# Connection Rules versus Differential Equations for Envelope Functions in Abrupt Heterostructures

Bradley A. Foreman*

*Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*

(Received 12 January 1998)

The controversial question of whether envelope functions are continuous or discontinuous at an abrupt heterojunction is addressed by developing a systematic procedure for obtaining interface connection rules from differential equations. The results show that even though modern envelope-function theories are smooth and continuous, their associated connection rules are discontinuous, in agreement with traditional concepts. This resolves the dispute in favor of *both* sides, by showing that the two views are physically equivalent. [S0031-9007(98)05915-8]

PACS numbers: 73.20.Dx, 63.20.Dj, 68.35.Ja, 73.61.Ey

Over the past two decades, envelope-function models have successfully been used to interpret a wide variety of experimental data on electrons and phonons in semiconductor nanostructures [1–4]. However, despite their popularity (or perhaps because of it), these models have been the source of much debate in the literature, and even today there remains a fundamental disagreement over the properties of envelope functions in the immediate vicinity of an abrupt heterojunction.

The traditional view is that, since bulk effective-mass theory [5] is not valid for rapidly varying potentials, one must take care to avoid using differential equations at the interface itself. The best one can do there is to connect the bulk envelopes across the interface using connection rules obtained from microscopic models [6–15]. Such rules indicate that the envelopes are discontinuous in general, so it seems clear that the use of differential equations is not justified near an abrupt junction (although it may be a reasonable approximation in a few special cases).

The opposite viewpoint has been advocated by those who have tried to go beyond bulk effective-mass theory and establish a rigorous set of differential equations for abrupt heterostructures. The fundamentals of this approach were developed a decade ago for electrons [16–18], and more recently for phonons [19]. A crucial element of such theories is that the envelopes are smooth and continuous functions, even near an abrupt junction; this is what enables them to be described by differential equations. Near an interface, the coefficients in these equations depend on microscopic properties of the interface, which cannot be expressed in terms of bulk effective-mass parameters. When these interfaces properties are included, the differential equations give excellent agreement with the underlying microscopic theory [18,19].

In spite of this, it is still frequently claimed that differential equations cannot rigorously be justified at an abrupt junction [4,10,11,13–15,20]. The sticking point seems to be a conceptual difficulty in reconciling the strong emphasis on *continuity* in the latter approach with an equally strong emphasis on *lack* of continuity in the former. The object of this paper is to demonstrate that there is, in fact, no physical difference between the two points of view. This is done by developing a systematic general procedure for obtaining connection rules from differential equations. The results differ sharply from the usual textbook connection rules [21] for second-order differential equations. In essence, this is because connection rules refer only to extrapolated *bulk* envelopes, so they may predict a substantial discontinuity even though the underlying differential equation exhibits none.

This paper deals exclusively with envelope-function equations; it does not consider the separate issue of deriving such equations from microscopic theory (see [17–19,22,23]). The equation of interest is the Sturm-Liouville eigenvalue equation [21]:

$$-\frac{d}{dx}\left[\frac{1}{p(x)}\frac{dy}{dx}\right] + q(x)y(x) = \lambda w(x)y(x). \quad (1)$$

Here $y$ is an envelope function, which may be complex; $\lambda$ is an eigenvalue; and $p$, $q$, and $w$ are material properties of the heterostructure (functions of $\lambda$ in general), which must be real if (1) is to be Hermitian. Within a limited range of eigenvalues $\lambda$, this equation can often be used to describe the motion of electrons [22] or phonons [19] along the growth axis of a [100] zinc-blende heterostructure. For electrons, $\lambda$ is the energy and $w = 1$; for acoustic or optical phonons, $\lambda$ is the square of the frequency and $w$ is a mass density, with $q = 0$ for acoustic modes.

Equation (1) is the most general second-order single-component envelope-function equation permitted by the symmetry of such systems [24]. As emphasized above, the values of $p$, $q$, and $w$ near an interface are not related to their values in the bulk. These functions, along with the envelope $y$, are smooth and infinitely differentiable [17–19], with a Fourier spectrum confined (either strictly [17,18] or in the Gaussian sense [22]) to the first Brillouin zone of the underlying lattice. Equation (1) is accurate only when the Fourier transform of $y$ is concentrated fairly close to the zone center; otherwise, higher-order differential operators (neglected here) become important.

