Relativistic velocity space, Wigner rotation and Thomas precession

John A. Rhodes*

Department of Mathematics, Bates College, Lewiston, Maine 04240

Mark D. Semon†

Department of Physics and Astronomy,

Bates College, Lewiston, Maine 04240

(Dated: September 21, 2003)

We develop a relativistic velocity space called *rapidity space* from the single assumption of Lorentz invariance, and use it to visualize and calculate effects resulting from the successive application of non-colinear Lorentz boosts. In particular, we show how rapidity space provides a geometric approach to Wigner rotation and Thomas precession in the same way that spacetime provides a geometrical approach to kinematic effects in special relativity.

I. INTRODUCTION

The most commonly used Lorentz transformation relates an inertial frame S to an inertial frame S' moving with a velocity $\vec{v}$ along the x-axis of S. If this Lorentz transformation also preserves the orientation of the spatial axes of S and leaves the sign of its time component unchanged (as is usually the case), then it is called a *boost*. In some cases an inertial frame S' is obtained from an inertial frame S by *two successive* boosts. If the two successive boosts are non-colinear then, contrary to what one might expect, the single Lorentz transformation that is their resultant is not a pure boost but rather is the product of a boost and a rotation. The unexpected rotation was discovered by Thomas[1] in 1926, and derived thirteen years later by Wigner[2] in his seminal article on representations of the Lorentz group. If successive non-colinear boosts return the spatial origin of S' to the spatial origin of S, then all of the Thomas-Wigner rotations along the way combine to produce a net rotation of S' with respect to S called the *Thomas precession.*[3–8]

Thomas precession is an essential part of quantum courses discussing relativistic correc-

tions to the Hamiltonian of a hydrogen atom because it changes the non-relativistic form of the spin-orbit term by a factor of one-half. Rather than derive this result, however, some texts state it without giving any references,[9] some state it and reference only Thomas' original article,[10] while others state it and appeal to the Dirac equation for its justification.[11,12]

The few texts and journal articles which do derive Thomas precession often use mathematics that is somewhat sophisticated,[13] such as infinitesimal generators of the Lorentz group,[4] "a weakly associative-commutative groupoid,"[14] "gyrogroups and gyrovector spaces,"[15] the Gibbs method for adding finite rotations,[16] holonomy group transformations and Clifford-Dirac algebra,[17] the tetrad formalism,[18] Fermi-Walker transport,[19] or unboosted Fermi-Walker frames,[20]. Although several texts do present straightforward algebraic derivations, they are long and somewhat tedious.[5,6] Recalling that many relativistic effects are easier to derive and understand when treated geometrically in spacetime, we wondered if there was a relativistic velocity space in which Thomas precession (and other effects involving successive non-colinear boosts) could be treated geometrically and in this way also made easier to understand.

Perhaps the most intriguing approach to constructing a relativistic velocity space was mentioned in the 1950's by Landau and Lifshitz.[21] They begin an exercise for the reader by noting that given two non-colinear relativistic velocities $\vec{v}$ and $\vec{v} + d\vec{v}$, the relative velocity $d\vec{v}$ can be considered as a line element in a three-dimensional velocity space in which each point is specified by the azimuthal and polar angles of $\vec{v}$ and a radial coordinate equal to a function of v called the *rapidity*. Landau and Lifshitz then ask the reader to show that this relativistic velocity space is non-Euclidean, with a hyperbolic geometry.

Landau and Lifshitz do not reference the origin of this exercise, so it is not clear if they discovered the velocity space themselves or are drawing upon work published previously in the Russian literature. Both Pauli[22] and Rosenfeld[23] credit a paper written in Russian by the Croatian mathematician Variçak as the first place in which relativistic velocity addition was related to the analog of vector addition in a hyperbolic space. Pauli cites four additional articles (also written in Russian) by Variçak, published between 1910 and 1919, and Rosenfeld notes that Variçak summarized and expanded upon his work in a book (written in Russian) published in 1924. Rosenfeld cites only two references to work on this subject that appeared after 1924, both of which are books written in Russian. One was published in 1963 and the other in 1965.

Given this history, it seems likely that Landau and Lifshitz's text was the first written in English to mention a relativistic velocity space with hyperbolic geometry. Indeed, in 1997, when Aravind[24] showed how the Thomas-Wigner rotation and Thomas precession had properties identical to those of areas in a hyperbolic space, he credited this discovery to "the crucial hint ... from Landau and Lifshitz...." More recently, Criado and Alamo[25] chose a hyperboloid in spacetime to represent a relativistic velocity space, which they mapped onto a unit disk with hyperbolic geometry (called the *Poincaré disk*). They then drew on results from non-Euclidean geometry, such as the law of cosines and the equations of geodesics in a hyperbolic space, to show how certain properties of hyperbolic triangles correspond to certain properties of relativistic velocities and velocity addition.

The interesting results in Refs. 24 and 25 are not readily accessible to many physicists because they assume a familiarity with formulas and theorems from non-Euclidean geometry. Furthermore, although these articles make the connection between relativistic velocity addition and hyperbolic geometry compelling, neither explains this connection nor develops it systematically from first principles.

The purpose of this paper is to derive a relativistic velocity space (called *rapidity space*) from first principles, and to demonstrate how it provides a geometric approach to solving problems involving the relativistic addition of non-colinear velocities and successive, non-colinear, Lorentz boosts. The development is self-contained and assumes no previous knowledge of hyperbolic geometry. Beginning with the single requirement of Lorentz invariance, we construct rapidity space using an approach that parallels the one used to establish spacetime. We find that just as many kinematic effects in special relativity are more easily and elegantly understood once the spacetime metric

$$ds^2 = dx^2 + dy^2 - c^2dt^2 \tag{1}$$

is established (in a spacetime with two spatial and one time dimension), so too are many aspects of the addition of non-colinear boosts more easily understood once the rapidity space metric

$$ds^2 = \left(\frac{2}{1-x^2-y^2}\right)^2 (dx^2 + dy^2) \tag{2}$$

is established (with x and y related to the usual components of velocity, as defined in Sec. V). In particular, once the main properties of rapidity space have been developed, exact expressions for the Thomas-Wigner rotation and Thomas precession can be found

geometrically. Furthermore, working in rapidity space allows various qualitative aspects of these effects to be deduced geometrically, some of which are more difficult to prove with algebraic equations alone. Indeed, we have found relativistic velocity (rapidity) space to be as useful for understanding the relativistic addition of non-colinear velocities and Lorentz boosts as spacetime has been for understanding kinematic effects in special relativity.

As mentioned above, this paper derives a 2D relativistic velocity space (called rapidity space) from the single assumption of Lorentz invariance. Although developing the material in this way is logically satisfying, and has the additional benefit of unifying (and occassionally correcting) results presented previously, it does make the paper rather long. Those who would rather bypass the derivations and proofs, and simply accept that there is a relativistic velocity space whose metric is given by Eq. (49c) and whose geodesics are the ones described at the end of Sec. VI, can proceed directly to the applications presented in Sec. VII. Sections II through VI present the proofs and derivations needed to establish the relativistic velocity space and its properties, while Sec. VIII is included for those interested in how some of what is presented here is expressed using group theory, quaternions, spinors, etc.

II. NOTATION AND BACKGROUND

Consider two inertial frames S and S' whose origins are coincident when $t = t' = 0$, and whose x and x' axes are aligned. (Two inertial frames related in this way are said to be in the *standard configuration*, and it is easy to show that the linearity of Lorentz transformations always makes this choice possible for any two inertial frames S and S'.[26]) As mentioned in the Introduction, a non-trivial Lorentz transformation from S to an inertial frame S' moving with a velocity $\vec{v}$ with respect to S is called a boost if it preserves the orientation of the spatial axes and leaves the sign of the time component unchanged. If the boost is in the x-direction, then the transformation equations are

$$x' = \gamma(x - vt), \tag{3a}$$

$$y' = y, \tag{3b}$$

$$z' = z, \tag{3c}$$

$$t' = \gamma\left(t - \frac{vx}{c^2}\right), \tag{3d}$$

with $\gamma = 1/\sqrt{1 - v^2/c^2}$. In what follows, we restrict ourselves to spacetimes with one time and two spatial dimensions because this is sufficient for understanding the most common cases of Thomas rotation and precession.[27]

For convenience, let $x_1 = x$, $x_2 = y$, $x_3 = ct$, and $\beta = v/c$. Using this notation, $\gamma = (1 - \beta^2)^{-\frac{1}{2}}$, and Eq. (3) becomes

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & -\gamma\beta \\ 0 & 1 & 0 \\ -\gamma\beta & 0 & \gamma \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{4}$$

If we let $\boldsymbol{x}$ represent the column matrix on the right-hand side of Eq. (4), then the length (norm) squared of $\boldsymbol{x}$ can be expressed as

$$x_1^2 + x_2^3 - x_3^2 = \boldsymbol{x}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{G} \boldsymbol{x}. \tag{5}$$

More generally, any linear transformation $\boldsymbol{\Lambda}$ is a Lorentz transformation if and only if it leaves the spacetime metric

$$\boldsymbol{x'}^T \boldsymbol{G} \boldsymbol{x'} = \boldsymbol{x}^T \boldsymbol{G} \boldsymbol{x} \tag{6}$$

invariant for all $\boldsymbol{x}$, or equivalently, if and only if

$$\boldsymbol{\Lambda}^T \boldsymbol{G} \boldsymbol{\Lambda} = \boldsymbol{G}. \tag{7}$$

III. RELATIVISTIC VELOCITY ADDITION AND THE RAPIDITY

The *rapidity* ϕ of a boost $\vec{\beta}$ is defined as

$$\phi \equiv \operatorname{arctanh} \beta. \tag{8}$$

Thus,

$$\beta = \tanh \phi, \tag{9a}$$

$$\gamma = \cosh \phi, \tag{9b}$$

$$\gamma\beta = \sinh \phi. \tag{9c}$$

Using the rapidity allows Lorentz boosts to be expressed in two alternative and interesting ways. In the first, if we substitute the rapidity into Eq. (4), we obtain

$$\begin{pmatrix} x_1' \\ \\ x_2' \\ \\ x_3' \end{pmatrix} = \begin{pmatrix} \cosh\phi & 0 & -\sinh\phi \\ \\ 0 & 1 & 0 \\ \\ -\sinh\phi & 0 & \cosh\phi \end{pmatrix} \begin{pmatrix} x_1 \\ \\ x_2 \\ \\ x_3 \end{pmatrix}, \tag{10}$$

which illustrates that the rapidity can be interpreted as an imaginary rotation angle in spacetime.

