Geometrical approach to scattering in one dimension

2004 J. Phys. A: Math. Gen. 37 1861

(http://iopscience.iop.org/0305-4470/37/5/026)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **37** (2004) 1861–1880 PII: S0305-4470(04)67133-2

Geometrical approach to scattering in one dimension

D W L Sprung[1], **G V Morozov**[1] **and J Martorell**[2]

[1] Department of Physics and Astronomy, McMaster University Hamilton, Ontario L8S 4M1, Canada
[2] Department d'Estructura i Constituents de la Materia, Facultat Física, University of Barcelona, Barcelona 08028, Spain

Received 1 August 2003, in final form 22 October 2003
Published 19 January 2004
Online at stacks.iop.org/JPhysA/37/1861 (DOI: 10.1088/0305-4470/37/5/026)

Abstract
The transfer matrix technique is much used to study one-dimensional scattering, such as electrons in semiconductor superlattices. We elaborate on the geometrical interpretation of the transfer matrix, in particular its relation to conformal mapping in the unit disc, and the analogy to Lorentz transformations on a Dirac spinor in 2 + 1 dimensions.

PACS numbers: 02.10.Y, 03.65.N, 73.25.A, 73.21.A

1. Introduction

Layered systems occur in many branches of physics. One effective technique for studying such one-dimensional systems is the transfer matrix method, in which the amplitudes of two fundamental solutions on either side of a layer (or a potential cell) are connected by a transfer matrix M, in the notation of [1]. In the context of periodic systems, the transfer matrix is often called the *monodromy* matrix [2], though strictly speaking it is the W-matrix form (cf [3] and [1]) of the transfer matrix that is so denoted.

Even when the properties of the medium are continuous, it may be advantageous to treat it as composed of separate layers as a method of solution [4, 5]. Griffiths and Steinke [6] have recently reviewed applications of the transfer matrix to many problems in both classical and quantum physics; this is a rich source of references. Among the currently active areas are ballistic transport of electrons through one-dimensional superlattices in semiconductor heterostructures, [7, 8] and photonic lattices [9–11] to cite only a few representative examples. Optics of course is a field in which uniform multilayers are ubiquitous, and the transfer matrix method is well established [12, 13].

We focus on cases where the one-dimensional Schrödinger equation applies. Hermiticity of the Hamiltonian implies flux conservation and time reversal symmetry, leading to a transfer matrix M with the properties $\det M = 1$, $M_{22} = M_{11}^*$, and $M_{12} = M_{21}^*$. No matter how complicated the shape of the potential function within a cell, at a given energy M can be parametrized in terms of three real parameters, one of which is conveniently chosen to be

the Bloch phase, defined by $2\cos\phi = \mathrm{Tr}\, M$. If the cell has reflection symmetry, just one additional parameter suffices.

Since any 2×2 matrix can be expressed in terms of the Pauli matrices, it is natural to think of the analogy to rotations of a spin-half system. In a recent paper [14] on the design of anti-reflection coatings (ARCs) for electrons in semiconductor heterostructures, we made use of this analogy to provide a simple and intuitively appealing picture of how such a coating works. A wave moving purely to the right corresponds to the state of spin-up along OZ. An arbitrary transfer matrix M causes the spin to precess by angle 2ϕ around an effective magnetic field direction in space, specified by the two additional parameters. The role of the upstream ARC is to rotate the spin-up state to lie along that magnetic field direction, so that the precession due to M merely gives an overall phase factor to the state, preserving the ratio of the left and right moving probabilities. The downstream ARC then restores the state to its original value (up to a phase).

In [14], we did not have space to develop the analogy in any detail; that is the purpose of the present work. Since any wave motion can be treated in similar manner, the same techniques can be applied to optics, as Monzón and collaborators [15, 16] and Kim's group [17, 18] have recently done. Our work differs from theirs in that they consider only the case of uniform monolayer cells (i.e. piecewise flat potentials), and they approach the subject from the viewpoint of group representations and in particular the Iwasawa decomposition. Our methods are valid for any shape of potential cell, and our presentation uses more elementary methods that should be familiar to most physicists.

In the next section we explain our notation for the transfer matrix, and study its action as a conformal mapping in the complex plane. We then take the case of a delta function potential as an illustration. In section 5 we interpret the action as a Lorentz transform on a Dirac spinor in $(2 + 1)$-dimensional spacetime. The two pictures are related by stereographic projection. The mathematical details are relegated to two appendices.

2. Transfer matrix and conformal mapping on the unit disc

In the single-band envelope function approximation, the electron wavefunction in a semiconductor satisfies the Schrödinger equation in a real potential and with an energy and position dependent (real) effective mass [19] $m^*(E, x)$:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{m}{m^*}\frac{\mathrm{d}\psi(x)}{\mathrm{d}x}\right] + \frac{2m}{\hbar^2}[E - V(x)]\psi(x) = 0. \tag{1}$$

Suppose that the potential $V = V_0$ and $m^* = m_0^*$ take constant values in both directions outside a cell located on $c < x < d$. The outside wave function may be written as

$$\begin{aligned} \psi(x) &= f_L(c)\,\mathrm{e}^{\mathrm{i}k(x-c)} + g_L(c)\,\mathrm{e}^{-\mathrm{i}k(x-c)} \qquad x \leqslant c \\ &= f_R(d)\,\mathrm{e}^{\mathrm{i}k(x-d)} + g_R(d)\,\mathrm{e}^{-\mathrm{i}k(x-d)} \qquad x \geqslant d \\ k^2 &= \frac{2m_0^*}{\hbar^2}[E - V_0] \end{aligned} \tag{2}$$

and the transfer matrix relates the amplitudes across the cell:

$$\begin{aligned} &\begin{pmatrix} f_L(c) \\ g_L(c) \end{pmatrix} = M \begin{pmatrix} f_R(d) \\ g_R(d) \end{pmatrix} \qquad \text{where} \\ &M = \begin{pmatrix} 1/t & r^*/t^* \\ r/t & 1/t^* \end{pmatrix} \end{aligned} \tag{3}$$

in terms of the reflection $r(k)$ and transmission $t(k)$ amplitudes with waves incident from the left.

It is easy to check that $M\sigma_z M^\dagger = \sigma_z$, and therefore the norm $|f|^2 - |g|^2$ is preserved by the action of M on the wave-function. Because of this, the (complex) ratio $z = g/f$ contains the essential information about the wave-function, and omits only an overall phase factor. The action of a transfer matrix is a mapping from the value on the right $z = g_R/f_R$ at $x = d$ to $w = g_L/f_L$ at $x = c$ on the left. For the usual scattering solution with incident waves from the left, $z = 0$ on the right, and $w = r$, the reflection amplitude, on the left. When $w = 0$, we have the necessary and sufficient condition for a transparent potential. For a general point z

$$\begin{aligned} w &= \frac{r/t + z/t^*}{1/t + zr^*/t^*} = \frac{-r^*/t^* + z/t^*}{1/t - zr/t} \\ &= \frac{t}{t^*}\frac{z - r^*}{1 - rz} = \frac{r}{r^*}\frac{r^* - z}{1 - rz} \end{aligned} \tag{4}$$

which is a bilinear or Möbius mapping of the complex plane [20] onto itself. It conformally maps the interior of the unit circle into itself, with r^* becoming the origin, and the origin going to r.

The invariant points of the mapping are most easily identified by finding the eigenstates of the operator M acting on two-component spinors. Before doing that, we adopt a useful parametrization of M which was developed by Kard [21] (but only for symmetric monolayers). In an allowed band, where ϕ is real, it takes the form

$$\begin{aligned} M_{11} &= \cos\phi - \mathrm{i}\sin\phi\cosh\mu = M_{22}^* \\ M_{21} &= -\mathrm{i}\,\mathrm{e}^{\mathrm{i}\chi}\sin\phi\sinh\mu = M_{12}^*. \end{aligned} \tag{5}$$

This form respects the relation $1 \leqslant 1/|t|^2 = 1 + \sin^2\phi\sinh^2\mu$ as well as $\det M = 1$. We call μ the impedance parameter, because in the case of a square well, e^μ is the impedance, the ratio of velocity outside to inside the cell [22]. The symmetric cell corresponds to the special case $\chi = 0$. Given M, ϕ is determined by the trace, χ by the phase of M_{21}/M_{12}, and $\tanh\mu$ by the ratio of $|M_{21}|$ to $\mathrm{Im}\,M_{11}$.

We can now write

$$\begin{aligned} M &= \begin{pmatrix} 1/t & r^*/t^* \\ r/t & 1/t^* \end{pmatrix} = \cos\phi - \mathrm{i}\sin\phi[\cosh\mu + \sinh\mu(\cos\chi\sigma_x + \sin\chi\sigma_y)]\sigma_z \\ &= \mathrm{e}^{-\mathrm{i}(\chi/2)\sigma_z}\,\mathrm{e}^{(\mu/2)\sigma_x}\,\mathrm{e}^{-\mathrm{i}\phi\sigma_z}\,\mathrm{e}^{-(\mu/2)\sigma_x}\,\mathrm{e}^{\mathrm{i}(\chi/2)\sigma_z}. \end{aligned} \tag{6}$$

The last form expresses M as a sequence of elementary transforms, which we will discuss in more detail later. For the moment we note that the effect of an asymmetric potential ($\chi \neq 0$) is just a rotation of the geometry by angle χ in the XY plane, as compared to the symmetric case. For simplicity we will sometimes discuss the $\chi = 0$ case, which implies the relation $r^*/t^* = -r/t$, making the off-diagonal elements of M pure imaginary.

