1}{2}\hbar^2m^{\alpha}\nabla m^{\beta}\nabla m^{\alpha}\psi+[E_c(\mathbf{r})+U(\mathbf{r})]\psi=E\psi\,. \tag{36}$$

As long as the potential U is everywhere finite, one sees immediately that unless[4]

$$m^{\alpha}\psi \text{ continuous},\quad m^{\beta}\nabla m^{\alpha}\psi \text{ continuous}\,, \tag{37}$$

the first term in (36) will contain stronger singularities than the other two terms.

We will now argue that consistency requires $\alpha=\gamma=0$, or, equivalently, $\beta=-1$. Again a simple test model will be used, in this case a one-dimensional δ well:

$$U(x)=-\mu\delta(x)\,. \tag{38}$$

It differs from the standard textbook example in that the well is surrounded by different materials, with effective masses m_1 and m_2, on the left-hand and the right-hand side, respectively [Fig. 5(a)]. For simplicity we assume zero band-edge discontinuity, $E_c=0$.

We approach the singular potential in a controlled manner by considering the δ well as the $\epsilon\rightarrow 0$ limit of the square-well potential in Fig. 5(b), viz.,

$$U(x)=\begin{cases}-\mu\epsilon^{-1}(a_1+a_2)^{-1} \text{ for } -a_1\epsilon<x<a_2\epsilon\,,\\ 0 \text{ otherwise}\,.\end{cases} \tag{39}$$

The result should be independent of the values of a_1 and a_2 as long as the $\epsilon\rightarrow 0$ limit is taken. The main advantage with the everywhere-finite potential (39) is that boundary conditions are unproblematic.

We seek bound states in the δ well. With the now piecewise constant potential the solution of the

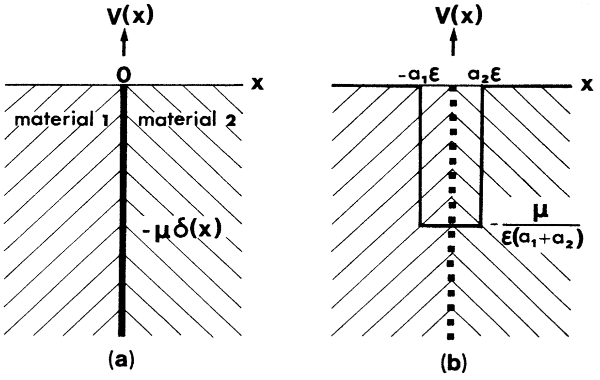

FIG. 5. (a) The δ well. (b) Square-well approximation to the δ well.

Schrödinger equation has the form

$$\psi=\begin{cases} A_r\exp(-q_r|x|) \text{ (well exterior) ,} \\ B_r\cos(Q_r\epsilon^{1/2}x+C_r) \text{ (well interior) .}\end{cases} \tag{40}$$

Here the subscripts $r=1,2$ refer to material r, while A_r, B_r, and C_r are constants. Moreover,

$$q_r=(-2m_rE)^{1/2}/\hbar \ , \tag{41}$$

and

$$Q_r=\{2m_r[\epsilon E+\mu/(a_1+a_2)]\}^{1/2}/\hbar \ . \tag{42}$$

Connecting the wave function and its derivative at the three boundaries $x=-a_1\epsilon$, 0, and $a_2\epsilon$, in accordance with the matching conditions (37), we obtain the following implicit equation for the bound-state energy E:

$$\sum_{r=1}^{2} m_r^\beta Q_r \frac{q_r-Q_r\epsilon^{-1/2}\tan(a_r\epsilon^{1/2}Q_r)}{Q_r+\epsilon^{1/2}q_r\tan(a_r\epsilon^{1/2}Q_r)}=0 \ . \tag{43}$$

In the limit $\epsilon\rightarrow 0$ this simplifies to

$$\sum_{r=1}^{2} m_r^\beta(q_r-a_rQ_r^2)=0 \ , \tag{44}$$

or

$$\sum_{r=1}^{2} m_r^\beta[q_r-2\mu a_r m_r(a_1+a_2)^{-1}\hbar^{-2}]=0 \ , \tag{45}$$

using the $\epsilon\rightarrow 0$ limit version of Eq. (42).