A second way of expressing Lorentz boosts is found by introducing the new coordinates (ξ, η), with[28]

$$\xi \equiv x_3 + x_1 \quad \text{and} \quad \eta \equiv x_3 - x_1. \tag{11}$$

If we use these coordinates, Eq. (10) can be expressed in the simple form

$$\xi' = e^{-\phi}\xi \tag{12a}$$

$$\eta' = e^{\phi}\eta. \tag{12b}$$

For future reference, note that Eq. (12) also can be written in the form

$$\xi' = (\gamma - \gamma\beta)\,\xi = \sqrt{\frac{1-\beta}{1+\beta}}\,\xi \tag{13a}$$

$$\eta' = (\gamma + \gamma\beta)\,\eta = \sqrt{\frac{1+\beta}{1-\beta}}\,\eta. \tag{13b}$$

The ξ and η coordinate axes lie on the light cone ($\sqrt{x^2+y^2} = \pm ct$) and transform into themselves under this type of Lorentz boost. Expressed in another way, these axes are eigenvectors of the boost in Eq. (10) with real eigenvalues $e^{\pm\phi}$ which, as can be seen from Eq. (12) and (13), are simply the blue- and redshift factors in the relativistic Doppler effect.[29]

The rapidity is most commonly used to simplify the addition of colinear relativistic velocities. As is well known, the relativistic addition of two colinear velocities $\vec{v}_1$ and $\vec{v}_2$ gives a resultant (colinear) velocity $\vec{v}$ with magnitude

$$v = \frac{v_1 + v_2}{1 + (v_1 v_2/c^2)} \iff \beta = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2}. \tag{14}$$

The correct generalization of Eq. (14) to the relativistic addition of n colinear velocities is not obvious. However, if we reexpress Eq. (14) using the rapidity, we find that

$$\phi = \operatorname{arctanh}\beta = \operatorname{arctanh}\left(\frac{\beta_1+\beta_2}{1+\beta_1\beta_2}\right). \tag{15}$$

Using the identity

$$\operatorname{arctanh}\alpha = \frac{1}{2}\ln\left(\frac{1+\alpha}{1-\alpha}\right), \tag{16}$$

we have

$$\phi = \frac{1}{2}\ln\frac{(1+\beta_1)(1+\beta_2)}{(1-\beta_1)(1-\beta_2)}. \tag{17}$$

Thus,

$$\begin{aligned}
\phi &= \frac{1}{2}\ln\left(\frac{1+\beta_1}{1-\beta_1}\right) + \frac{1}{2}\ln\left(\frac{1+\beta_2}{1-\beta_2}\right) && \text{(18a)}\\
&= \operatorname{arctanh}\beta_1 + \operatorname{arctanh}\beta_2 && \text{(18b)}\\
&= \phi_1 + \phi_2. && \text{(18c)}
\end{aligned}$$

If we express the sum of two colinear velocities in the form of of Eq. (18) rather than in the usual form of Eq. (14), the relativistic sum of n colinear velocities $\vec{\beta_1}, \vec{\beta_2}, \ldots, \vec{\beta_n}$ is easily shown to have magnitude β, with

$$\begin{aligned}
\phi &= \operatorname{arctanh}\beta && \text{(19a)}\\
&= \frac{1}{2}\ln\frac{(1+\beta_1)(1+\beta_2)\ldots(1+\beta_n)}{(1-\beta_1)(1-\beta_2)\ldots(1-\beta_n)} && \text{(19b)}\\
&= \frac{1}{2}\ln\left(\frac{1+\beta_1}{1-\beta_1}\right) + \ldots + \frac{1}{2}\ln\left(\frac{1+\beta_n}{1-\beta_n}\right) && \text{(19c)}\\
&= \operatorname{arctanh}\beta_1 + \cdots + \operatorname{arctanh}\beta_n && \text{(19d)}\\
&= \phi_1 + \phi_2 + \ldots + \phi_n. && \text{(19e)}
\end{aligned}$$

Thus, the rapidity provides an easy way to express the relativistic sum of n colinear velocities when $n \geq 2$.

Note that using the Lorentz transformation Eq. (12) allows us to express Eq. (19) in another useful form.[30] If we combine Eqs. (12b) and (13b), we have

$$e^{\phi} = \sqrt{\frac{1+\beta}{1-\beta}}. \tag{20}$$

Using this relation in Eq. (19e) we find that

$$e^{\phi} = e^{\phi_1}e^{\phi_2}\ldots e^{\phi_n}, \tag{21}$$

which implies that

$$\left(\frac{1+\beta}{1-\beta}\right) = \left(\frac{1+\beta_1}{1-\beta_1}\right)\left(\frac{1+\beta_2}{1-\beta_2}\right)\cdots\left(\frac{1+\beta_n}{1-\beta_n}\right). \tag{22}$$

Equation (22) provides a surprisingly easy way to find the resultant β of the relativistic sum of n colinear boosts.

IV. RAPIDITY SPACE

In this section we derive a 2D relativistic velocity space, called rapidity space, directly from the 3D spacetime of special relativity (that is, from the spacetime with two spatial and one time dimension). If we use the coordinates x_1, x_2, and x_3 defined in Sec. II, the square of the line element for this spacetime is

$$ds^2 = (dx_1)^2 + (dx_2)^2 - (dx_3)^2. \tag{23}$$

We choose this particular form of ds^2 because in the $x_3 = 0$ plane, it reduces to the usual Euclidean relation

$$ds_E^2 = (dx_1)^2 + (dx_2)^2. \tag{24}$$

Suppose we fix ourselves in one inertial frame and consider another with the same spacetime origin but moving with a velocity $\vec{v}$ relative to the first. The spatial origin of this second frame appears to us as following a straight line $(x(t), y(t)) = \vec{v}t$, where $(v/c) < 1$. Thus its trajectory is a line emanating from the origin and lying within the light cone. This line also can be described in x_i-coordinates as the one formed by all the scalar multiples of the vector $(\vec{\beta}, 1)$, where $\vec{\beta} = \vec{v}/c$. If we turn this statement around, we can say that every straight line through the origin that lies within the light cone represents the trajectory of the origin of some inertial frame traveling with a velocity $\vec{v}$ relative to the fixed inertial frame represented by the spacetime.

Another way of describing all the straight lines through the origin and within the light cone is to note that each can be viewed as the x_3' axis of some inertial frame obtained from the original (x_1, x_2, x_3) frame by a unique boost. Because each x_3' axis corresponds to one particular velocity (and vice versa), we can create a model of velocity space by choosing one point from each x_3' axis. The set of all such points is a velocity space because each point in

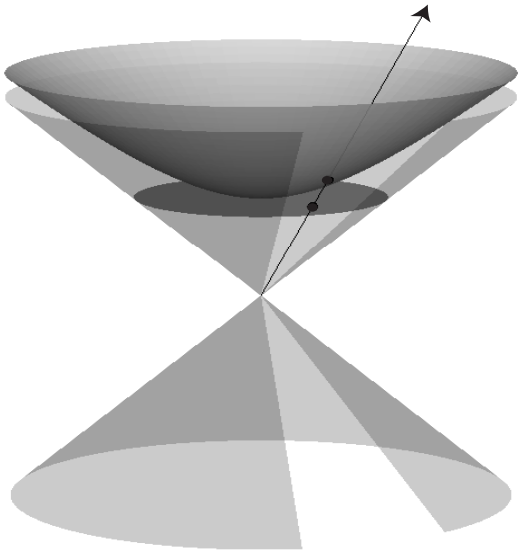


FIG. 1: The hyperboloid and Klein (or simultaneity) models of rapidity space. Note that any possible time-axis intersects each surface exactly once, so the point of intersection can be used to represent that axis. Part of the light cone has been removed to make the figure easier to understand.

it represents a unique velocity, and because all velocities $\vec{\beta}$ with magnitude $\beta < 1$ will be represented.

There are several natural ways to construct a relativistic velocity space from the set of points defined in the previous paragraph. For example, we could start with all the points lying in the plane $x_3 = 1$ and inside the lightcone; alternatively, we could start with all the points lying on the hyperboloid with $x_3 > 0$.

As shown in Fig. 1, the first choice is the *simultaneity plane* $x_3 = 1$ for an observer in the inertial frame represented by our spacetime, while the second is the set of all points for which the proper time $\tau = 1$. The first choice results in a velocity space known as the Klein model, which is not the best choice for our purposes because angles in this model do not appear like Euclidean angles (that is, the Klein model is not conformal.[31]) The reason we would like a conformal model is that the relativistic addition of non-colinear velocities, the Thomas-Wigner rotation and the Thomas precession, are built on understanding angles between successive boosts, so only those spaces in which angles behave like Euclidean angles can be expected to offer the geometric insight we seek.