  

This condition is surprisingly easy to satisfy, even for discontinuous envelopes [23].

Before discussing the interface region in detail, it is helpful to establish some general properties of the solutions to (1). A useful technique for such analysis is the transfer-matrix method [10,25,26], which is usually presented in terms of two exact but unspecified solutions. Here an alternative approach is developed, using a perturbative expansion of the integral equations associated with (1). Upon integrating (1) from $x_i$ to $x_f$ one obtains

$$z(x_f) = z(x_i) + \int_{x_i}^{x_f} Q(x)y(x)\,dx\,, \tag{2}$$

where $z(x) = p^{-1}(x)dy/dx$ and $Q(x) = q(x) - \lambda w(x)$. Multiplying (2) by $p(x_f)$ and integrating once more yields

$$y(x_f) = y(x_i) + z(x_i)g(x_f, x_i) + \int_{x_i}^{x_f} g(x_f, x)Q(x)y(x)\,dx\,, \tag{3}$$

in which $g(x_f, x_i) = \int_{x_i}^{x_f} p(x)\,dx$. Repeated substitution of (3) into the integrands of (2) and (3) permits one to express $y$ and $z$ at $x_f$ in terms of their values at $x_i$:

$$\mathbf{y}(x_f) = T(x_f, x_i)\mathbf{y}(x_i)\,, \tag{4}$$

where $\mathbf{y}(x) = [\,y(x)\ z(x)]^{\mathrm{T}}$. The transfer matrix $T$ is written as $T = \sum_{n=0}^{\infty} T^n$, where $n$ indicates the power of $Q$ appearing in each term. The $n = 0$ matrix is given by

$$T^0(x_f, x_i) = \begin{bmatrix} 1 & g(x_f, x_i) \\ 0 & 1 \end{bmatrix}, \tag{5}$$

while the matrices for $n > 0$ are obtained by iteration:

$$\begin{aligned} T^n_{1j}(x_f, x_i) &= \int_{x_i}^{x_f} g(x_f, x)Q(x)T^{n-1}_{1j}(x, x_i)\,dx\,, \\ T^n_{2j}(x_f, x_i) &= \int_{x_i}^{x_f} Q(x)T^{n-1}_{1j}(x, x_i)\,dx \\ &= \frac{1}{p(x_f)}\,\frac{\partial}{\partial x_f}\,T^n_{1j}(x_f, x_i)\,, \end{aligned} \tag{6}$$

in which $j = 1$ or 2.

From Eq. (1), one can easily verify the current-density conservation law $dJ/dx = 0$, where $J = \mathrm{Im}(y^* z)$. Since $T$ is real, this implies that $\det T = 1$ [6,9,26]. The inverse of the matrix $T(x_f, x_i)$ is therefore given by

$$T(x_i, x_f) = \begin{bmatrix} T_{22}(x_f, x_i) & -T_{12}(x_f, x_i) \\ -T_{21}(x_f, x_i) & T_{11}(x_f, x_i) \end{bmatrix}. \tag{7}$$

The result (7) also holds separately for each matrix $T^n$, but $\det T = 1$ is exact only for $T$ and $T^0$, not for any other finite-order approximation to $T$ [e.g., $\det(T^0 + T^1)$ differs from 1 by terms of order $Q^2$].

If $p$, $q$, and $w$ are constant between $x_i$ and $x_f$, then the integrals (6) are easy to evaluate explicitly:

$$T^n(x_f, x_i) = \begin{bmatrix} \frac{\alpha^{2n} d^{2n}}{(2n)!} & \frac{p\alpha^{2n} d^{2n+1}}{(2n+1)!} \\ \frac{\alpha^{2n} d^{2n-1}}{p(2n-1)!} & \frac{\alpha^{2n} d^{2n}}{(2n)!} \end{bmatrix}, \tag{8}$$

where $\alpha^2 = p(q - \lambda w)$ and $d = x_f - x_i$ [note that $T^0_{21} = 0$ since $(-1)! = \infty$]. The total transfer matrix is thus

$$T(x_f, x_i) = \begin{bmatrix} \cosh \alpha d & (p/\alpha)\sinh \alpha d \\ (\alpha/p)\sinh \alpha d & \cosh \alpha d \end{bmatrix}, \tag{9}$$

which is identical to the usual textbook result [26].