We can now discuss the invariant points of the bilinear mapping. The matrix whose two columns are the eigenstates is $U = \mathrm{e}^{-\mathrm{i}(\chi/2)\sigma_z}\,\mathrm{e}^{(\mu/2)\sigma_x}$. This is easily verified by direct substitution into equation (6):

$$\begin{aligned} MU &= U\,\mathrm{e}^{-\mathrm{i}\phi\sigma_z} \\ U &= \begin{pmatrix} \mathrm{e}^{-\mathrm{i}\chi/2}\cosh\mu/2 & \mathrm{e}^{-\mathrm{i}\chi/2}\sinh\mu/2 \\ \mathrm{e}^{\mathrm{i}\chi/2}\sinh\mu/2 & \mathrm{e}^{\mathrm{i}\chi/2}\cosh\mu/2 \end{pmatrix}. \end{aligned} \tag{7}$$

which shows that the eigenvalues are $\mathrm{e}^{\mp\mathrm{i}\phi}$. In terms of the mapping, the invariant values of z are the ratios of the elements in each column of U: $z = a = \mathrm{e}^{\mathrm{i}\chi}\tanh\mu/2$ and $z = b = \mathrm{e}^{\mathrm{i}\chi}\coth\mu/2$. They are collinear with the origin. Since $ab^* = 1$, a and b^* are inverse points with respect to the unit circle. They do not depend on the value of the Bloch phase ϕ, and for a symmetric

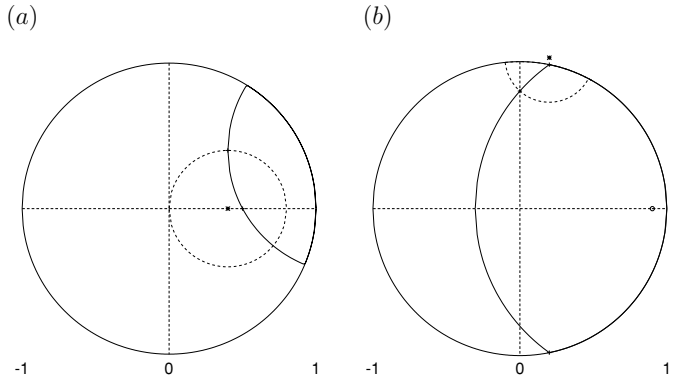


Figure 1. Examples of Kard paths (dashed lines) for mapping on unit disc when invariant points (*a*) are real or (*b*) lie on unit circle. Centre of the (solid line) circle of family A is indicated by an open circle, and the centre of the (dashed line) coaxal circle of family B by a star.

cell they are real. Physically, the eigenstates are the Bloch waves of an infinite periodic array of cells.

The general theory of the bilinear mapping [23] tells us that when we have two distinct invariant points, there are two families of circles with the property that points on a given circle transform into points on the same circle. The first family (A) are circles which pass through the invariant points, and the second 'coaxal' set (B) are orthogonal trajectories to the first set. They are described by the relations

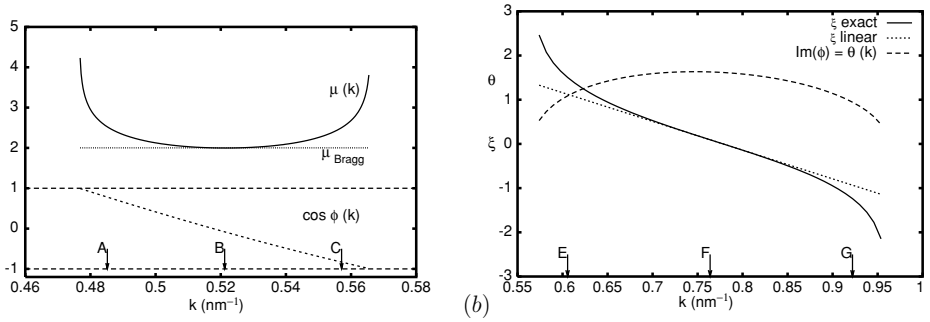
$$\begin{aligned} &\arg\left(\frac{z-a}{z-b}\right) = \text{constant} \quad \text{(A)} \\ &\left|\frac{z-a}{z-b}\right| = \text{constant} \quad \text{(B)} \end{aligned} \tag{8}$$

respectively. The centres of the orthogonal trajectories lie on the line connecting the invariant points. The two invariant points plus one initial point specify a unique circle of either family. Two examples are shown in figure 1. In (*a*), the fixed points lie on the real axis, while in (*b*), the invariant points lie on the unit circle near $0.2 \pm 0.98\mathrm{i}$. The coaxal circles, shown as dashed lines, are paths appropriate to motion in (*a*) an allowed band or (*b*) a forbidden band, in the vicinity of the crossing point of the two circles.

It is useful to discuss the mapping when M is written in the Kard form, equation (6). In the design of ARCs for electrons, we found [14] it was a good approximation to consider $\mu \sim \text{constant} = \mu_{\text{Bragg}}$, while ϕ varies from zero to π as the energy of the incident electron increases across an allowed band. An example is shown in figure 2(*a*), drawn from [24]. A mapping for fixed μ, χ and variable ϕ will be called a Kard path.

Starting from *any* initial point z_0, w will follow an 'orbit' in the complex plane which is one of the coaxal circles. Suppose that we begin from the point $z_0 = 0$, which corresponds to a state moving purely to the right; this we call the *principal orbit*. From equations (5) and (6), one has

$$\begin{aligned} r &= \frac{\sinh\mu}{\cosh\mu + \mathrm{i}\cot\phi}\,\mathrm{e}^{\mathrm{i}\chi} \\ \frac{r}{r^*} &= \frac{\tan\phi\cosh\mu - \mathrm{i}}{\tan\phi\cosh\mu + \mathrm{i}}. \end{aligned} \tag{9}$$



(*a*) (*b*)

Figure 2. Kard parameters for a symmetric delta cell of strength $\Omega = 1.88$ nm^{-1} and width $d = 5.55$ nm, used in [24]. (*a*): first allowed band, (*b*): forbidden band.

When $\phi = 0, r = 0$ so the origin maps onto itself. As ϕ increases, z_0 will map onto points w which lie on a circle of family (B) passing through the origin and enclosing the fixed point $a = \tanh(\mu/2)\,\mathrm{e}^{\mathrm{i}\chi}$. At the Bragg point $\phi = \pi/2$, $w = \tanh\mu\mathrm{e}^{\mathrm{i}\chi}$ which is again on the same ray. This is the diametrically opposite point on the orbit, and puts the centre of the circle midway to the origin.

If we begin from a different initial point, but still on the same ray, it will be mapped, when $\phi = \pi/2$, into the opposite side of the circle. This identifies the centre and radius of a circular orbit. If the initial point is very close to the invariant point, $a = \tanh(\mu/2)\,\mathrm{e}^{\mathrm{i}\chi}$, the infinitesimal circle has its centre at a. In the opposite extreme, an initial point on the unit circle will traverse the unit circle. For infinite μ, the invariant points both approach $z = \mathrm{e}^{\mathrm{i}\chi}$, and the orbits will be on circles interior to the unit circle but tangent to it at the double invariant point.

As remarked earlier, the origin corresponds to a wave moving purely to the right. Under the bilinear mapping, points inside the unit circle are mapped among themselves, as are the points outside the unit circle. A time-reversal invariant potential cannot connect the two sets. However, they are related by the inversion $z \to 1/z^*$, which interchanges the two invariant points. Physically the inversion corresponds to interchanging states which are intrinsically right-moving with those that are left-moving, and vice versa. When the initial state is right-moving, only points inside or on the unit circle are physical. The reflection operation Π on the potential cell corresponds to replacing $M \to M^{-1*} \equiv M^{(\pi)}$; if the cell is reflection symmetric this leaves M invariant.

In the design of a quarter-wave impedance transformer, we are concerned with behaviour of the transmission in the neighbourhood of the Bragg point, $\phi = \pi/2$. In that vicinity, μ has a broad minimum, and the approximation of constant $\mu = \mu_B$ was found to be a good one [14] (see figure 2(*a*)). Then the mapping causes w to move on an arc of a coaxal circle. We should emphasize that at a given energy, only two points on the orbit matter, the initial point and its image. But when one is concerned with transmission over a range of energies, the orbit is a convenient construct for seeing how the transmission will vary with energy. Successive potential cells move the system point in steps, within the unit circle. For identical cells, the mapping moves around a fixed circle.