Inserting the connection (41) between q_r and E, we finally obtain the energy eigenvalue

$$E=-\frac{2\mu^2}{\hbar^2}\frac{(a_1m_1^{\beta+1}+a_2m_2^{\beta+1})^2}{(m_1^{\beta+1/2}+m_2^{\beta+1/2})^2(a_1+a_2)^2} \tag{46}$$

in the δ well. In the homogeneous limit $m_1=m_2$ this reduces to the well-known result $E=-\frac{1}{2}\mu^2m\hbar^{-2}$ independent of β, and of a_1,a_2.

The crucial observation is that for a heterostructure ($m_1\neq m_2$) the energy level E is unique, i.e., independent of a_1 and a_2, *if and only if* $\beta=-1$.

V. CONCLUDING REMARKS

We have in the present article explored the consequences of the different ways of ordering the momentum and the mass operators in the effective-mass Hamiltonian (2). Assuming the values of α, β, and γ to be universal, we have through exactly solvable test cases shown that only for $\alpha=\gamma=0$, $\beta=-1$ are results finite and self-consistent. For a more general starting point than the operator (2), which leads to the same conclusion, see the Appendix.

The resulting kinetic operator

$$-\tfrac{1}{2}\hbar^2\nabla m^{-1}\nabla \ , \tag{47}$$

was apparently first used by BenDaniel and Duke.[8]

It is interesting that a very recent study[9] finds that photoluminescence spectra for GaAs/(Al,Ga)As quantum wells are best fitted with a value of β in the neighborhood of -1.

The remaining question is of course the relation of effective-mass theory to a more complete treatment of the underlying crystal lattice. Needless to say, the application to heterostructures must at least have the same restrictions as for homogeneous materials. It is also known[10] that for heterostructures the validity is more restricted. However, exact model calculations have shown[10] that *when* the effective-mass approximation is valid, the operator ordering is given by the values $\alpha=\gamma=0, \beta=-1$.

APPENDIX

In this appendix we consider a more general starting point than Eq. (2). For a Hermitian kinetic operator one may take

$$\tfrac{1}{4}\sum_i c_i(m^{\alpha_i}pm^{\beta_i}pm^{\gamma_i}+m^{\gamma_i}pm^{\beta_i}pm^{\alpha_i}) \ , \tag{A1}$$

where $\alpha_i+\beta_i+\gamma_i=-1$ and where the coefficients c_i sum to unity. Introducing the wave function (6) we are again led to the effective potential equation (7), but with new coefficients that are given by the replacements

$$\beta=\sum_i c_i\beta_i \tag{A2}$$

and

$$(\alpha-\gamma)^2-\beta^2=\sum_i c_i[(\alpha_i-\gamma_i)^2-\beta_i^2] \ . \tag{A3}$$

This demonstrates that the many-term operator (A1) is in fact identical to a two-term operator of the form studied in the main text. The equations (A2) and (A3), together with $\alpha+\beta+\gamma=-1$, determine the "effective" values of α, β, and γ. The only new feature that may possibly occur is that the coefficient $(\alpha-\gamma)^2$, which plays such a crucial role in the discussion in Sec. III, can also be negative. (Thus, the effective values of α and γ are not necessarily real.) For $(\alpha-\gamma)^2$ negative the junction acts as an impenetrable *repulsive* barrier, rather than a singular attractive well. This situation is just as unphysical, and can only be avoided if $\alpha=\gamma$, as before.

[1] S. T. Pantelides, Rev. Mod. Phys. **50**, 797 (1978).
[2] J. C. Slater, Phys. Rev. **76**, 1592 (1949); H. M. James, *ibid.* **76**, 1611 (1949).
[3] For a review see L. Esaki, IEEE J. Quantum Electron **QE-22**, 1611 (1986).
[4] R. A. Morrow and K. R. Brownstein, Phys. Rev. B **30**, 678 (1984).
[5] R. A. Morrow, Phys. Rev. B **35**, 8074 (1987); **36**, 4836 (1987); P. Enders, Phys. Status Solidi B **139**, K113 (1987); H. C. Liu, Superlatt. Microstruct. **3**, 413 (1987).
[6] O. von Roos, Phys. Rev. B **27**, 7547 (1983).
[7] G. N. Watson, *Theory of Bessel functions* (Cambridge University Press, Cambridge, England, 1922).
[8] D. J. BenDaniel and C. B. Duke, Phys. Rev. **152**, 683 (1966).
[9] I. Galbraith and G. Duggan, Phys. Rev. B **38**, 10057 (1988).
[10] G. T. Einevoll and P. C. Hemmer, J. Phys. C **21**, L1193 (1988); G. T. Einevoll, P. C. Hemmer, and J. Thomsen (unpublished).