With these preliminaries established, we may now investigate what happens at an abrupt heterojunction, as depicted in Fig. 1. For an interface nominally at $x = 0$, the interface is assumed to occupy a region of finite width, $|x| < \epsilon$. Inside this region the material properties are smooth but arbitrary, while outside it they are constant [e.g., $p(x) = p_-$ for $x \le -\epsilon$ and $p(x) = p_+$ for $x \ge \epsilon$]. The restriction to finite $\epsilon$ is used only to simplify the ensuing discussion; it can be discarded if the interface perturbations are exponentially localized.

There is no difficulty in solving Eq. (1) numerically just as it stands, regardless of how complicated the interface dependence of $p$, $q$, and $w$ may be. However, it is seldom possible to find analytical solutions unless one replaces these parameters with *extrapolated* bulk functions $\{p_e, q_e, w_e\}$ that are piecewise constant (see Fig. 1). One can then try to reproduce $y(x)$ by imposing connection rules on $y_e(x)$ at $x = 0$, where $y_e(x \neq 0)$ satisfies (1) with $\{p, q, w\} \to \{p_e, q_e, w_e\}$. (This is the same as the traditional approach [6–15], except for the existence of an underlying differential equation.)

Clearly no connection rules for $y_e$ can reproduce $y$ exactly. The best one can do is require that $y_e(x) = y(x)$ when $|x| \ge \epsilon$ (see Fig. 1), which also implies $\lambda_e = \lambda$. This yields the connection rule

$$\mathbf{y}_e(0^+) = T^i(\lambda)\mathbf{y}_e(0^-)\,, \tag{10}$$

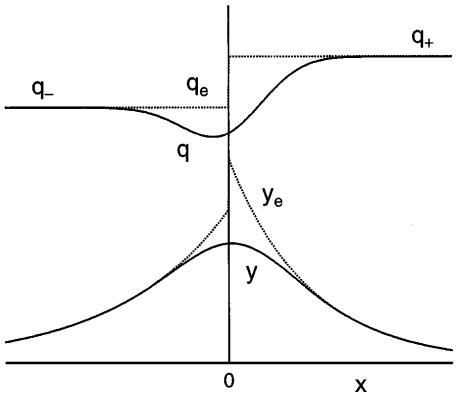


FIG. 1. Schematic illustration of the relation between actual functions $(q, y)$ and extrapolated functions $(q_e, y_e)$ in the region near an interface.

where the interface matrix $T^i(\lambda)$ is defined by

$$T^i = T^e(0^+, \epsilon)T(\epsilon, -\epsilon)T^e(-\epsilon, 0^-). \tag{11}$$

This is obtained simply by matching $\mathbf{y}_e = \mathbf{y}$ at $x = \pm\epsilon$, then propagating $\mathbf{y}$ and $\mathbf{y}_e$ within the interface region via $T[p, q, w]$ and $T^e[p_e, q_e, w_e]$.

As an example, consider a square quantum well from a perspective in which the entire well is just an "interface" to be replaced by appropriate connection rules. For simplicity let $p = w = 1$, with $q = 0$ when $|x| < \epsilon$ and $q = q_0$ when $|x| > \epsilon$. In Eq. (11), the matrix $T^e$ is given by (9) with $\alpha = (q_0 - \lambda)^{1/2}$, while $T$ is given by (9) with $\alpha \to ik$, where $k = \lambda^{1/2}$. The elements of $T^i$ are then