While μ varies only slowly as one moves away from the band centre, the variation becomes much more rapid near a band edge, $\phi \to n\pi$. That is because M_{21} is a smoothly varying function of energy, but the second of equation (5) shows that when $\sin\phi \to 0$, $\sinh\mu$ must diverge in order to give the correct value of M_{21}. What this means is the following: in the vicinity of the Bragg point w travels on the circle centred at $(1/2)\tanh\mu$. But as ϕ moves away

from the Bragg point, the centre of the circle moves slowly to the right. The image point w moves with less curvature and on a wider circle. As $\phi \to n\pi$, w will not approach the origin, as the Kard form equation (5) appears to imply, rather it will be on a circle touching the unit circle. The variation of μ near the band edge allows the mapping to follow the correct path which is not exactly circular. The Kard parameterization with *fixed* μ is more valid the closer one is to the Bragg point. To make this point clear, we discuss an example in the following section. It would apply to a system with thin barriers and wide wells, such as that of Gomez *et al* [25].

3. Delta potential as an example

For a cell of width d consisting of a single delta barrier with free propagation on each side (d_1 on left, d_2 on right) we have

$$\begin{aligned} M &= \begin{pmatrix} \mathrm{e}^{-\mathrm{i}kd_1} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}kd_1} \end{pmatrix} \begin{pmatrix} 1+\mathrm{i}\Omega/k & \mathrm{i}\Omega/k \\ -\mathrm{i}\Omega/k & 1-\mathrm{i}\Omega/k \end{pmatrix} \begin{pmatrix} \mathrm{e}^{-\mathrm{i}kd_2} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}kd_2} \end{pmatrix} \\ &= \begin{pmatrix} (1+\mathrm{i}\alpha)\,\mathrm{e}^{-\mathrm{i}kd} & \mathrm{i}\alpha\,\mathrm{e}^{-\mathrm{i}kD} \\ -\mathrm{i}\alpha\,\mathrm{e}^{\mathrm{i}kD} & (1-\mathrm{i}\alpha)\,\mathrm{e}^{\mathrm{i}kd} \end{pmatrix} \end{aligned} \tag{10}$$

where $d = d_1 + d_2$, $D = d_1 - d_2$, and $\alpha = \Omega/k$ decreases smoothly to zero over the infinite energy interval $0 < E < \infty$. The role of χ is taken by kD, while α is $\sin\phi \sinh\mu$. For a symmetric cell, $D = 0$, and

$$\begin{aligned} &\cos\phi = \cos kd + (\Omega/k)\sin kd \\ &r = \frac{-\mathrm{i}\alpha}{1+\mathrm{i}\alpha}\,\mathrm{e}^{\mathrm{i}kd}. \end{aligned} \tag{11}$$

For a barrier, $\Omega > 0$. Then $\cos\phi(k=0) = 1 + \Omega d > 1$. The first allowed band begins when kd is in the second quadrant and satisfies $\cos kd + \alpha \sin kd = 1$. It ends at $kd = \pi = \phi$. The Bragg point is at $\tan kd = -1/\alpha = -kd/\Omega d$. Larger Ωd squeezes the band into a narrower energy range.

It is convenient to define $\tan\beta = 1/\alpha$; β rises from zero to $\pi/2$ over the same infinite energy interval. We can then write

$$\begin{aligned} &\cos\phi = \frac{\sin(kd+\beta)}{\sin\beta} - r = \frac{1}{1-\mathrm{i}/\alpha}\,\mathrm{e}^{\mathrm{i}kd} = \frac{1+\mathrm{i}\tan\beta}{1+\tan^2\beta}\,\mathrm{e}^{\mathrm{i}kd} \\ &r = -\frac{\alpha}{\sqrt{1+\alpha^2}}\,\mathrm{e}^{\mathrm{i}(kd+\beta)} = -\cos\beta\,\mathrm{e}^{\mathrm{i}(kd+\beta)}. \end{aligned} \tag{12}$$

Since the principal orbit follows $w = r$, it spirals in towards the origin, starting from -1 at $k = 0$, with the radius $\cos\beta$ slowly decaying. This is illustrated in figure 3 by the solid line which is followed through three allowed bands. The question is, how does the Kard parametrization capture this behaviour? At the Bragg point, $kd + \beta_0 = \pi$. Then

$$\begin{aligned} &\frac{r}{t} = -\mathrm{i}\alpha_0 = -\mathrm{i}\sin\phi\sinh\mu \\ &\sinh\mu = \alpha_0 = \cot\beta_0 \\ &r(Bragg) = \tanh\mu = \frac{\alpha_0}{\sqrt{1+\alpha_0^2}} = \cos\beta_0. \end{aligned} \tag{13}$$

In the first allowed zone, kd lies between $\pi/2$ and π; assuming large Ωd, $\beta_0 \sim 1/\alpha_0 \sim kd/\Omega d$ will be a small angle. Using the first of equation (12) we find the following estimates

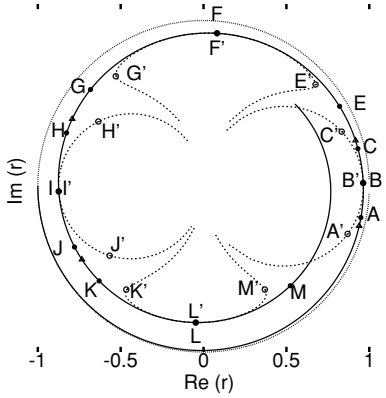


Figure 3. Kard path compared to exact path in two allowed zones and two forbidden zones for the delta-potential cell of figure 2.

for the lower and upper band edges and the Bragg point (the two filled triangles and point B at the right of figure 3):

$$\begin{aligned} k_l d &\approx \pi - 2\beta_l \qquad & \beta_l &\sim \frac{\pi}{\Omega d}\left(1 - \frac{2}{\Omega d}\right) \\ k_B d &\approx \pi - \beta_B \qquad & \beta_B &\sim \frac{\pi}{\Omega d}\left(1 - \frac{1}{\Omega d}\right) \\ k_u d &= \pi \qquad & \beta_u &\sim \frac{\pi}{\Omega d}. \end{aligned} \tag{14}$$

At the upper band edge $k_u d = \pi$, and

$$r(upper) = \frac{\alpha_u}{\sqrt{1+\alpha_u^2}}\, \mathrm{e}^{\mathrm{i}\beta_u}. \tag{15}$$

At the lower band edge the reflection amplitude is

$$r(lower) = \frac{\alpha_l}{\sqrt{1+\alpha_l^2}}\, \mathrm{e}^{-\mathrm{i}\beta_l}. \tag{16}$$

The radius of r has changed little. To leading order in $1/(\Omega d)$, the phase β is constant across the band, so the motion is symmetrical with respect to the Bragg point; an arc of a circle of approximately constant radius has been traversed. Since the true motion is a spiral, the continuation of the path must be traversed while the energy passes through the forbidden band.

When we take the Kard parametrization with fixed μ and variable ϕ, the circle begins at the origin and passes through the Bragg point B at $x = \tanh \mu_0$. The starting point is wrong, and this can only be accommodated by allowing μ to diverge at the band edge. The divergence is necessary if a good representation of the trajectory is to hold over the entire band width. We also need the angular motion of the reflection amplitude to be rapid compared to the rate of decrease of the radius. Then the orbit will be well described by a portion of the Kard circle, at least in the middle of the band. For the application to design ARCs [14], this is exactly what is needed. The points labelled A and C in figure 3 are at 10% and 90% of the band width; they correspond to A′ and C′ on the Kard trajectory for fixed $\mu = \mu_B$. (They also correspond to A and C as marked on figure 2(*a*).) This shows that even with a fixed μ, the Kard path remains close to the true path over most of an allowed band. Similar comments apply to points H, I, J in the second allowed band, which has $\mu < 0$.

4. Kard representation in a forbidden band

For simplicity we will discuss only the symmetric cell in this section. At a band edge, $\phi \to p\pi$, and in a forbidden band, $\cos\phi$ must be replaced by $(-)^p \cosh\theta$. It turns out that one must replace $\mu \to \xi + \mathrm{i}\pi/2$, and $\phi \to p\pi + \mathrm{i}\theta$. As we saw in the last section, the Kard parameter μ must diverge at a band edge in order to have a good representation of the actual transfer matrix; so does ξ. Numerical studies show that in a forbidden band, ξ usually changes sign, falling from say $+\infty$ to $-\infty$ at the next allowed band edge. It is never a good approximation to consider $\xi \approx$ constant; on the other hand, a constant rate of change with energy is quite reasonable over much of the forbidden zone; an example is shown in figure 2(*b*). With the above replacement of parameters, one has $\cos\phi \to (-)^p \cosh\theta$, while $\sinh\phi \to (-)^p\,\mathrm{i}\sinh\theta$. Similarly $\cosh\mu \to \mathrm{i}\sinh\xi$ and $\sinh\mu \to \mathrm{i}\cosh\xi$.