The second choice, of all the points that lie on the hyperboloid of revolution, does result

in a conformal velocity space and is in fact the one chosen in Ref. 25. This choice is both reasonable and convenient because the metric on this space and the geometric properties that follow from it are well-known to mathematicians. However, few of us are good at judging angles on a curved surface.

So how do we motivate or justify choosing one model over another? Given that we are free to choose any surface created by any method of choosing one point from each x_3' axis, why choose the hyperboloid? How do we know there isn't some other way of choosing a point from each x_3' axis that will lead to an even more convenient or appropriate model of velocity space?

Rather than trying to justify one choice over another after the fact, we construct a model of relativistic velocity space from first principles by following the method used to derive spacetime and the spacetime metric. In that case, spacetime is developed from the physical requirement that the speed of light is independent of the motion of the source and is the same in all inertial frames. The mathematical statement of this property, $r^2 = (ct)^2$, implies that

$$x^2 + y^2 - (ct)^2 = 0, \tag{25}$$

and leads to interpreting the quantity

$$x^2 + y^2 - (ct)^2 \tag{26}$$

as the square of a distance in a 3D spacetime where Lorentz transformations are represented by linear coordinate transformations $\mathbf{\Lambda}$ satisfying Eq. (7). Thus, rather that choosing *a priori* the nature of spacetime, a physical invariance is used to deduce a metric that determines its mathematical properties.

We use this same approach to deduce the geometry of a relativistic velocity space. We note that Lorentz transformations acting on spacetime also act on the set of rays inside the light cone emanating from the origin, and that each of these rays has a one-to-one correspondence with a rapidity. Thus, the invariance of the spacetime metric Eq. (23) under Lorentz transformations can be used to define a metric on the rays (or rapidities) that also is Lorentz invariant.

To find this metric, first note that because we are choosing the points in rapidity space to correspond to rays inside the light cone that emanate from the spacetime origin, the metric in rapidity space should be expressible in terms of the spacetime coordinates. That is, there

should be functions $f_{i,j}$ such that the line element squared in rapidity space can be expressed in the form

$$ds^2 = \sum_{i,j=1}^{3} f_{i,j}(x_1, x_2, x_3) dx_i dx_j. \tag{27}$$

However, because any two points on the same ray in spacetime specify the same rapidity, the line element in rapidity space must be the same regardless of which spacetime points on the ray we choose. Thus, for any λ, we require

$$ds^2(x_1, x_2, x_3) = ds^2(\lambda x_1, \lambda x_2, \lambda x_3). \tag{28}$$

The spacetime form $ds^2 = dx_1^2 + dx_2^2 - dx_3^2$ does not have this property because

$$ds^2(\lambda x_1, \lambda x_2, \lambda x_3) = \lambda^2 ds^2(x_1, x_2, x_3). \tag{29}$$

However, we can obtain a ds^2 with the property given in Eq. (28) by using a simple but clever trick: first take the logarithm of both sides of Eq. (29) (which changes the multiplication by λ^2 into the addition of $\ln \lambda^2$), and then differentiate (so that the $\ln \lambda^2$ term disappears).

More formally, the spacetime inner product,

$$q(x, y) = -x \cdot y = -x_1 y_1 - x_2 y_2 + x_3 y_3, \tag{30}$$

is positive for rays within the light cone, and has the property

$$q(\lambda_1 x, \lambda_2 y) = \lambda_1 \lambda_2 q(x, y). \tag{31}$$

By taking the logarithm of both sides, we find

$$\ln q(\lambda_1 x, \lambda_2 y) = \ln \lambda_1 + \ln \lambda_2 + \ln q(x, y). \tag{32}$$

Finally, taking the differential of both sides respect to x and y, we obtain

$$d_x d_y \ln q(\lambda_1 x, \lambda_2 y) = d_x d_y \ln q(x, y). \tag{33}$$

In this way we are led to look for a rapidity space metric whose inner product has the form

$$d_x d_y \ln(-x \cdot y) = d_x d_y [\ln(x_3 y_3 - x_1 y_1 - x_2 y_2)] \tag{34}$$

$$= d_x \left[\frac{x_3 dy_3 - x_1 dy_1 - x_2 dy_2}{x_3 y_3 - x_1 y_1 - x_2 y_2} \right] \tag{35}$$

$$= (x \cdot y)^{-2} [(dx \cdot dy)(x \cdot y) - (x \cdot dy)(y \cdot dx)]. \tag{36}$$

Setting $x = y$ we see that the line element squared in this relativistic velocity space should have the form

$$ds^2 = K(x \cdot x)^{-2}\big[(dx \cdot dx)(x \cdot x) - (x \cdot dx)^2\big], \tag{37}$$

where K is an arbitrary constant.

Recall that even though the line element in Eq. (37) is expressed in terms of spacetime coordinates, it also is a function of the rays inside the light cone and through the origin on which those spacetime points lie.

As mentioned above, in order to better visualize the geometry that follows from ds^2 in Eq. (37), we may choose as a model of relativistic velocity space any surface that intersects each ray in exactly one point. If such a surface is expressed as

$$x_3 = g(x_1, x_2), \tag{38}$$

then by substituting this expression for x_3 into Eq. (37), we can express the metric in terms of the two coordinates x_1 and x_2.

Although we have great freedom in selecting the function g, several judicious choices will greatly simplify our model. First, because Lorentz transformations include rotations, and because we seek a metric that is invariant under Lorentz transformations, it is natural to choose a surface that has rotational symmetry. Thus, we require that

$$x_3 = g\left(\sqrt{(x_1)^2 + (x_2)^2}\right) \equiv g(r). \tag{39}$$

To find dx_3 in terms of x_1 and x_2, we apply the chain rule to Eq. (39), giving

$$dx_3 = \frac{g'}{r}\left(x_1 dx_1 + x_2 dx_2\right), \tag{40}$$

with

$$g' \equiv \frac{dg}{dr}. \tag{41}$$

Because

$$-x \cdot x = x_3^2 - x_1^2 - x_2^2 = g^2 - r^2, \tag{42}$$

we can rewrite Eq. (37) as

$$\begin{aligned} ds^2 = K(g^2 - r^2)^{-2}\Big[&(dx_3^2 - dx_1^2 - dx_2^2)(g^2 - r^2) \\ &- (x_3 dx_3 - x_1 dx_1 - x_2 dx_2)^2\Big]. \end{aligned} \tag{43}$$

If we use Eqs. (39) and (40) in Eq. (43) and do some algebra, we find that

$$ds^2 = \left[\frac{K}{r^2-g^2}\right](dx_1^2+dx_2^2) - K\left[\frac{g'^2(r^2-g^2)}{r^2} + \left(1-\frac{gg'}{r}\right)^2\right]\left(\frac{x_1dx_1+x_2dx_2}{r^2-g^2}\right)^2. \quad (44)$$

Although there is still much freedom in our choice of the surface g, we now impose our desire to have a model that is conformal.[32] We note that a metric will be conformal if it is a multiple of the Euclidean metric, even if the multiplicative factor varies from point to point. Thus, Eq. (44) will be a conformal metric if cross terms like dx_1dx_2 are not present. To this end, we look for a surface g for which

$$\frac{g'^2(r^2-g^2)}{r^2} + \left(1-\frac{gg'}{r}\right)^2 = 0, \quad (45a)$$

which implies that

$$g'(g'r-2g)+r=0. \quad (45b)$$

One way to solve this first-order non-linear differential equation for g is to look for solutions of the form

$$g = Ar^2 + Br + C. \quad (46)$$

If we substitute thus form for g into Eq. (45b), we find that g will be a solution if the coefficients A, B and C satisfy the conditions

$$2AB = 0, \quad (47a)$$

$$B^2 + 4AC = 1, \quad (47b)$$

$$2BC = 0. \quad (47c)$$

One set of coefficients that satisfies these conditions is $A = \pm\frac{1}{2}$, $C = \pm\frac{1}{2}$, and $B = 0$, in which case[33]

$$g = \pm\left(\frac{1+r^2}{2}\right). \quad (48)$$

Therefore, we can use Eq. (48) in Eq. (44) and find that

$$ds^2 = \left(\frac{K}{r^2-g^2}\right)(dx_1^2+dx_2^2) \quad (49a)$$

$$= \left(\frac{4K}{2r^2-r^4-1}\right)(dx_1^2+dx_2^2) \quad (49b)$$

which implies that

$$ds^2 = \left(\frac{2}{1-r^2}\right)^2(dx_1^2+dx_2^2). \quad (49c)$$

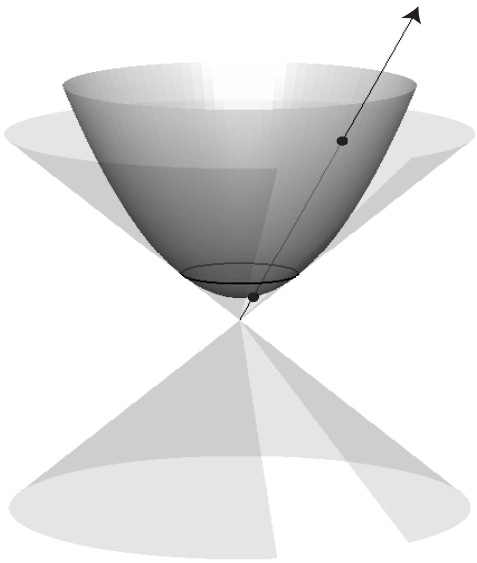


FIG. 2: The paraboloid model. Because each non-vertical time-axis intersects the paraboloid in two points, the lower point is chosen as its representation in this model.