$$\begin{aligned}
T^i_{11} &= \cos 2k\epsilon \cosh 2\alpha\epsilon \\
&\quad + \tfrac{1}{2}(k/\alpha - \alpha/k)\sin 2k\epsilon \sinh 2\alpha\epsilon, \\
T^i_{12} &= (\alpha^2 k)^{-1}(\alpha^2 \cosh^2 \alpha\epsilon - k^2 \sinh^2 \alpha\epsilon)\sin 2k\epsilon \\
&\quad - \alpha^{-1} \cos 2k\epsilon \sinh 2\alpha\epsilon, \\
T^i_{21} &= k^{-1}(\alpha^2 \sinh^2 \alpha\epsilon - k^2 \cosh^2 \alpha\epsilon)\sin 2k\epsilon \\
&\quad - \alpha \cos 2k\epsilon \sinh 2\alpha\epsilon,
\end{aligned} \tag{12}$$

and $T^i_{22} = T^i_{11}$. The bound states of this interface are found by letting $y_e(x) = A_- e^{\alpha x}$ for $x < 0$ and $y_e(x) = A_+ e^{-\alpha x}$ for $x > 0$, which implies

$$\alpha(T^i_{11} + T^i_{22}) + \alpha^2 T^i_{12} + T^i_{21} = 0. \tag{13}$$

After some algebra this reduces to

$$(\cot k\epsilon - k/\alpha)(\cot k\epsilon + \alpha/k) = 0, \tag{14}$$

which is the well-known dispersion relation for a square quantum well of width $2\epsilon$.

Therefore, provided we include enough terms in the perturbative expansion, $T^i$ can generate exact eigenvalues even for strongly confining systems. The same cannot be said for the eigenfunctions, however, since for strong confinement $y_e(x)$ will not resemble $y(x)$ even qualitatively. Connection rules are consequently useful only if the interface is a *weak perturbation*, with at most one bound interface state. Fortunately this is true for most high-quality semiconductor heterojunctions [9], where the interface width is of the order of a lattice constant, and the interface fluctuations of $p$, $q$, and $w$ are not large [18].

Thus, for any case in which connection rules are physically acceptable, a first-order approximation to $T^i$ should be sufficient to provide an accurate description of the interface behavior. Such an approximation is readily calculated from Eqs. (6) and (11), with the result

$$T^i = (1 - b^2 - ac)^{-1/2}\begin{bmatrix} 1 + b & c \\ a & 1 - b \end{bmatrix}, \tag{15}$$

in which

$$\begin{aligned}
a &= \int_{-\epsilon}^{\epsilon} Q_i \, dx, \\
b &= \int_{-\epsilon}^{\epsilon} [-x p_e Q_i + g_i(\epsilon \operatorname{sgn} x, x) Q_e] \, dx, \\
c &= \int_{-\epsilon}^{\epsilon} [p_i - x^2 p_e^2 Q_i + 2x g_i(\epsilon \operatorname{sgn} x, x) p_e Q_e] \, dx.
\end{aligned} \tag{16}$$

Here $f_i(x) = f(x) - f_e(x)$, and no terms beyond the first order (e.g., $Q^2$ or $p_i Q_i$) were retained. The prefactor $(1 - b^2 - ac)^{-1/2}$ in Eq. (15) is the only exception: It was added to ensure that $T^i$ satisfies current conservation (i.e., $\det T^i = 1$) not just to first order but *exactly*. For weak perturbations this factor differs from 1 only by small terms of the second order.

The simplest interface perturbation is a smoothing of the interface, which is often invoked to explain discrepancies between experimental data and square-well calculations [27]. A smooth variation of $p$, $q$, and $w$ is, in fact, a *derived* feature of *ab initio* envelope-function theories, even at microscopically abrupt interfaces [18,22,23]. Interface smoothing is conveniently described by an error-function profile $Q(x) = \overline{Q} + \frac{1}{2}\Delta Q \operatorname{erf}(x/\sigma)$, where $\overline{Q} = \frac{1}{2}(Q_+ + Q_-)$, $\Delta Q = Q_+ - Q_-$, and $\sigma$ is the Gaussian half-width of the interface region (which should be no less than half a monolayer [18,22]); this gives $Q_i(x) = -\frac{1}{2}\Delta Q \operatorname{sgn}(x) \operatorname{erfc}(|x|/\sigma)$. If we assume for the moment that $p(x)$ is constant, then $a = c = 0$, but $T^i$ is not the unit matrix, since