For a symmetric cell, ($\chi = 0$), the result is an overall factor $(-)^p$ on the Kard representation:

$$M = \begin{pmatrix} 1/t & r^*/t^* \\ r/t & 1/t^* \end{pmatrix} = (-)^p \cosh\theta + \mathrm{i}(-)^p \sinh\theta[\sinh\xi + \cosh\xi\sigma_x]\sigma_z. \tag{17}$$

In particular, $M_{21} = (-)^p\,\mathrm{i}\sinh\theta\cosh\xi$ has a sign which is determined by the sign of θ. For a repulsive delta potential it is a negative imaginary value, so the sign of θ must be $(-)^{p+1}$ depending on the band selected.

The factorized form of M is the last line of equation (6). It is convenient to work symbolically. Define $V = \mathrm{e}^{\mathrm{i}(\pi/4)\sigma_x}$ which is unitary, and the boost $B = \mathrm{e}^{+(\xi/2)\sigma_x}$, which is not. (A boost is also called a hyperbolic rotation; see for example [26].) The operators V and B commute. Then, with $\chi = 0$,

$$M = (-)^p BV\,\mathrm{e}^{\theta\sigma_z} V^\dagger B^{-1} = (-)^p B\,\mathrm{e}^{+\theta\sigma_y} B^{-1}. \tag{18}$$

The matrix of eigenstates is now

$$U = VB = BV = \frac{1}{\sqrt{2}}\begin{pmatrix} \cosh\xi/2 + \mathrm{i}\sinh\xi/2 & \cosh\xi/2 - \mathrm{i}\sinh\xi/2 \\ \cosh\xi/2 - \mathrm{i}\sinh\xi/2 & \cosh\xi/2 + \mathrm{i}\sinh\xi/2 \end{pmatrix}. \tag{19}$$

One easily sees that

$$\begin{aligned} (-)^p MU &= BV\,\mathrm{e}^{-\theta\sigma_z} B^{-1} V^\dagger VB \\ &= BV\,\mathrm{e}^{\theta\sigma_z} = U\,\mathrm{e}^{\theta\sigma_z} \end{aligned} \tag{20}$$

showing that the eigenvalues are $(-)^p\,\mathrm{e}^{\pm\theta}$. Correspondingly, the Bloch states vary by an exponential growth or decay factor across a cell. Corresponding to equation (9) we now have ($\mathrm{i}\tanh\theta = \tan\phi$; $\mathrm{i}\sinh\xi = \cosh\mu$)

$$\begin{aligned} r^* &= \frac{\cosh\xi}{\sinh\xi + \mathrm{i}\coth\theta} \\ \frac{r}{r^*} &= \frac{\tanh\theta\sinh\xi + \mathrm{i}}{\tanh\theta\sinh\xi - \mathrm{i}}. \end{aligned} \tag{21}$$

What is the orbit corresponding to a forbidden zone? As the energy approaches a band edge, the two fixed points approach $z = +1$. They then continue on the unit circle as a complex conjugate pair, maintaining their status as reciprocal points with respect to the unit circle. Taking the first forbidden zone with $p = 1$, $M_{11} = -\cosh\theta - \mathrm{i}\sinh\theta\sinh\xi$, $M_{21} = -\mathrm{i}\sinh\theta\cosh\xi$. From equation (19), the fixed points are at

$$z_F = \frac{\sinh\xi \pm \mathrm{i}}{\cosh\xi}. \tag{22}$$

The line joining the fixed points is now vertical, so the circles of both families A and B extend outside the unit circle. Those of family B have their centres on the vertical line ($x = \tanh\xi$) joining the fixed points. For an orbit point in the upper half plane, the circle of family B is concave upwards, as in figure 1(*b*). As $|\xi|$ falls from ∞, the fixed points rotate around the unit circle, towards $z = \pm i$, which is reached at $\xi = 0$. Typical behaviour of ξ and θ in the first forbidden band is shown in figure 2(*b*); ξ shows a tangent-like behaviour which is well approximated by the dotted line, the tangent where it crosses the axis. The imaginary phase $\theta = \mathrm{Im}(\phi)$ vanishes at the band edge and has a maximum where ξ vanishes, which could define the centre of the FZ. Points E, F, G in figure 2(*b*) correspond to those in figure 3.

With ξ negative the fixed points move into the third/fourth quadrants, still on the unit circle. As $\xi \to -\infty$, they reach $z = -1$ together. The Kard circle which is tangent to the orbit appears to roll along the orbit as the energy passes through the forbidden band. It is the linear variation of ξ with energy or k that allows this to occur. This behaviour contrasts sharply with the situation in an allowed band where $\mu \sim$ constant over much of the band, and a single Kard circle is a good approximation to the orbit.

In figure 3, points E, F, G on the true path correspond to 10%, 50% and 90% of the first forbidden zone, while E′, F′, and G′ are the corresponding points on the Kard path assuming linear variation of ξ. Apparently the Kard form in the forbidden band works even better than in an allowed band, because the dotted line through E′, F′, G′ stays very close to the true solid curve. This is because it is a locus of the points of contact of the coaxal circles which (under the linear variation of ξ assumption) lie close to the true curve. Near to the band edge, ξ is diverging much faster than linearly so G′ is getting further from G, and E′ from E. Fortunately, for the design of ARCs we are interested only in the orbits within allowed bands, where the $\mu = \mu_B$ approximation works well.

5. Three-dimensional geometry

5.1. Relation to Lorentz transformations

From the example of the delta-function potential, it is evident that there are two distinct problems: to describe the orbit of the Kard matrix, with a fixed value of μ (a coaxal circle) and secondly, to understand how, by variation of the impedance parameter, the actual orbit is followed as one passes through an allowed or a forbidden band. It will be helpful at this point to extend the discussion out of the complex plane into three-dimensional geometry.

Our original insight came from noting that the Kard operator takes a form similar to a rotation of a spin-half about a magnetic field, albeit one pointing in a complex direction [14]. Actually it is a Lorentz transform on a Dirac spinor in 2 + 1 dimensions, and the hyperbolic functions arise for this reason. We summarize the effect of Lorentz transformations on the Dirac equation in appendix A. The change of coordinates $x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu}$ induces the same transformation on the Dirac spin operators, equation (A.13). If we adopt the simple representation of equation (A.21), the elementary operators reduce to

$$\begin{aligned} B_x(\mu) &= \mathrm{e}^{\frac{1}{2}\mu\sigma_x} \qquad \text{boost along} \quad \mathrm{OX} \\ B_y(\mu) &= \mathrm{e}^{\frac{1}{2}\mu\sigma_y} \qquad \text{boost along} \quad \mathrm{OY} \\ &= \cosh(\mu/2) + \sinh(\mu/2)\sigma_y \\ R_z(\phi) &= \mathrm{e}^{-\mathrm{i}\frac{1}{2}\phi\sigma_z} \qquad \text{rotation in XY plane} \\ &= \cos(\phi/2) - \mathrm{i}\sin(\phi/2)\sigma_z. \end{aligned} \tag{23}$$

Any motion can be represented as a sequence of these elementary operations [27].

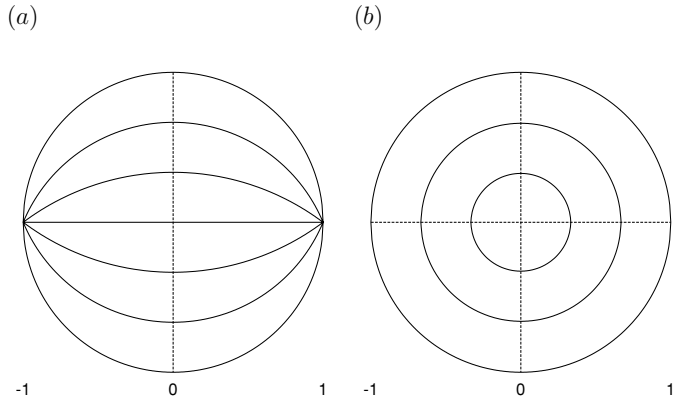


Figure 4. Examples of paths in unit disk under the action of: (*a*) the boost $B_x(\mu)$ and (*b*) the rotation $R_z(\phi)$ of equation (23).

For the boost $B_x(\mu)$, the invariant points are at ± 1; under this operator a point in the unit disc moves on one of the circles of family A through the fixed points, as illustrated in figure 4(*a*). For example, the point at the origin moves along the horizontal axis, while a point on the unit circle moves along it. For the boost $B_y(\mu)$, simply rotate the drawing by 90°. The rotation operator $R_z(\phi)$ has its invariant points at zero and ∞; a point in the unit disc moves on one of the concentric circles of family B, as illustrated in figure 4(*b*).

In figure 6, the boost $B_x(\mu)$ takes the point N to V on the unit hyperboloid, while the rotation $R_z(\phi)$ rotates the point B counterclockwise around the vertical t-axis labelled OT. (It also rotates the projected point L on a circle which is the intersection of the unit sphere with a horizontal plane through KL.) Under either operation the norm of the position vector $x^\mu x_\mu = t^2 - x^2 - y^2 = 1$ is preserved.