Note that because we prefer distances in velocity space to be non-negative and real, we have chosen $K = -1$ in Eq. (49c).

The surface described by Eq. (48) is

$$x_3 \equiv g = \frac{1+r^2}{2} = \frac{1+x_1^2+x_2^2}{2}, \tag{50}$$

which is a paraboloid of revolution about the x_3 axis with vertex at $x_1 = x_2 = 0$ and $x_3 = 1/2$. The paraboloid also goes through points with $r^2 = x_1^2 + x_2^2 = 1$ and $x_3 = 1$, which are on the light cone. Not only does the paraboloid touch the light cone at $r = 1$, but the light cone is tangent to the paraboloid at this point because

$$\frac{\Delta(ct)}{\Delta(r)}\Big|_{r=1} \to \frac{d}{dr}(ct)\Big|_{r=1} = \frac{d}{dr}\big(\frac{1+r^2}{2}\big)\Big|_{r=1} = 1, \tag{51}$$

which is exactly the slope of the light cone.

As shown in Fig. 2, each ray emanating from the origin and inside the light cone (with the exception of the x_3-axis) intersects the paraboloid twice: once below the disk $x_3 = 1$ and once above it. Because we only need a single point from each ray to create a velocity space, we only use that part of the paraboloid that is below the disk (that is, points on the paraboloid with $x_3 < 1$), as shown in Fig. 3.

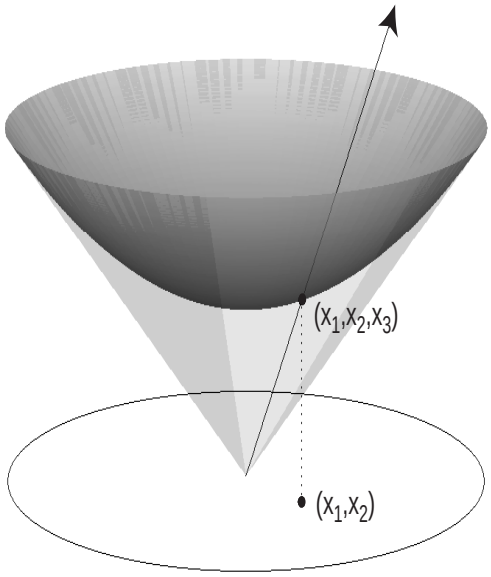


FIG. 3: Possible time axes correspond to their unique points of intersection with the lower part of the paraboloid, which then correspond by downward orthogonal projection to unique points in the Poincaré disk.

Finally, note that the line element squared given in Eq. (49c) can be expressed in plane polar coordinates as

$$ds^2 = \left(\frac{2}{1-r^2}\right)^2 (dr^2 + r^2 d\theta^2). \tag{52}$$

V. VELOCITY SPACE, RAPIDITY SPACE AND THE POINCARÉ DISK

Although the paraboloid of revolution derived in Sec. IV is a valid model of relativistic velocity space, in most cases it is much easier to work in the space obtained by projecting this paraboloid onto the (x_1, x_2) plane by the projection $(x_1, x_2, x_3) \mapsto (x_1, x_2)$ shown in Fig. 3.

The space created by this projection is a unit disk with the metric in Eq. (49c), and is known to mathematicians as the *Poincaré disk*. Although both the Poincaré and hyperboloid models are conformal (the first in two dimensions and the second in three), the Poincaré model is superior for building intuition about the Thomas-Wigner rotation and Thomas precession because it can be drawn in two dimensions, which makes line segments and angles easier to visualize.

As we shall see, the distance from the origin to any point on the Poincaré disk (as determined by the line element in Eq. (49c)) is just the rapidity associated with that point, which is why we refer to this disk as rapidity space.[34] Because points on the edge of the disk are defined by the projection of points on the intersection of the paraboloid and the light cone, they represent velocities with speed $v = c$. We shall see that these points are an infinite (Poincaré) distance away from any point inside the disk, reflecting the fact that speeds can approach but never reach the speed of light.

To simplify the notation, we rename the coordinates on the disk $x\,(\equiv x_1)$ and $y\,(\equiv x_2)$. We identify the physical significance of the distance s of any point from the origin by evaluating[35]

$$s = \int_0^R ds = \int_0^R \Big(\frac{2}{1-r^2}\Big)\sqrt{dx^2+dy^2} \tag{53a}$$

$$= \int_0^R \Big(\frac{2}{1-r^2}\Big)\sqrt{dr^2+r^2d\theta^2} \tag{53b}$$

$$= \int_0^R \Big(\frac{2}{1-r^2}\Big)dr \tag{53c}$$

$$= \ln\Big(\frac{1+R}{1-R}\Big), \tag{53d}$$

which implies that

$$s = 2\,\mathrm{arctanh}\,R. \tag{54}$$

However a point on the disk whose radial coordinate is $R = \sqrt{a^2+b^2}$ corresponds to the spacetime point $(a, b, \frac{a^2+b^2+1}{2})$, which lies on the ray in spacetime with slope

$$\frac{ct}{R} = \frac{(1+R^2)/2}{R}, \tag{55}$$

which implies that

$$\frac{c}{v} = \frac{1+R^2}{2R} \tag{56}$$

$$\beta = \frac{2R}{1+R^2} = \frac{2\tanh(s/2)}{1+\tanh(s/2)\tanh(s/2)} \tag{57}$$

$$\beta = \tanh(s/2+s/2) = \tanh s. \tag{58}$$

We thus conclude that the (Poincaré distance) s of any point from the origin is

$$s = \mathrm{arctanh}\,\beta = \phi. \qquad \text{(the rapidity)} \tag{59}$$

We can further clarify the nature of rapidity space by relating its coordinates (x, y) to the Euclidean velocity components v_x and v_y. First, note that a point (x, y) on the disk

corresponds to a point $(x, y, \frac{1+r^2}{2})$ on the paraboloid, which means that

$$\beta_x = \Big(\frac{2}{1+r^2}\Big)x, \quad \beta_y = \Big(\frac{2}{1+r^2}\Big)y, \tag{60}$$

and

$$\beta = \Big(\frac{2}{1+r^2}\Big)r. \tag{61}$$

Equations (60) and (61) give β (and its components) in terms of any radial coordinate r (and its components). To find the inverse relation (that is, the r associated with a given β), we first solve Eq. (61) for the magnitude of r:

$$\beta r^2 - 2r + \beta = 0 \tag{62}$$

$$r = \frac{1 - \sqrt{1-\beta^2}}{\beta} \tag{63}$$

$$r = \frac{\gamma - 1}{\gamma\beta}. \tag{64}$$

(We use only the negative root in the quadratic formula because $r < 1$). We substitute Eq. (64) in Eq. (60) and use the identity,

$$\gamma^2 - 1 = \frac{\beta^2}{1-\beta^2} = (\gamma\beta)^2, \tag{65}$$

to find

$$x = \Big(\frac{1+r^2}{2}\Big)\beta_x = \Big(\frac{\gamma}{\gamma+1}\Big)\beta_x \tag{66a}$$

$$y = \Big(\frac{1+r^2}{2}\Big)\beta_y = \Big(\frac{\gamma}{\gamma+1}\Big)\beta_y, \tag{66b}$$

and

$$r = \Big(\frac{\gamma}{\gamma+1}\Big)\beta. \tag{67}$$

From Eq. (66) we see that the x coordinate in rapidity space is proportional to β_x and the y coordinate is proportional to β_y. The proportionality factor in both cases is $\gamma/(\gamma+1)$, which goes to unity as $v \to c$ and to $1/2$ as $v \to 0$. Thus, in the limit as $v \to 0$, the line element squared in rapidity space (Eq. (49c)) reduces to

$$ds^2 = (d\beta_x)^2 + (d\beta_y)^2, \tag{68}$$

which, of course, is the line element squared of the usual Euclidean non-relativistic velocity space.

To summarize, we have shown that requiring the velocity space metric to be invariant under Lorentz transformations leads to a model of relativistic velocity space that can be represented either as a paraboloid of revolution with its vertex at $x_1 = x_2 = 0, x_3 = \frac{1}{2}$ and top edge at $x_3 = 1$, or as a unit disk with the metric given in Eq. (49c). In this paper we refer to this (Poincaré) disk as rapidity space because the distance from the origin to any point on the disk is its rapidity. Note that from the definition of rapidity, it is easy to see that any point on the edge of the disk is infinitely far from any point in the disk. Also, any point in rapidity space is related to the components of any velocity by Eqs. (60) and (66). The conformal property of rapidity space can be seen explicitly by noting that if $\arctan(v_y/v_x)$ is the angle made by a velocity vector $\vec{v}$ with the horizontal axis in real space, then, using Eq. (60), we have

$$\arctan\frac{v_y}{v_x} = \arctan\frac{y}{x} = \theta, \tag{69}$$

which means that the angle of $\vec{\beta}$ in real space is the same as the angle θ in rapidity space. This property of rapidity space is what makes it so useful for understanding the Thomas-Wigner rotation and Thomas precession.