$$b = \tfrac{1}{4}\sigma^2 p \Delta Q. \tag{17}$$

Interface smoothing therefore generates a discontinuity in the extrapolated envelope $y_e(x)$, with $y_e(0^+) > y_e(0^-)$ when $p\Delta Q > 0$. This makes sense physically, since for conduction electrons (where $p > 0$ and $\Delta w = 0$), $\Delta q > 0$ means the interface potential $q_i(x)$ is repulsive for $x < 0$ and attractive for $x > 0$.

This type of diagonal interface matrix was proposed in Ref. [11] as a replacement for the interface-smoothing concept used in Ref. [27]. The author of [11] noted that the two approaches gave equivalent results, but claimed that interface smoothing is invalid because $\operatorname{erf}(x/\sigma)$ is too rapidly varying at the interface to be permitted in an envelope-function calculation. Such claims can no longer be supported in light of modern advances in envelope-function theory [18,23]. Equation (17) proves that the two approaches are indeed equivalent, despite the seeming paradox in which a *smoother* interface gives rise to a *stronger* discontinuity in $y_e(x)$. (A paradox exists only when $y_e$ is mistaken for the true envelope $y$.)

Microscopic interface effects may be represented by a series of terms proportional to $\delta_\sigma(x)$ and its derivatives, where $\delta_\sigma(x) = (\sigma\sqrt{\pi})^{-1}\exp(-x^2/\sigma^2)$ is a Gaussian representation of a finite-width $\delta$ function. If we let $Q_i(x) = Q_0(\lambda)\delta_\sigma(x)$ and again hold $p(x)$ constant, then

$b$ is not affected, but $T^i$ acquires new off-diagonal terms:

$$a = Q_0, \qquad c = -\tfrac{1}{2}\sigma^2 p^2 Q_0 . \tag{18}$$

The element $T^i_{21}$ is normally interpreted as a $\delta$-function potential [9,14]; Eqs. (16) and (18) confirm this interpretation. The interpretation of $T^i_{12} \simeq c$ is less clear [9,14], but Eq. (18) shows that one effect described by $c$ is the finite width of any real interface.

Terms proportional to $\delta'_\sigma(x)$ have odd parity, so they are qualitatively the same as interface smoothing, except the magnitude and sign of $b$ are independent of the bulk properties of the heterostructure. The association of $\delta'$ potentials with $T^i_{12}$ in Ref. [14] is therefore incorrect. Terms proportional to $\delta''_\sigma(x)$ contribute to $c$ but not to $a$ or $b$.

If $p$ is only piecewise constant (i.e., $\Delta p \neq 0$ but $p_i = 0$), the effects described above become more entangled. For example, a smooth interface with a $\delta_\sigma$ term has

$$\begin{aligned} a &= Q_0 , \\ b &= \frac{\sigma^2}{4}\,\overline{p}\Delta Q - \frac{\sigma}{2\sqrt{\pi}}\,\Delta p Q_0 , \\ c &= \frac{\sigma^2}{4}\,\overline{p^2} Q_0 + \frac{\sigma^3}{6\sqrt{\pi}}\,\Delta(p^2)\Delta Q . \end{aligned} \tag{19}$$

Thus, a clear identification of $b$ and $c$ with particular types of interface perturbations is no longer possible.

Interface fluctuations of $p$ have no effect on $a$, but they do alter $b$ and $c$. If we include smoothing plus a $\delta_\sigma$ term in $p$, then $b$ and $c$ have, in addition to (19), contributions of the form

$$\begin{aligned} b &= -\frac{\sigma^2}{4}\,\Delta p\overline{Q} + \frac{\sigma}{2\sqrt{\pi}}\,p_0\Delta Q , \\ c &= p_0(1 + \tfrac{1}{2}\sigma^2\overline{pQ}) - \frac{\sigma^3}{6\sqrt{\pi}}\,\Delta p\Delta(pQ) , \end{aligned} \tag{20}$$

The most straightforward interpretation of $c$ is, therefore, that of a (smooth and finite) $\delta$ function in $p$, although many other effects also contribute to $c$. Which effect is dominant will depend on the details of the interface.