The Kard representation of the transfer matrix, equation (6), can also be expressed (in the symmetric case) as

$$\begin{aligned} M &= \mathrm{e}^{(\mu/2)\sigma_x}\,\mathrm{e}^{-\mathrm{i}\phi\sigma_z}\,\mathrm{e}^{(-\mu/2)\sigma_x} \\ &= B_x(\mu)R_z(2\phi)B_x(-\mu). \end{aligned} \tag{24}$$

Geometrically, the first factor on the right is a reverse boost, which moves the vector SFV to point along the vertical axis. The R_z then rotates the system by angle 2ϕ around the vertical axis. Finally the boost transforms the vertical axis back to lie along SFV. The vector SFV is invariant under this sequence of operations, so it is analogous to an axis of rotation. Vectors close to SFV move on an orbit around that direction, remaining on the hyperboloid. One can show that the orbit is planar, and the plane is parallel to the tangent plane (to the hyperboloid) at the point V.

The elementary transforms utilized by Yonte *et al* [15] are, in our language,

$$\begin{aligned} K(\phi) &= R_z(-\phi) \\ A(\xi) &= B_y(-\xi) \\ N(\nu) &= \lim_{\phi\to 0} B_y(\mu)R_z(-\phi)B_y(-\mu) \\ &= 1 + (\nu/2)[\sigma_x - i\sigma_z] \end{aligned} \tag{25}$$

with the limit of $\phi\,\mathrm{e}^{\mu} \equiv \nu$. Their $N(\nu)$ therefore corresponds to scattering at a band edge. Although they do not include the boost along OX, the operation $N(\nu)$ does involve motion of the mapped point with a component along the x direction.

5.2. Stereographic projection

As discussed in section 2, a time-reversal symmetric scattering process can be visualized geometrically as a Möbius transform in the equatorial plane. It can also be interpreted as a Lorentz transform of a Dirac spinor in 2 + 1 dimensions, as discussed in appendix A. Such a transform is associated with a Lorentz transform of the coordinates in (t, x, y) space. Under a Lorentz transform, the north pole moves to a point on the upper sheet of the unit hyperboloid. The unit hyperboloid, also called the pseudosphere, plays an equivalent role to the unit sphere in the case of rotations R_3. The relation between motion on a hyperbolic surface and the bilinear mapping in the plane has a rich history, going back to Poincaré and Klein; it is discussed in a lucid review by Balazs and Voros [28].

The two geometrical interpretations are related by stereographic projection from the south pole as described in appendix B. Here we restrict the discussion to reflection symmetric potentials, and work in the approximation of a more or less constant μ, and $\cos\phi$ which varies linearly across an allowed band.

In figure 5 we illustrate the stereographic projection by a perspective view. A Kard circle in the equatorial plane surrounds the invariant point at $a = \tanh(\mu/2)$; the centre $x_c' < a$ (labelled by C) is close by. A cone with apex at the south pole S through the circle projects it up onto an ellipse on the unit hyperboloid. This ellipse is the intersection of a plane (white) with the upper sheet of the hyperboloid. If the radius of the Kard circle were shrunk to zero, $x_c' \to a$, the plane would become the tangent plane to the hyperboloid at the point where the invariant line intersects it. For any radius of the Kard circle, the planes are parallel. The projection cone also intersects the unit sphere in a circle, whose centre is on OR. This circle is the intersection of a plane perpendicular to OR, with the sphere. However, if the radius of the Kard circle in the equatorial plane is varied, these planes are *not* parallel (as they would be if the motion were the precession of a Pauli spinor about a fixed field direction). That explains why the interpretation of the transfer matrix is an analogy to the transformation of a Dirac spinor, and not a Pauli spinor.

In figure 6 we see a vertical section of figure 5 through the OX axis. The invariant point $z = a$ of the Möbius transform, a distance $\tanh(\mu/2)$ from the origin projects up to the point V with coordinates $(t = \cosh\mu, x = \sinh\mu, y = 0)$ on the unit hyperboloid. The vector SV from the south pole is the invariant axis of the Lorentz transform. Orbits with different initial points on the unit hyperboloid are conic sections defined by the intersection of a plane with the hyperboloid. The orbit of zero radius is the point where the invariant axis intersects the hyperboloid. The tangent plane touches the hyperboloid at this point, and all the other orbits lie on planes parallel to this one. The corresponding orbits in the equatorial plane are circles of family B lying within the unit circle.

For finite μ, the orbits are ellipses, which in the limit $\mu \to \infty$ become parabolae. This limit corresponds to scattering at a band edge. Moving into a forbidden zone, μ is replaced by $\xi + \mathrm{i}\pi/2$; the plane is then inclined at more than $\pi/4$ to the horizontal, and the orbits become hyperbolae. The fixed points of the Möbius transformation move onto the unit circle; they are the projections of the asymptotes of the hyperbolae. The corresponding orbits in the equatorial plane are circles which extend outside the unit circle; only the inside arc is physical.

The line connecting a point on the hyperboloid to the south pole S passes through the unit sphere. For a given ellipse on the hyperboloid, the orbit on the sphere is a circle, defined by the intersection of a plane with the unit sphere. For the orbit of zero radius, the centre of the circle is at polar angle θ, where $\tan(\theta/2) = \tanh(\mu/2)$. The motion on the sphere can be interpreted as the precession of a Pauli spin-half system in a magnetic field inclined at angle θ, when the spin is 'up' along that direction. The only weak point of this analogy is that when

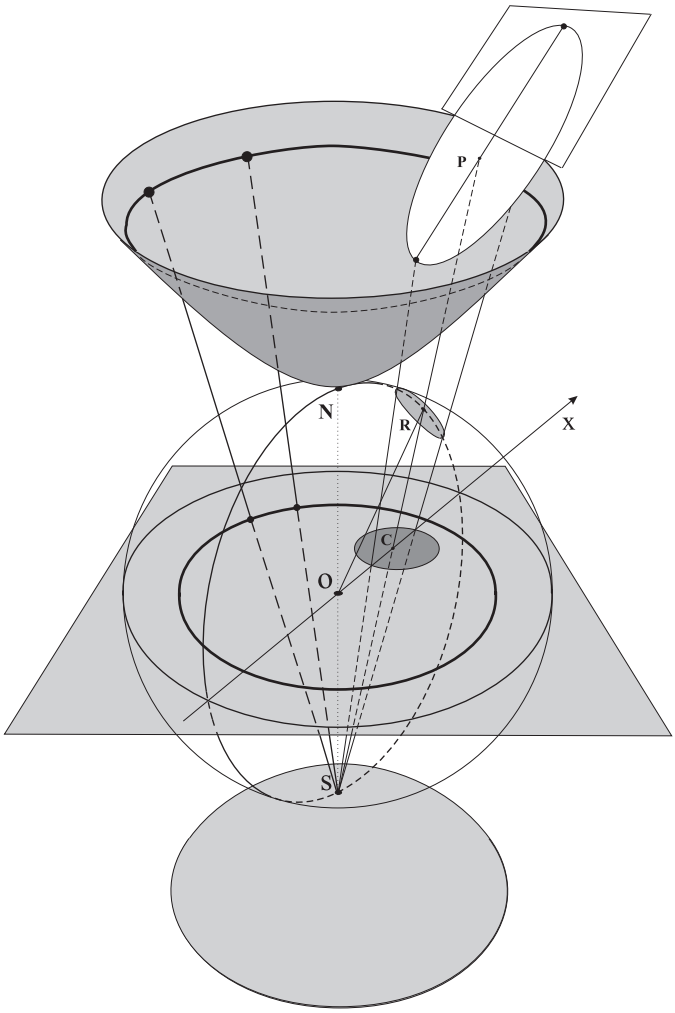


Figure 5. Perspective view of stereographic projection, showing the unit sphere, the equatorial circle, and parts of the unit hyperboloid. Hidden portions of lines are dashed. A Kard circle in the plane (dark shaded) is projected up onto a circle formed by a plane (dark shaded) intersecing the unit sphere, and onto an ellipse formed by a plane (white) intersecting the unit hyperboloid. See text for further discussion.

we consider a different orbit on the hyperboloid, all of which belong to a family of parallel planes, the corresponding orbits on the sphere do *not* have a common axis of rotation.

However, if we restrict our attention to orbits which are smaller than the principal orbit (meaning that the initial point is to the right of the origin), the centres of the circles are displaced by only a small amount, (between $0.5 \tanh \mu$ and $\tanh \mu/2$), and the analogy remains qualitatively correct, though wrong in detail. This is clear from the discussion in appendix B, as the axis of the principal orbit, OD, is not far from the axis of the infinitesimal orbit OU.