VI. GEODESICS IN RAPIDITY SPACE

We need to understand one more aspect of rapidity space before we can use it to investigate the relativistic addition of non-colinear velocities and boosts. When we boost from one inertial frame traveling with a velocity $\vec{v}_1$ to another traveling with a velocity $\vec{v}_2$, we pass through the minimum number of velocities whose speeds are greater than v_1 and less than v_2. The corresponding path in rapidity space between the points representing these velocities is the shortest one which, by definition, is the geodesic connecting them. Hence, to study successive non-colinear boosts, we need to identify the geodesics in rapidity space. We shall see that these geodesics are straight lines if the origin (the point representing zero velocity) is one of the values taken on during the boost. In all other cases the geodesics are not straight lines but rather are the arcs of circles. Following these arcs from one point in rapidity space to another will bring out most of the interesting features of relativistic velocity addition, successive Lorentz boosts, Thomas-Wigner rotations and Thomas precession.

We begin by showing that any geodesic that includes the origin of rapidity space is a straight line. Our proof parallels the traditional one showing that the shortest distance

between two points in a Euclidean plane is a straight line, and is accomplished by finding the path of minimum distance connecting two points on the disk. The length of any path connecting the origin and a point (A, B) in rapidity space is

$$\int_{(0,0)}^{(A,B)} ds = \int_{(0,0)}^{(A,B)} \Big(\frac{2}{1-r^2}\Big)\sqrt{dr^2 + r^2 d\theta^2} \tag{70a}$$

$$= \int_0^R \Big(\frac{2}{1-r^2}\Big)\sqrt{1 + r^2\theta'^2}\, dr, \tag{70b}$$

with $R = \sqrt{A^2 + B^2}$, and

$$\theta' \equiv \frac{d\theta}{dr}. \tag{71}$$

If we use the Euler-Lagrange equation, we know that the integral in Eq. (70b) will be an extremum when

$$\frac{\partial}{\partial\theta}\Big[\Big(\frac{2}{1-r^2}\Big)\sqrt{1 + r^2\theta'^2}\Big] - \frac{d}{dr}\frac{\partial}{\partial\theta'}\Big[\Big(\frac{2}{1-r^2}\Big)\sqrt{1 + r^2\theta'^2}\Big] = 0. \tag{72}$$

The first term on the left-hand side is zero, so Eq. (72) reduces to

$$\frac{d}{dr}\frac{\partial}{\partial\theta'}\Big[\Big(\frac{2}{1-r^2}\Big)\sqrt{1 + r^2\theta'^2}\Big] = 0, \tag{73}$$

which implies that

$$\frac{d}{dr}\Big[\Big(\frac{2}{1-r^2}\Big)\frac{r^2\theta'}{\sqrt{1 + r^2\theta'^2}}\Big] = 0, \tag{74}$$

and

$$\Big(\frac{2}{1-r^2}\Big)\frac{r^2\theta'}{\sqrt{1 + r^2\theta'^2}} = h(\theta'). \tag{75}$$

If we square both sides of Eq. (75) and rearrange the terms, we find $h^2(Ar^6 + Br^4 + Cr^2 + 1) = 4r^4\theta'^2$, with A, B, and C functions of θ'. In order for this equation to be satisfied for all r, including $r = 0$, $h(\theta')$ must be zero. Hence, from Eq. (75) we see that θ' also must be zero, and thus θ is a constant whenever $r = 0$ lies on the path. Therefore, any geodesic in rapidity space that includes the origin is a straight line, as shown in Fig. 6.

Next we derive the geodesics in rapidity space that do not include the origin. Because the rapidity space metric is Lorentz invariant by construction, Lorentz transformations must send geodesics to geodesics. Thus, we can obtain a geodesic that does not include the origin by applying the same boost to every point on a geodesic that does include the origin. Rather than doing this directly, it is easier to obtain the final geodesic geometrically by projecting back and forth between the disk and the paraboloid. In this approach, we first identify the

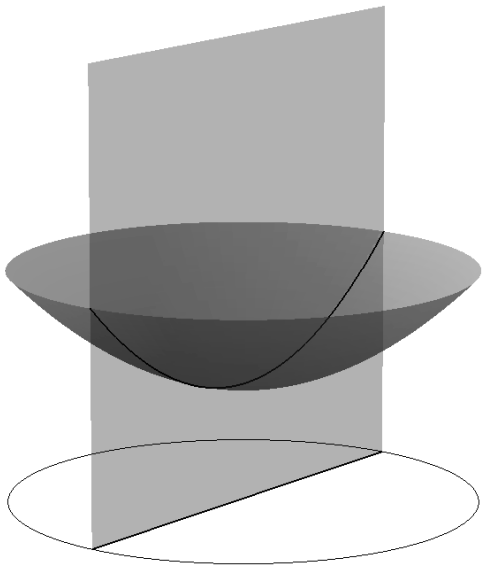


FIG. 4: Geodesics through the vertex of the paraboloid are formed by intersections of the paraboloid and vertical planes. These project to Euclidean-straight lines through the origin in the Poincaré model.

geodesics on the paraboloid that correspond to geodesics through the origin of the disk. We then apply the same boost to every point on one of these geodesics on the paraboloid. Finally, we project the Lorentz transformed geodesic on the paraboloid onto the disk and find the equation that describes it.

We already have proven that any geodesic that includes the origin of the disk is a straight line. If we project one of these straight lines back up to the paraboloid, we see that it corresponds to a parabola through the vertex of the paraboloid, as shown in Fig. 4. Thus, any such (vertically oriented) parabola is a geodesic. Equivalently, any one of these (vertically oriented) parabolas can be regarded as the curve formed at the intersection of the paraboloid and the plane defined by two axes x_3' and x_3'', each of which has a point on the paraboloids' geodesic.

If we now perform the same pure boost on every frame represented by a point on a parabola passing through the vertex of the paraboloid, we obtain a new geodesic that does not include the vertex. Because boosts are linear transformations, planes through the origin transform into other planes through the origin. Thus, the new geodesic can be described as the curve created by the intersection of the paraboloid and the new plane formed by the

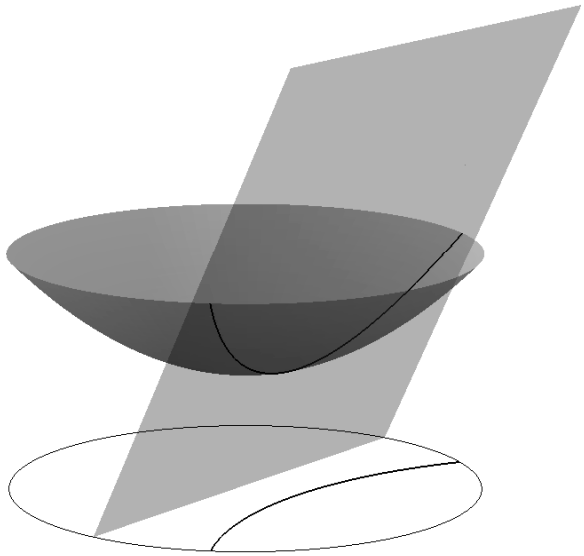


FIG. 5: All geodesics on the paraboloid are formed by its intersection with planes through the origin. These project downward onto circular arcs, orthogonal to the unit disk, in the Poincaré model.

boosted x_3' and x_3'' axes, as shown in Fig. 5. If we project this curve onto the disk, we can find the equation of an arbitrary geodesic (on the disk) that does not include the origin:

Because the boosted x_3' and x_3'' axes define the plane whose intersection with the paraboloid determines the new geodesic, this plane must lie within the light cone. It follows that there is a normal to the plane making an angle θ with the original x_3 axis, with $\pi/2 > \theta > \pi/4$:

$$0 < \cos\theta < \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}. \tag{76}$$

On the other hand, if we express the normal to the plane as $(a, b, 1)$, then any point (x_1, x_2, x_3) on the plane satisfies

$$(a, b, 1) \cdot (x_1, x_2, x_3) = 0, \tag{77}$$

or $ax_1 + bx_2 + x_3 = 0$. However, Eq. (76) tells us that

$$\cos\theta = \frac{(a, b, 1) \cdot (0, 0, 1)}{\sqrt{a^2 + b^2 + 1}} < \frac{1}{\sqrt{2}}, \tag{78}$$

which implies that $\sqrt{a^2 + b^2 + 1} > \sqrt{2}$ and $a^2 + b^2 > 1$. We thus conclude that the plane shown in Fig. 5 is specified by

$$ax_1 + bx_2 + x_3 = 0 \quad \text{with} \quad a^2 + b^2 > 1. \tag{79}$$

Because any such plane can be obtained by boosting the appropriate x_3' and x_3'', we conclude that planes of the form in Eq. (79) determine all the geodesics on the paraboloid. That is, a curve is a geodesic if and only if it lies on the intersection of the paraboloid and any plane including the origin.

Our main interest is in the points shown in Fig. 5 that lie on the intersection of the plane and the paraboloid. Any point on the paraboloid satisfies Eq. (50),

$$x_3 = \frac{1 + x_1^2 + x_2^2}{2}. \tag{80}$$

If we combine Eq. (80) with Eq. (79), we see that points on the curve formed at the intersection of the plane and the paraboloid (that is, points on a geodesic on the paraboloid that do not pass through its vertex) satisfy

$$-ax_1 - bx_2 = \frac{x_1^2 + x_2^2 + 1}{2}, \qquad (a^2 + b^2 > 1) \tag{81}$$

which implies that

$$(x_1 + a)^2 + (x_2 + b)^2 = a^2 + b^2 - 1 > 0. \tag{82}$$

If we project back down onto the disk as shown in Fig. 5, we see that any geodesic that does not include the origin is the arc of a circle centered at $(-a, -b)$ with radius $\sqrt{a^2 + b^2 - 1}$. Note that the center of any of these circles always lies outside of the unit disk because $a^2 + b^2 > 1$, and that any point (a, b) outside of the unit disk is the center of a circle on which some geodesic inside the disk lies. Also, because a and b can be positive or negative, it doesn't really matter whether we denote the center of the circle by $(-a, -b)$ or (a, b).