Previous derivations of $T^i$ from differential equations [19,22,23] tended to focus mostly on Eq. (2) and $T^i_{21}$, although Refs. [18] and [23] did note the correlation between $\delta'$ potentials and envelope discontinuities. However, as shown here, a consistent first-order analysis of the interface must include all four components of $T^i$. There is often good reason to neglect $T^i_{12}$ [9], but this should not be assumed without proof.

In conclusion, connection rules and differential equations are equally valid representations of interface behavior in the usual case where the interface is a weak perturbation. However, textbook methods for deriving connection rules do not reflect this equivalence. The method developed here shows that the apparent qualitative discrepancy between these representations is illusory—their physical content is the same.

This work was supported by Hong Kong RGC Grant No. DAG96/97.SC38.

*Electronic address: phbaf@ust.hk

[1] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Wiley, New York, 1988).
[2] B. K. Ridley, *Electrons and Phonons in Semiconductor Multilayers* (Cambridge University, Cambridge, England, 1997).
[3] E. L. Ivchenko and G. E. Pikus, *Superlattices and Other Heterostructures: Symmetry and Optical Phenomena* (Springer, Berlin, 1997), 2nd ed.
[4] V. R. Velasco and F. García-Moliner, Surf. Sci. Rep. **28**, 123 (1997).
[5] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[6] T. Ando and S. Mori, Surf. Sci. **113**, 124 (1982).
[7] Q.-G. Zhu and H. Kroemer, Phys. Rev. B **27**, 3519 (1983).
[8] H. Akera and T. Ando, Phys. Rev. B **40**, 2914 (1989).
[9] T. Ando, S. Wakahara, and H. Akera, Phys. Rev. B **40**, 11 609 (1989).
[10] W. Trzeciakowski, Phys. Rev. B **38**, 12 493 (1988).
[11] W. Trzeciakowski, Semicond. Sci. Technol. **10**, 768 (1995).
[12] J. P. Cuypers and W. van Haeringen, J. Phys. Condens. Matter **4**, 2587 (1992); Phys. Rev. B **47**, 10 310 (1993).
[13] T. Yamanaka *et al.*, J. Appl. Phys. **76**, 2347 (1994).
[14] R. Balian, D. Bessis, and G. A. Mezincescu, Phys. Rev. B **51**, 17 624 (1995).
[15] R. Balian, D. Bessis, and G. A. Mezincescu, J. Phys. I (France) **6**, 1377 (1996).
[16] M. G. Burt, Semicond. Sci. Technol. **2**, 460 (1987).
[17] M. G. Burt, Semicond. Sci. Technol. **3**, 739 (1988).
[18] M. G. Burt, J. Phys. Condens. Matter **4**, 6651 (1992).
[19] B. A. Foreman, Phys. Rev. B **52**, 12 260 (1995).
[20] C. Trallero-Giner, F. García-Moliner, V. R. Velasco, and M. Cardona, Phys. Rev. B **45**, 11 944 (1992).
[21] B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956).
[22] B. A. Foreman, Phys. Rev. B **52**, 12 241 (1995).
[23] B. A. Foreman, Phys. Rev. B **54**, 1909 (1996).
[24] It is well known [21] that every *real* Hermitian second-order differential equation may be written in the form (1), but for electrons an imaginary first-order term might also exist. Such a term is ruled out by time-reversal symmetry for spinless $\Gamma$ electrons in zinc-blende heterostructures [22]; it is permitted for electrons with spin, but these also require a two-component envelope.
[25] H. M. James, Phys. Rev. **76**, 1602 (1949).
[26] R. A. Smith, *Wave Mechanics of Crystalline Solids* (Chapman and Hall, London, 1969), 2nd ed.
[27] D. F. Nelson, R. C. Miller, C. W. Tu, and S. K. Sputz, Phys. Rev. B **36**, 8063 (1987).