Our initial interest in this problem arose from the design of (ARCs), and in this application the most relevant orbits do lie within the principal orbit. Fortunately, our work in [14] did not rely on the analogy to spin precession in more than a qualitative and pictorial sense: all the equations were worked out for the hyperbolic geometry of Lorentz transformations on the unit

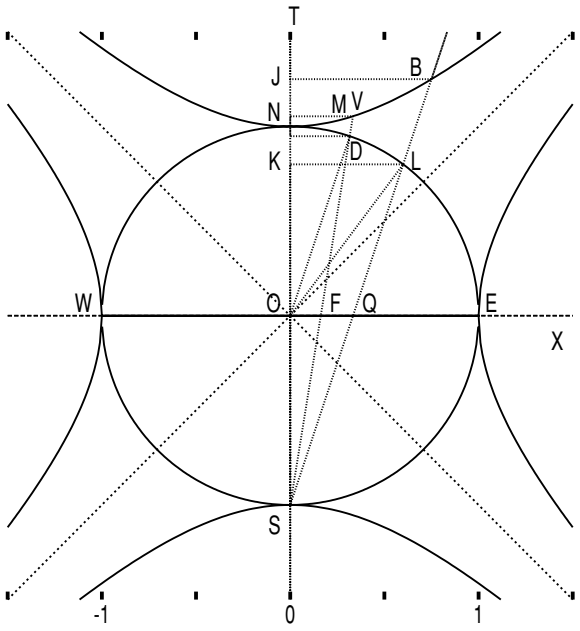


Figure 6. Stereographic projection from the unit hyperboloid to unit sphere and to unit circle in (t, x, y) space. A section through the TX plane is shown.

hyperboloid. The orbits should therefore have been drawn as ellipses on a hyperbolic surface rather than circles on a sphere, but with the same initial and Bragg points. In the small angle limit it makes no difference. The other point that needs revision in [14] is the statement that the boost $B_x(\mu)$ acts in the TY-plane. We should have drawn the centres of the Kard orbits on the OX axis. This does not affect any of the calculations.

For asymmetric potential cells, there is a third Kard parameter χ; see equation (6). The difference from the $\chi = 0$ case is that the geometry is rotated by azimuthal angle χ away from OX. If there are several potential cells in sequence, with the same χ, then one rotation takes care of them all. But combining cells with different χ becomes more complicated, because the fixed points of successive cells do not lie on the same ray from the origin. The advantage of dealing entirely with reflection symmetric cells is that the boosts are all in the TX-plane, at least in allowed zones.

6. Conclusion

We have interpreted the action of a transfer matrix on a scattering wavefunction geometrically in two ways. Firstly as a conformal mapping on the unit disc, and secondly as a Lorentz transform on the unit hyperboloid in t, x, y space. These are related by stereographic projection from the south pole. For an arbitrary shaped potential cell, the transfer matrix can be parametrized in the Kard form with three parameters (ϕ, μ, χ). Two polar coordinates, μ, χ, define a point on the unit hyperboloid, and 2ϕ is an 'angle of rotation' on an elliptical orbit around that direction.

For electrons passing through a layered system, we have previously found that over most of an allowed band, it is a good approximation to consider the direction fixed, while ϕ is free

to vary. Then the orbits are coaxal circles in the equatorial plane, and ellipses on the unit hyperboloid. This picture was key to the design of anti-reflection coatings in [14].

Acknowledgments

We are grateful to NSERC-Canada for Discovery grant SAPIN-3198 (DWLS, GM) and to DGES-Spain for continued support through grant BFM2001-3710 (JM). This work was carried out as part of CERION-2, EU thematic network IST-2001-39059.

Appendix A. Dirac equation and Lorentz invariance

We use tensor notation, following Spain [29]. Time and space $\mathbf{r} = (x, y, z)$ are combined into a contravariant position 4-vector

$$\begin{aligned} &x^0 = ct, x^k = r_k \qquad && k = 1, 2, 3 \\ &x^\mu = (ct, \mathbf{r}) && \mu = 0, \ldots, 3. \end{aligned} \tag{A.1}$$

Greek indices run over values $0, \ldots, 3$ while Latin indices are restricted to spacelike values. The diagonal metric tensor $g_{\mu\nu} = g^{\mu\nu}$ with $g_{00} = 1$, $g_{kk} = -1$ can be used to raise and lower indices. The summation convention will hold for repeated indices. The norm of a vector is

$$\begin{aligned} |x|^2 &= g_{\mu\nu} x^\mu x^\nu = x^\mu x_\mu \\ &= c^2t^2 - x^2 - y^2 - z^2. \end{aligned} \tag{A.2}$$

The summed product $g_{\mu\nu}x^\nu$ is the covariant position vector with lowered index: $x_\mu = (ct, -\mathbf{r})$.

The covariant 4-momentum operator is introduced by

$$p_\mu = \mathrm{i}\hbar \frac{\partial}{\partial x^\mu} = (E/c, -\mathbf{p}) \tag{A.3}$$

in terms of which the Dirac equation [27] reads

$$(\alpha^\mu p_\mu - mc\beta)|\psi\rangle = 0. \tag{A.4}$$

Here $\alpha^0 = 1$, while α^k and β are the usual Dirac matrices. Providing that they satisfy anti-commutation relations

$$\begin{aligned} &\{\alpha^k, \alpha^m\} = 2\delta_{km} \\ &\{\alpha^k, \beta\} = 0 \end{aligned} \tag{A.5}$$

multiplication of equation (A.4) by $(p_0 - \alpha^k p_k + \beta mc)$ shows that each component of the wavefunction $|\psi\rangle$ is a solution of the Klein–Gordon equation, and is consistent with the energy–momentum relation of special relativity:

$$(p_0 - \alpha^k p_k + \beta mc)(p_0 + \alpha^r p_r - \beta mc) = (p_0)^2 - \mathbf{p}\cdot\mathbf{p} - m^2c^2 = 0. \tag{A.6}$$

Since the order of the factors can be reversed, the plane-wave solutions are the columns of the matrix

$$(p_0 - \alpha^k p_k + \beta mc) = \begin{pmatrix} E/c + mc & \sigma\cdot\mathbf{p} \\ \sigma\cdot\mathbf{p} & E/c - mc \end{pmatrix} \tag{A.7}$$

in 2×2 block form, when we adopt the simple Dirac representation of the Dirac matrices.

The 4-current vector $j^\mu = c\langle\psi|\alpha^\mu|\psi\rangle$ satisfies the equation of continuity

$$\frac{\partial j^\mu}{\partial x^\mu} = 0 \tag{A.8}$$

where

$$j^\mu = (c\rho, \mathbf{j}).$$

It should be emphasized that the α^μ are *not* components of a vector, rather they are spinoperators that are the same at all points in space. On the other hand, the components of the 4-current are observables, the probability density and probability flux density, so under a change of the coordinate axes they must transform like a 4-vector.

A Lorentz transform Λ of the coordinate system is written

$$\begin{aligned} x'^\mu &= \Lambda^\mu{}_\nu x^\nu \\ x'_\mu &= g_{\mu\sigma}\Lambda^\sigma{}_\nu x^\nu = g_{\mu\sigma}\Lambda^\sigma{}_\nu g^{\nu\tau}x_\tau \\ &= \Lambda_\mu{}^\tau x_\tau. \end{aligned} \tag{A.9}$$

The norm of the vector is unchanged providing that

$$\begin{aligned} x'^\mu x'_\mu &= \Lambda^\mu{}_\nu x^\nu \Lambda_\mu{}^\tau x_\tau \\ &= \left[\Lambda^\mu{}_\nu \Lambda_\mu{}^\tau\right] x^\nu x_\tau = \delta_\nu{}^\tau x^\nu x_\tau. \end{aligned} \tag{A.10}$$

This says that the matrix Λ describing the Lorentz transform, becomes its inverse when the character of the indices is exchanged, and the matrix is transposed (i.e. the columns are orthogonal vectors). The inverse transform therefore reads

$$\begin{aligned} x_\mu &= x'_\nu \Lambda^\nu{}_\mu \\ x^\mu &= x'^\nu \Lambda_\nu{}^\mu. \end{aligned} \tag{A.11}$$

Under a Lorentz transform, equation (A.4) becomes

$$(\alpha^\mu \Lambda^\nu{}_\mu p'_\nu - \beta mc)|\psi\rangle = 0$$

rather than

$$(\alpha^\mu p'_\mu - \beta mc)|\psi'\rangle = 0. \tag{A.12}$$

Pauli's fundamental theorem states that we can always find a transformation $M(\Lambda)$ which acts on the spin degrees of freedom, with the property

$$\begin{aligned} M^\dagger \alpha^\mu M &= \Lambda^\mu{}_\nu \alpha^\nu \\ M^\dagger \beta M &= \pm\beta. \end{aligned} \tag{A.13}$$

(The + sign applies for orthochronous transforms, which preserve the sense of time.) Then, under the transformation $|\psi'\rangle = M|\psi\rangle$, the two forms in equation (A.12) are equivalent. Hence we can solve the Dirac equation using the *same* α-matrices in any frame of reference. Essentially, it says that if you apply the Lorentz transform both to the coordinates and to the spin degrees of freedom, the equation is invariant.

The final step is to check that the 4-current, calculated in the new frame from the new solutions, is indeed the Lorentz transform of the current calculated in the original frame:

$$\begin{aligned} j'^\mu &= c\langle\psi'|\alpha^\mu|\psi'\rangle = c\langle\psi|M^\dagger\alpha^\mu M|\psi\rangle \\ &= \Lambda^\mu{}_\nu c\langle\psi|\alpha^\nu|\psi\rangle = \Lambda^\mu{}_\nu j^\nu \end{aligned} \tag{A.14}$$

and according to equation (A.9) this is the desired result.