Finally, we can prove that the circles derived in Eq. (82) are perpendicular to the edge of the disk at their points of intersection. We recall from Sec. V that points on the disk are denoted by (x, y), points on the edge of the disk satisfy

$$x^2 + y^2 = 1, \tag{83}$$

and points on a geodesic on the disk satisfy

$$(x + a)^2 + (y + b)^2 = a^2 + b^2 - 1. \tag{84}$$

By taking the differential of Eq. (83) and Eq. (84), we obtain

$$2xdx + 2ydy = 0 \tag{85}$$

$$2(x + a)dx + 2(y + b)dy = 0, \tag{86}$$

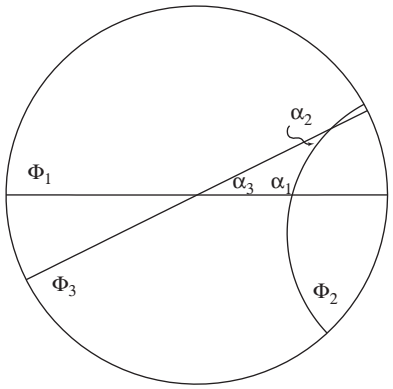


FIG. 6: Three geodesics in rapidity space and a non-Euclidean triangle. The geodesics are labelled as Φ_i and the angles of intersection as α_i

which can be rewritten as

$$\frac{dy}{dx} = -\frac{x}{y} \tag{87}$$

$$\frac{dy}{dx} = -\frac{x+a}{y+b}. \tag{88}$$

Equation (87) gives the slope of the tangent to any point on the edge of the disk, and Eq. (88) gives the slope of the tangent to any point on the geodesic. To prove that the geodesic is perpendicular to the edge of the disk at their point of intersection, we must show that these two tangents are perpendicular to each other; that is, that

$$-\frac{x}{y} = \frac{y+b}{x+a}. \tag{89}$$

To do this, we note that any point on both the edge of the disk and on the geodesic satisfies both Eqs. (82) and (83), and by adding these two equations together we find

$$2x^2 + 2ax + 2y^2 + 2by = 0 \tag{90}$$

$$x(x+a) + y(y+b) = 0 \tag{91}$$

$$-\frac{x}{y} = \frac{y+b}{x+a}, \tag{92}$$

as required. Note that this argument can be reversed to show that any circular arc orthogonal to the unit circle at their points of intersection is a geodesic.

The main results of this section are the following. As shown in Figs. 4 and 6, geodesics through the origin of rapidity space are straight lines (in the Euclidean sense), and any straight line through the rapidity space origin is a geodesic.

As shown in Figs. 5 and 6, geodesics in rapidity space which do not include the origin are arcs of circles whose centers lie outside the disk and are perpendicular to the edge of the disk at their points of intersection. Conversely, any point outside the disk is the center of some circular arc within the disk that is a geodesic. As Fig. 6 shows, the new feature is that rapidity space contains geodesics that are not "straight" in the Euclidean sense. One way to understand these curved geodesics is to note that the metric of Eq. (49c) tells us that segments near the edge of the disk are much longer then they appear, so the shortest path between two points near the disks' edge must be bowed inward rather than straight.

VII. APPLICATIONS

Now that we have found the Lorentz invariant metric and geodesics in rapidity space, we can investigate the relativistic addition of velocities and the various consequences of successive non-colinear Lorentz boosts.

A. Qualitative Results

Before deriving quantitative expressions for the Thomas-Wigner rotation and Thomas precession, we first discuss results that can be deduced geometrically without the use of any equations.

Velocities are represented by points in rapidity space. A boost from one velocity to another is represented in rapidity space by the geodesic connecting them. Because a pure boost doesn't involve any rotation of the reference frame being boosted, the coordinate axes representing the boost in rapidity space maintain a fixed angle with respect to the geodesic they follow, as shown in Fig. 7. When coordinate axes maintain a fixed angle with a geodesic as they move along it from one point to another, they are said to undergo *parallel transport*. The most familiar example of parallel transport occurs in nonrelativistic velocity space, in which all the geodesics are straight lines.

We begin by considering a set of colinear boosts in real space. If we assume this set contains the zero velocity frame, then the first boost is represented in rapidity space by a segment of a straight line geodesic that includes the rapidity space origin. Without any loss of generality, we take that direction as the horizontal axis in both real and rapidity space.

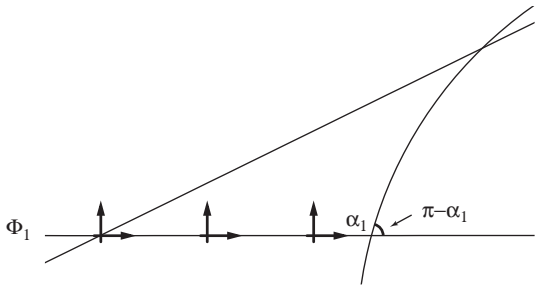


FIG. 7: The first boost, along the geodesic Φ_1.

Because straight lines through the origin of rapidity space are geodesics, each colinear boost is represented by a segment of the same (horizonal) line. If we represent a coordinate system in rapidity space by two small perpendicular lines (crosshairs) centered on the point of interest, then as shown in Fig. 7, when we boost from one velocity to another, the orientation of the crosshairs remains fixed with respect to the geodesic connecting them. Thus, the orientation of the crosshairs is unchanged no matter how many colinear boosts it undergoes because, in each case, it is moving along the same straight line geodesic in rapidity space.

Furthermore, because the distance (as measured with the rapidity space metric) from the origin to any point in rapidity space is the rapidity of that point, we see that when successive boosts are colinear, the corresponding rapidities add and subtract like ordinary numbers. Thus, rapidity space provides an easy geometrical way to obtain Eq. (19e) and to prove that frames boosted in the same direction do not rotate with respect to each other. Working in rapidity space also provides an easy proof that no matter how many colinear velocities are added together, the magnitude of their sum always will be less than the speed of light.

A boost that does not include the zero velocity frame corresponds to a geodesic in rapidity space that does not include the rapidity space origin. As we proved in Sec. VI, this geodesic lies on the arc of a circle whose center is outside the disk. As shown in Fig. 8, crosshairs moving along this type of geodesic maintain their orientation with respect to it. Therefore, crosshairs moving back to the origin along a closed path that includes one (or more) of these geodesics will be rotated with respect to their initial orientation. An example of this rotation is shown in Fig. 9. Suppose we boost from rest to a velocity $\vec{v}$ along the x-axis (in real space). Then we perform a non-colinear boost from a frame with velocity $\vec{v}$ to a frame with velocity

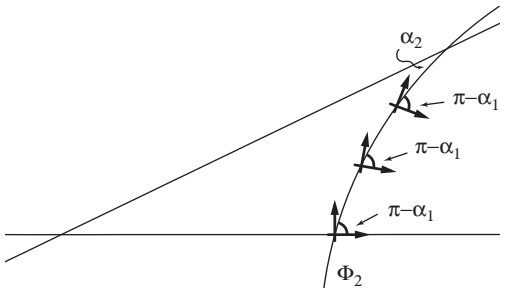


FIG. 8: The second boost, along the geodesic Φ_2.

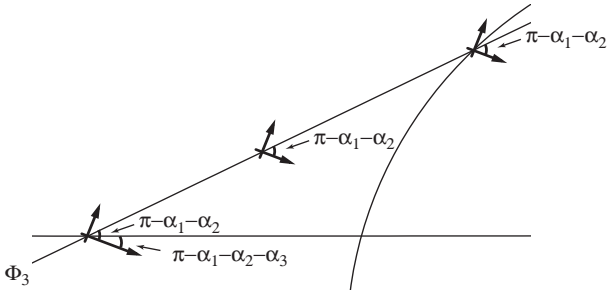


FIG. 9: The third boost, along the geodesic Φ_3.

$\vec{v'}$, and finally, we boost from a frame with velocity $\vec{v'}$ back to the original rest frame. If we look at the corresponding points in rapidity space as shown in Fig. 9, we see that the frame obtained at the end of these three boosts is rotated with respect to the one that stayed at the origin. This rotation is the Thomas-Wigner rotation, which we denote by TWR, and the geometry of rapidity space shows that the TWR is in the clockwise (negative) direction when a frame is moving in rapidity space in the counterclockwise (positive) direction (and vice versa).

It also is easy to see that there is an upper limit on the TWR angle. Without any loss of generality, suppose we first boost along the x-axis (in real space) to a frame whose speed is very close to the speed of light. There is no orientation change of the boosted frame because the geodesic it follows in rapidity space is a straight line. If we next perform a non-colinear boost to a speed even closer to the speed of light, which makes an angle slightly less than π with the x-axis, then as shown in Fig. 10, the geodesic representing this second boost will lie on the arc of a circle that is perpendicular to the edge of the disk (where $v = c$) at its two points of intersection. In the limit that both speeds approach the speed of light and the

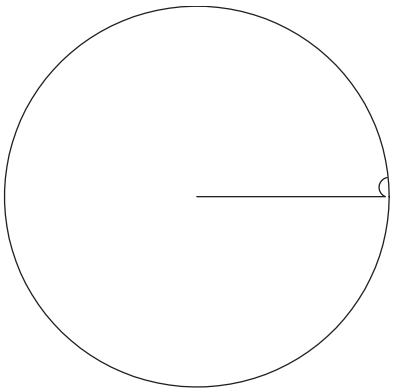


FIG. 10: When a second boost is applied at an angle approaching π with the first, the Thomas-Wigner rotation angle approaches its maximum value of π.

angle between the boosts approaches π, the arc representing the second boost approaches a half circle. Thus, the change in orientation of a reference frame following the arc approaches π. Consequently, any Thomas-Wigner rotation angle has an upper limit of π, and the limit of π is approached only when the two boosts involved have speeds very close to the speed of light and are almost opposite to each other. In all other cases the TWR angle will be less than π.

On the other hand, if we perform the same two non-colinear boosts as before, but now give each a nonrelativistic speed ($\beta \ll 1$), then as shown in Fig. 11, even though the geodesic representing the second boost still lies on the arc of a circle, that arc is indistinguishable from a straight line because it is located near the rapidity space origin. Therefore, when the boost speeds are nonrelativistic, there is essentially no rotation of the frame in following the two geodesics.

Next consider a reference frame in real space undergoing circular motion in the counter-clockwise (positive) direction with a constant, nonrelativistic speed. Classically, this situation is treated by representing it as the limiting case of a set of small, non-colinear boosts. That is, circular motion is approximated as motion along a polygon with an ever increasing number of sides. Because the boosts involved all have the same nonrelativistic speed, all the circular arcs representing them in rapidity space are essentially indistinguishable from straight lines. Hence, as long as the speed of the frame in circular motion is nonrelativistic, it undergoes essentially no change in orientation upon its return to the origin (as we would

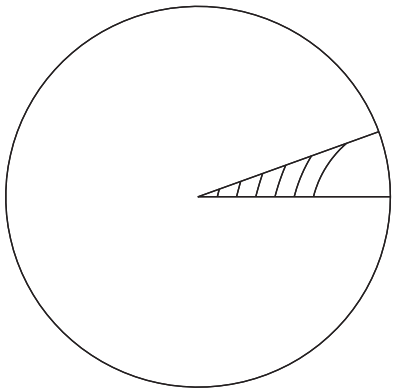


FIG. 11: All of these non-Euclidean triangles have the same two base angles, but the smaller a triangle is, the closer it is to appearing Euclidean.

expect).

Now suppose the reference frame undergoing circular motion has a constant speed that is relativistic. The geodesics in rapidity space that form the polygon are now small arcs that lie on circles whose intersection with the edge of the disk is orthogonal, as shown in Fig. 12. Consequently, the frame being boosted undergoes a definite change in orientation with each boost. Thus, when the reference frame returns to its starting point, it will have undergone a clockwise (negative) rotation with respect to its initial orientation. This sum of all the rotations experienced along the way is the Thomas precession. Furthermore, we see from the geometry of rapidity space that the amount of rotation will be a function of the speed of the circular motion (that is, the rapidity space distance from the rapidity space origin), and will increase without bound as this speed approaches the speed of light, as shown in Fig. 12.

We also can use the geometry of rapidity space to show that non-colinear, relativistic boosts are also (in general) noncommutative. Suppose we first boost a reference frame from speed zero to a speed close to the speed of light, and then boost the frame through a second non-colinear velocity. From the rapidity space diagram in Fig. 13, it is clear that we will end up at a completely different point if we do the same boosts in reverse order. However, if we look closely at the rapidity space diagram, we see that although the two resultant velocities are represented by different points in rapidity space, they both are the same (rapidity space) distance from the rapidity space origin. Consequently, they both have the same rapidity (and hence speed), but not the same direction. The result that the final speed resulting from two

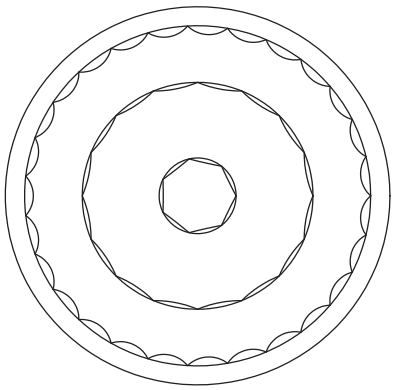


FIG. 12: Polygonal approximations to curved paths in rapidity space.

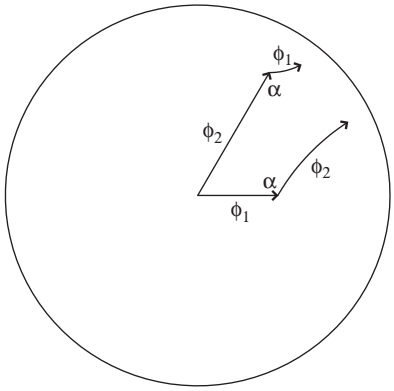


FIG. 13: Two boosts, with corresponding rapidities ϕ_i in non-parallel directions, do not commute. In both cases, the angle between the boosts is α.

successive non-colinear boosts is independent of the order in which the boosts are applied is usually proved by a somewhat long algebraic calculation.[36]

Finally, it is easy to see that the sum of any number of non-colinear relativistic velocities always results in a velocity whose magnitude is less than the speed of light. Although obvious when viewed on a rapidity space diagram, the corresponding algebraic proof is somewhat complex.

B. Relation of the Thomas-Wigner rotation to the rapidity space triangle

Because angles in rapidity space behave exactly like angles in Euclidean space, it is relatively easy to quantify the arguments of Sec. VII A. As shown in Fig. 6, the angles

between the boosts are α_3, α_1, and α_2. The straight line geodesic making an angle α_3 with the horizontal axis is called Φ_3, and the other two geodesics are Φ_1 and Φ_2. The length (rapidity) of the segment of the geodesic Φ_i representing each boost is denoted by ϕ_i. We see in Fig. 7 that when we boost from the origin along Φ_1, there is no change in the orientation of the crosshairs. As shown in Fig. 8, when we boost along Φ_2, the x-axis of the crosshairs maintains its angle $\pi - \alpha_1$ with respect to Φ_2 because Φ_2 is a geodesic. Finally, as shown in Fig. 9, boosting back to the origin along Φ_3, the x-axis of the crosshairs maintains its orientation of $(\pi - \alpha_1) - \alpha_2$ with respect to the geodesic Φ_3. Thus, as Fig. 9 shows, when the coordinate system returns to the rapidity space origin, its x-axis will have rotated from its initial orientation in the clockwise (negative) direction by the Thomas-Wigner rotation angle

$$\text{TWR} = -\big[\pi - (\alpha_1 + \alpha_2 + \alpha_3)\big]. \tag{93}$$

Note that if the two rapidities ϕ_1 and ϕ_2 are small (that is, if the speeds they represent are nonrelativistic) then, as shown in Fig. 11, the figure formed by the segments of the three geodesics Φ_i is indistinguishable from a Euclidean triangle (because in this case $\alpha_1+\alpha_2+\alpha_3 \approx \pi$) and, as expected, the TWR angle is essentially zero. On the other hand, if the two rapidities ϕ_1 and ϕ_2 are large, then the resulting TWR angle can approach the upper limit of π, as discussed above.

The absolute value of the right-hand side of Eq. (93) is known to mathematicians as the "angular defect," because it is a measure of how much the sum of the angles inside a triangle differs from the corresponding sum in ordinary Euclidean space (π). The theorem that the angular defect of a triangle is equal to its area is proven in hyperbolic geometry courses. Rather than simply invoking this result, we can establish it from first principles. Even if we had no previous knowledge of hyperbolic geometry, we might suspect that the area enclosed by the rapidity space triangle is proportional to the TWR angle, because our method for finding the TWR angle involves traveling around a closed path and summing up the angular change along the way. This sum corresponds to evaluating an integral of the form

$$\oint_C d\theta. \tag{94}$$

An integral like (94) appears in Green's theorem, which relates an integral around a closed curve to an integral over the two-dimensional area enclosed by that curve. We now show

that the integral Eq. (94) is actually the left-hand side of Green's theorem for a particular choice of the integrand, and that this choice makes the right-hand side of Green's theorem equal to the area enclosed by the curve.

We begin by writing Green's theorem as

$$\oint_C \vec{F} \cdot d\vec{s} = \iint_\Sigma \nabla \times \vec{F} \cdot \hat{n}\, d\sigma, \tag{95}$$

where C signifies any closed (2D) curve traversed in the counterclockwise direction, $\hat{n}$ is the unit vector normal to the plane of this curve (according to the right-hand rule), Σ stands for the region enclosed by the curve, and $\vec{F}$ is any vector defined on rapidity space.

To apply Green's theorem in rapidity space, we first must find explicit expressions for each integrand. To do this, recall that

$$ds^2 = \frac{4}{(1-r^2)^2}\Big(dx^2 + dy^2\Big) = h_1^2\, dx^2 + h_2^2\, dy^2, \tag{96}$$

and

$$d\vec{s} = \Big(\frac{2}{1-r^2}\Big)dx\, \hat{\imath} + \Big(\frac{2}{1-r^2}\Big)dy\, \hat{\jmath}. \tag{97}$$

From Eq. (97), we find that the area element in Cartesian and plane polar coordinates is

$$d\sigma = \Big[\Big(\frac{2}{1-r^2}\Big)dx\Big]\Big[\Big(\frac{2}{1-r^2}\Big)dy\Big] \tag{98}$$

$$= \Big(\frac{2}{1-r^2}\Big)^2 dx\, dy = \Big(\frac{2}{1-r^2}\Big)^2 r dr d\theta. \tag{99}$$

We also can give a more informal derivation of Eq. (99). Because the metric Eq. (97) is conformal, it is locally a multiple of the Euclidean metric (even though this multiple varies from point to point). Therefore, we expect the area element also to be a multiple of the Euclidean area element, with the multiplying factor equal to $\big(2/(1-r^2)\big)^2$, because the area is the product of the infinitesimal length in each of the two orthogonal directions and each length is the Euclidean length multiplied by the factor $2/(1-r^2)$.