The lesson is that associated with each Lorentz transform on the coordinates, there is a related Lorentz transform on the spin degrees of freedom. This not only transforms the solution $|\psi\rangle$, but it also induces a bilinear transformation on the spin operators, equation (A.13).

As examples, Dirac [27] considered the elementary transforms which act in one plane of the 4-space, labelled by μ, ν. There are six possibilities, three boost operations in the $0, k$

planes and three real rotations involving the k, m planes. He showed that the corresponding operator with $\mu < \nu$ is

$$M(\mu, \nu) = \mathrm{e}^{+\frac{1}{2}\omega\alpha^{\nu}\alpha^{\mu}} \tag{A.15}$$

where ω is the 'angle of rotation'. In the case of a boost along OX,

$$M(0, x) = \mathrm{e}^{\frac{1}{2}\omega\alpha^{x}} = M^{\dagger}. \tag{A.16}$$

One easily sees that, because of the anti-commutation relations, $\alpha^x f(\alpha^k) = f(-\alpha^k)\alpha^x$ for $k = y$ or z, so only α^0 and α^x are affected. For them,

$$\begin{aligned} M^{\dagger}\alpha^0 M &= \cosh\omega\,\alpha^0 + \sinh\omega\,\alpha^x \sim \Lambda^0{}_{\nu}\alpha^{\nu} \\ M^{\dagger}\alpha^x M &= +\sinh\omega\,\alpha^0 + \cosh\omega\,\alpha^x \end{aligned} \tag{A.17}$$

The corresponding transformation of the coordinates (leaving x and y invariant) is

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh\omega & \sinh\omega \\ \sinh\omega & \cosh\omega \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \sim \Lambda^{\mu}{}_{\nu} x^{\nu} \tag{A.18}$$

which takes a particle from rest to motion at relative velocity $v/c = \tanh\omega$, starting at $t = 0$. (For consistency with the convention for rotations in quantum mechanics, where the 'active' viewpoint is preferred, we have chosen the sign so that for small velocities, $x' \sim x + vt, t' \sim \gamma(t + vx/c^2)$, where $\gamma^2 = 1/(1 - v^2/c^2)$. A particle at rest in the moving frame shows an increasing new coordinate x' due to the motion at velocity v.)

Correspondingly, when both indices are spacelike, a rotation in the x, y plane has

$$M(x, y) = \mathrm{e}^{-\frac{1}{2}\omega\alpha^{x}\alpha^{y}} = \mathrm{e}^{-\frac{1}{2}\omega i\Sigma_z}. \tag{A.19}$$

In the Dirac representation, Σ_z has the Pauli spin-matrix on the diagonal in both blocks. Unlike the boost example, here $M^{\dagger} = M^{-1}$ is unitary. Because there are *two* α's in the exponent, both α^0 and α^z are unaffected. One finds

$$\begin{aligned} M^{\dagger}\alpha^x M &= \cos\omega\,\alpha^x - \sin\omega\,\alpha^y \sim \Lambda^x{}_{\nu}\alpha^{\nu} \\ M^{\dagger}\alpha^y M &= \sin\omega\,\alpha^x + \cos\omega\,\alpha^y \end{aligned} \tag{A.20}$$

corresponding to a counterclockwise rotation of the system by angle ω, in the 'active' viewpoint. In principle, an arbitrary transform can be accomplished by a finite number of elementary transforms of the above type, and the corresponding product of the matrices Λ gives the total Lorentz transform [27].

For application to the potential scattering problem, we do not require the four-dimensional space of the Dirac equation in 3D. Rather we can drop one coordinate, say z, to leave x^{μ}, $\mu = 0, 1, 2$. In this case, it suffices to choose a 2×2 representation of the Dirac matrices as follows:

$$\alpha^x = \sigma_x \qquad \alpha^y = \sigma_y \qquad \beta = \sigma_z. \tag{A.21}$$

This means that the spin up/down states are correlated with the particle/hole states, even for finite mass, because of the reduced dimensionality.

The transfer matrix M can be represented as a linear combination of $\alpha^0 = 1$ plus the Pauli matrices. Using equation (A.13) it is easy to work out the elements of the Lorentz transform associated with a given transfer matrix. This in turn can be visualized as a transformation of points in the $x^{\mu} = (ct, x, y)$ space.

In the transfer matrix formalism, a scattering state is represented by a two-component spinor, acted upon by M. The state which is 'spin-up' along some direction can be obtained starting from the state which is 'up' along the ct-axis, and applying a suitable Lorentz transform to it. Correspondingly, in the 3-space, the associated direction is rotated from $(1, 0, 0)$ to a

new position (ct', x', y'). This new position lies on the unit hyperboloid of revolution passing through the north pole N in figure 6. The situation is analogous to rotation of a spin-half state in real space R_3, except that for real rotations, the motion is on the unit sphere. In either case, by sterographic projection from the south pole, the orbit can be projected onto the equatorial unit circle. For spin-half systems, the position on the unit sphere is interpreted as the mean direction for measurements of the spin. For the scattering state we can think of the position on the unit hyperboloid in a similar way. It is just a difference between spherical and hyperbolic geometry. The root of the distinction is that a spin-state is normalized so that $\langle\chi|\chi\rangle = 1$. But the scattering states of the transfer matrix formalism have a different metric, namely

$$\langle\psi|\sigma_z|\psi\rangle = |a|^2 - |b|^2 = 1. \tag{A.22}$$

The second of equations (A.13), showing the invariance of the matrix $\beta = \sigma_z$ guarantees the conservation of this normalization. The choice of the positive sign is what confines the motion to the upper sheet of the hyperboloid. Conservation of flux of particles is what rules out the time-reversed type of transfer matrix, which would take a point from the upper/lower sheeted hyperboloid, to the one (the funnel) which crosses the OX and OY axes.

Appendix B. Orbits and stereographic projection

In this appendix we describe the geometrical interpretation of the action of a transfer matrix on a spinor. Figure 6 shows a cross section of the unit sphere and unit hyperboloid in the T–X plane, the vertical axis corresponding to ct in relativity. We will set the velocity of light to $c = 1$ in this section as it has no physical signficance here. Three points labelled N, M, and B are shown on the section of the unit hyperboloid.

Under stereographic projection from the south pole S, N is related to O, and B to Q respectively in the equatorial plane. Let us consider the *principal trajectory*, which in the Kard parametrization is a circle in the complex plane centred at $x'_C = 0.5\tanh\mu$ on the x-axis, close to the letter F. The coordinates of a general point on the hyperboloid (t, x, y) are related to those in the plane $(t' = 0, x', y')$ by

$$\begin{aligned} &\frac{x}{x'} = \frac{y}{y'} = \frac{t+1}{1} \quad \text{while} \\ &t^2 - 1 = x^2 + y^2 = (t+1)^2(x'^2 + y'^2) \quad \text{giving} \\ &t = \frac{1 + x'^2 + y'^2}{1 - x'^2 - y'^2}. \end{aligned} \tag{B.1}$$

The projection B of the extremal point Q (the Bragg point, with $\phi = \pi/2$). is therefore at $(t = \cosh 2\mu, x = \sinh 2\mu, y = 0)$. The point L on the unit sphere has a polar angle $\mathrm{KOL} = 2\eta$. This is the exterior angle to the isosceles triangle OSL, whose angle $\mathrm{KSL} = \eta$.

The triangles JSB and KSL are similar, so

$$\tan\eta = \frac{\sinh 2\mu}{1 + \cosh 2\mu} = \frac{\sin 2\eta}{1 + \cos 2\eta} = \tanh\mu. \tag{B.2}$$

This shows *inter alia* that the maximum value of η is $\pi/4$, corresponding to $\mu \to \infty$, and the point B moving away towards the asymptotic cone, whose intersection with the plane of the figure is along the dotted lines. The coordinates of L are

$$\begin{aligned} &t = \cos 2\eta = \frac{1 - \tan^2\eta}{1 + \tan^2\eta} = \frac{1 - \tanh^2\mu}{1 + \tanh^2\mu} = \frac{1}{\cosh 2\mu} \\ &x = \sin 2\eta = \tanh 2\mu. \end{aligned} \tag{B.3}$$

Next, draw a line SDM which bisects the angle KSL; the triangle OSD is also isosceles and therefore the polar angle of line OD is η. It follows that OD is parallel to SL. The coordinates of D are

$$\begin{pmatrix} t \\ x \\ y \end{pmatrix}_D = \begin{pmatrix} \cos\eta \\ \sin\eta \\ 0 \end{pmatrix} = \frac{1}{\sqrt{\cosh 2\mu}} \begin{pmatrix} \cosh\mu \\ \sinh\mu \\ 0 \end{pmatrix}. \tag{B.4}$$

The stereographic projection of D is M at

$$\begin{pmatrix} t \\ x \\ y \end{pmatrix}_M = \frac{1}{\cosh\mu} \begin{pmatrix} \sqrt{\cosh 2\mu} \\ \sinh\mu \\ 0 \end{pmatrix}. \tag{B.5}$$

Close and to the right of M on the hyperbola is another point V (not marked), which is the extension of the line OD and has coordinates $(\cosh\mu, \sinh\mu, 0)$. Interestingly V is also the stereographic projection of the fixed point F of the Möbius transform in the equatorial plane, at $x_F = \tanh(\mu/2)$. F lies to the right of the centre, $x'_C = 0.5\tanh\mu$, of the principal trajectory.

The line SDM intersects the equatorial plane at the projection of M; this point has a coordinate $x' = \sinh\mu/(\cosh\mu + \sqrt{\cosh 2\mu}) < x'_C < x_F$.

For a given Kard transfer matrix with fixed μ, the orbits are circles of family B in the equatorial plane, which enclose the fixed point F. At one extreme, the orbit of zero radius is located at F; its projection on the hyperboloid is the point V.

The principal orbit starts at the origin, reaches Q when $\phi = \pi/2$, and then continues back to O. Projected onto the hyperboloid it runs from N to B $= (\cosh 2\mu, \sinh 2\mu, 0)$, and then continues back to N. Both descriptions of the orbit lie on a cone with apex at S. The cone intersects the equatorial plane in a circle, and the hyperboloid in an ellipse. Surprisingly, this ellipse is the intersection of the hyperboloid with a plane, perpendicular to the page, and passing through the line BN. The normal to this plane is $\hat{n}_H = (\cosh\mu, -\sinh\mu, 0)/\sqrt{\cosh 2\mu}$, which is also the normal to the tangent to the hyperbola at the point V (the projection from the fixed point F). Indeed, all the orbits associated with the Kard parametrization are the intersections of the hyperboloid with one of a family of parallel planes having the same normal $\hat{n}_H$.

To understand the analogy to spin, we also consider the point L on the unit sphere. The cone associated with the principal orbit, intersects the unit sphere in a curve passing through N and L. Again this curve is planar, and therefore is a circle. To prove this we must carry out the projection for a general point G on the orbit.

The line NL defines a plane perpendicular to the t, x-plane, which in turn cuts the sphere in a circular orbit. The task is to prove that any point on the true orbit projects into a point G on this circular orbit. Similar to equation (B.1), a point $(0, x', y')$ in the plane projects onto the unit sphere with

$$\begin{aligned} &\frac{x}{x'} = \frac{y}{y'} = \frac{t+1}{1} \qquad \text{while} \\ &t^2 - 1 = -(x^2 + y^2) = -(t+1)^2(x'^2 + y'^2) \quad \text{giving} \\ &t = \frac{1 - x'^2 - y'^2}{1 + x'^2 + y'^2}. \end{aligned} \tag{B.6}$$

(Note that this t is the inverse of that of equation (B.1).) For a point on the principal orbit, $(x'^2 + y'^2) = x'\tanh\mu = x'\tan\eta$. From this one quickly finds that at G, $t = 1 - x\tan\eta$, and the vector NG = $(-x\tan\eta, x, y)$ is perpendicular to OD, which is what we had to prove. One can therefore regard the motion on the sphere as a precession around the axis OD, like a spin precessing around a magnetic field.

In a similar manner, using equation (B.1), one finds that for the principal orbit projected up onto the unit hyperboloid, the general point G has position $\mathrm{OG} = (t, x, y)$ with $t = 1+x\tanh\mu$. Then $\mathrm{NG} = (x\tanh\mu, x, y)$ is perpendicular to the normal $\hat{n}_H = (\cosh\mu, -\sinh\mu, 0)$ to the tangent at the point V.

If we increase the size of the orbit, the orbit on the hyperboloid is always planar and the planes are all parallel for a given μ. The extreme points of the ellipse move further to the left and the right. In the extreme limit, they both approach the asymptotes. The corresponding projected orbit on the sphere becomes the equatorial circle in that limit, with its axis of rotation on the line ON. It is clear from this that the direction in space defining the axis of rotation is not fixed for the orbit on the sphere, unlike the fixed direction which defines the orbits on the hyperboloid. The analogy to spin precession is flawed.

The infinitesimal orbit is another one which shows that the axis of rotation varies with the size of the orbit. We have already remarked that the point V on the hyperboloid is the image of the fixed point F. The line SFV intersects the unit sphere at a point U, not indicated, which lies just to the right of D. The corresponding axis of rotation, OU, has a polar angle θ to the vertical. This angle θ is exterior to the isosceles triangle OSU, showing that $\tan(\theta/2) = \tanh(\mu/2)$. The plane of the infinitesimal orbit at U has slope $-\tan\theta = -\sinh\mu$. This slope is more negative than that of the chord NL, namely $-\tanh\mu$, because U must lie to the right of D. Indeed, in the limit $\mu \to \infty$, the point U moves to E, with infinite negative slope, while the chord NE would be inclined at 45°.

For finite μ the orbits on the hyperboloid are ellipses. As $\mu \to \infty$, the orientation of the planes approaches $\pi/4$ and in the limit the orbits become parabolae. This situation corresponds to scattering at a band edge. Both fixed points in the plane approach +1, and the planar orbits are circles tangent to the unit circle at the double fixed point.

Moving into a forbidden zone, the fixed points are complex conjugates, on the unit circle. The tilt of the plane defining the orbits on the hyperboloid now exceeds $\pi/4$, so the orbits become hyperbolae. The branch on the upper sheet projects in the equatorial plane onto an arc of a circle (i.e. the portion lying inside the unit circle) and enclosing one of the fixed points. The family of parallel planes defines the various possible orbits of different sizes. The particular case $\xi = 0$ corresponds to vertical planes, one of which passes through ON and gives a straight line orbit in the plane; the fixed points lie at $z = \pm\mathrm{i}$.

References

[1] Sprung D W L, Wu H and Martorell J 1993 *Am. J. Phys.* **61** 1118–24
[2] Yakubovich V A and Starzhinskii V M 1975 *Linear Differential Equations with Periodic Coefficients* vol 1 English edn (New York: Wiley) chapter 2
[3] Titchmarsh E C 1958 *Eigenfunction Expansions Associated with Second-order Differential Equations* vol 2 (Oxford: Oxford University Press) chapter 21
[4] Kalotas T M and Lee A R 1992 *Phys. Scr.* **45** 552–9
[5] Kalotas T M and Lee A R 1995 *Eur. J. Phys.* **16** 119–23 and references therein
[6] Griffiths D J and Steinke C A 2001 *Am. J. Phys.* **69** 137–54
[7] Weber T A 1994 *Solid State Comm.* **90** 713–6
[8] Sprung D W L, Sigetich J D, Jagiello P and Martorell J 2003 *Phys. Rev.* B **67** 085318 (11 pp)
[9] Bendickson J M, Dowling J P and Scalora M 1996 *Phys. Rev.* E **53** 4107–21
[10] Bulgakov S A and Nieto-Vesperinas M 1998 *J. Opt. Soc. Am.* A **15** 503–10
[11] Tsai Y C, Shung K W and Gou S C 1998 *J. Mod. Opt.* **45** 2147–54
[12] Yeh P, Yariv A and Hong T K C 1977 *J. Opt. Soc. Am.* **67** 423–38
[13] Yeh Pochi 1988 *Optical Waves in Layered Media* (New York: Wiley)
[14] Sprung D W L, Morozov G V and Martorell J 2003 scattering and spin precession *J. App. Phys.* **93** 4395–406
[15] Yonte T, Monzón J J, Sánchez-Soto L L, Cariñena J F and López-Lacasta C 2002 *J. Opt. Soc. Am.* A **19** 603
[16] Monzón J J, Yonte T, Sánchez-Soto L L and Cariñena J F 2002 *J. Opt. Soc. Am.* A **19** 985

[17] Georgieva E and Kim Y S 2001 *Phys. Rev.* E **64** 026602 (6 pp)
[18] Başkal S and Kim Y S 2003 *Phys. Rev.* E **67** 056601 (8 pp)
[19] Nelson D F, Miller R C and Kleinman D A 1987 *Phys. Rev.* B **35** 7770–3
[20] Phillips E G 1957 *Functions of a Complex Variable* 4th edn (Edinburgh: Oliver and Boyd) chapter II
[21] Kard P G 1957 *Opt. Spektrosk.* **2** 236–44
[22] Chang C C and Kuo C S 1999 *J. Phys. D: App. Phys.* **32** 139–46
[23] Kober H 1957 *Dictionary of Conformal Representations* (New York: Dover) pp 1–34
[24] Morozov G V, Sprung D W L and Martorell J 2002 *J. Phys. D: Appl. Phys.* **35** 3052–9
[25] Gomez I, Dominguez-Adame F, Diez E and Bellani V 1999 *J. Appl. Phys.* **85** 3916–8
[26] Monzón J J and Sánchez-Soto L L 2002 *Eur. J. Phys.* **23** 1–9
[27] Dirac P A M 1947 *Quantum Mechanics* 3rd edn (Oxford: Oxford UP University Press) chapter XI
[28] Balazs N L and Voros A 1986 Chaos on the Pseudosphere *Phys. Rep.* **143** 109–240 chapter II
[29] Spain B 1960 *Tensor Calculus* 3rd edn (Edinburgh: Oliver and Boyd) chapter I