To evaluate the integral on the right-hand side of Eq. (95), we need to express the curl and dot product in the coordinates of rapidity space. From Boas[37] we have

$$\nabla \times \vec{F} \cdot \hat{n} \;=\; \frac{1}{h_1 h_2}\Big[\frac{\partial}{\partial x}\Big(h_2 F_2\Big) - \frac{\partial}{\partial y}\Big(h_1 F_1\Big)\Big] \tag{100a}$$

$$\;=\; \Big(\frac{1-r^2}{2}\Big)^2 \frac{\partial}{\partial x}\Big[\Big(\frac{2}{1-r^2}\Big)F_y\Big] - \frac{\partial}{\partial y}\Big[\Big(\frac{2}{1-r^2}\Big)F_x\Big]. \tag{100b}$$

Because Green's theorem is true for any vector $\vec{F}$, it holds for the particular vector $\vec{F} = -y\hat{\imath} + x\hat{\jmath}$. For this choice, the right-hand side of Eq. (100b) reduces to 1.

Thus, when $\vec{F} = -y\hat{\imath} + x\hat{\jmath}$, Green's theorem becomes

$$\text{Area}(\Sigma) = \oint_C \vec{F} \cdot d\vec{s}, \tag{101}$$

where the area enclosed by the closed curve C is calculated using the rapidity space area element Eq. (99). If we substitute our choice for $\vec{F}$ into the right-hand side of Eq. (101), we find

$$\oint_C \vec{F} \cdot d\vec{s} = \oint_C (-y\hat{\imath} + x\hat{\jmath}) \cdot \Big(\frac{2}{1-r^2}\Big)\Big(dx\hat{\imath} + dy\hat{\jmath}\Big) \tag{102a}$$

$$= \oint_C \Big(\frac{2}{1-r^2}\Big)\Big(x dy - y dx\Big). \tag{102b}$$

Therefore, Eq. (101) can be written as

$$\text{Area}(\Sigma) = \oint_C \Big(\frac{2}{1-r^2}\Big)(x\, dy - y\, dx). \tag{103}$$

We now show that the right-hand side of Eq. (103) is equal to an integral of the form given in Eq. (94). Let r and θ be the usual plane polar coordinates measured from the center of the disk. In terms of these coordinates,

$$x = r\cos\theta \quad \text{and} \quad dx = -r\sin\theta d\theta + dr\cos\theta, \tag{104a}$$

$$y = r\sin\theta \quad \text{and} \quad dy = \quad r\cos\theta d\theta + dr\sin\theta. \tag{104b}$$

We substitute these coordinates into Eq. (103) and find that

$$x\, dy - y dx = r^2 d\theta, \tag{105}$$

which implies that

$$\text{Area}(\Sigma) = \oint_C \left(\frac{2r^2}{1-r^2}\right) d\theta. \tag{106}$$

By looking at the hyperbolic triangle in Fig. 14 representing the three boosts, we see that two of the three sides are straight lines emanating from the origin, which means that $d\theta = 0$ for these geodesic segments. Consequently, the integral along each of these segments makes no contribution to the path integral. Therefore, only the integral over the curved

geodesic contributes to the right-hand side of Eq. (106). We can evaluate this integral by changing it from one in terms of the polar coordinates (r, θ) measured from the center of the disk to an integral in terms of coordinates measured from the center (a, b) of the circle on which the curved arc lies (see Fig. 14). If $\sqrt{a^2 + b^2 - 1}$ is the radius of this circle and ω the corresponding angular coordinate (defined as positive in the counterclockwise direction), then the desired coordinate transformation is

$$x = a - \sqrt{a^2 + b^2 - 1}\cos\omega, \tag{107a}$$

$$y = b - \sqrt{a^2 + b^2 - 1}\sin\omega. \tag{107b}$$

After some algebra, we find that

$$\frac{2}{1 - r^2} = \frac{1}{1 - (a^2 + b^2) + \sqrt{a^2 + b^2 - 1}\big(a\cos\omega + b\sin\omega\big)}, \tag{108}$$

and

$$x dy - y dx = -\big[1 - (a^2 + b^2) + \sqrt{a^2 + b^2 - 1}\big(a\cos\omega + b\sin\omega\big)\big] d\omega. \tag{109}$$

If we use Eqs. (108) and (109) in the integrand of Eq. (103), we conclude that

$$\Big(\frac{2}{1 - r^2}\Big)(x dy - y dx) = -d\omega. \tag{110}$$

Using Eq. (110) in the integral in Eq. (103), we see that the integral in Eq. (103) is zero on the straight lines through the origin, while on the other geodesic, it is (to within a sign) the angular extent of the geodesic segment about its center.

If we look at Fig. 14, we see that the two radii of the circle centered at (a, b) together with the two sides of the triangle which are straight lines, form a four-sided figure (that is, a Euclidean quadrilateral). Because the sum of the angles in a Euclidean quadrilateral is 2π,[44] we have that

$$2\pi = \alpha_3 + \Big(\alpha_1 + \frac{\pi}{2}\Big) + \omega + \Big(\alpha_2 + \frac{\pi}{2}\Big) \tag{111a}$$

$$= (\alpha_1 + \alpha_2 + \alpha_3) + \omega + \pi, \tag{111b}$$

which implies that

$$\omega = \pi - (\alpha_1 + \alpha_2 + \alpha_3). \tag{112}$$

Therefore, the area enclosed by the triangle is

$$\text{Area}(\Sigma) = -\int_{\omega_1}^{\omega_2} d\omega = \int_{\omega_2}^{\omega_1} d\omega \tag{113a}$$

$$= \pi - (\alpha_1 + \alpha_2 + \alpha_3) = -(\text{TWR}). \tag{113b}$$

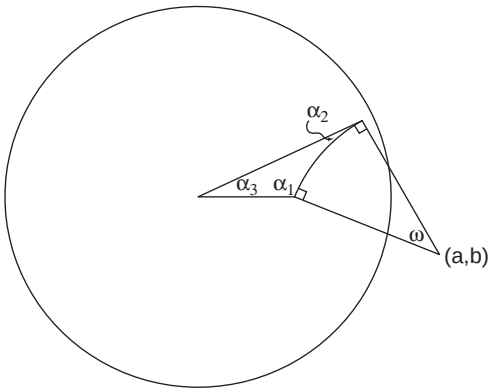


FIG. 14: The non-Euclidean area of a non-Euclidean triangle is equal to its angular defect $\pi - (\alpha_1 + \alpha_2 + \alpha_3) = \omega$.

Note that because we have proved the area of our special triangle is just the angular sweep of its one curved side, it is easy to prove that the area of any geodesic-sided polygon is the sum of the angular sweeps of its sides (about their various centers of curvature).

We have thus proved the main result of this section, that the negative of the Thomas-Wigner rotation is equal to both the (rapidity space) area enclosed by the rapidity space triangle and the angular defect (π minus the sum of the interior angles of the rapidity space triangle). Although we have derived this result from first principles, it was pointed out by Aravind,[24] and later discussed in a slightly different context by Criado and Alamo.[25]

Although the result (113b) is interesting in its own right, it also suggests that another way to evaluate and compare Thomas-Wigner rotations is to look at the areas of the corresponding triangles in rapidity space. Although possible in principle, in practice this is not very easy to do because areas in rapidity space depend on where they are located, and thus are not readily compared using our Euclidean-trained eyes. More specifically, as Eq. (49c) for the area element $d\sigma$ shows, although two regions in different parts of rapidity space may appear to have the same area to our Euclidean-trained eyes, the area of the one closest to the edge of the disk is larger.

C. Various Equations for the Thomas-Wigner Rotation Angle

We now are in a position to derive the various expressions for the Thomas-Wigner rotation angle that have appeared in the literature. Because most of these expressions are for the

magnitude of the TWR, we can use Eqs. (103) and (113b) to write

$$\Big|\text{TWR}\Big| = \Big|\oint \Big(\frac{2}{1-r^2}\Big)\Big(xdy - ydx\Big)\Big| \tag{114a}$$

$$= \oint \Big(\frac{2}{1-r^2}\Big)|(\vec{r}\times d\vec{r})\cdot\hat{k}| \tag{114b}$$

$$= \oint \Big(\frac{2r^2}{1-r^2}\Big)\Big|\frac{(\vec{r}\times d\vec{r})\cdot\hat{k}}{r^2}\Big|. \tag{114c}$$

If we use Eqs. (64) and (65), we find that

$$\frac{2r^2}{1-r^2} = \gamma - 1. \tag{115}$$

Equation (67) can then be used to show that for any infinitesimal segment of the path,

$$\left|\frac{(\vec{r}\times d\vec{r})\cdot\hat{k}}{r^2}\right| = \left|\frac{(\vec{v}\times d\vec{v})\cdot\hat{k}}{v^2}\right|, \tag{116}$$

which means that Eq. (114c) can be rewritten as

$$|\text{TWR}| = \oint_C (\gamma - 1)\left|\frac{\vec{v}\times d\vec{v}}{v^2}\right|. \tag{117}$$

Equation (117) can be re-expressed in various forms. For example, if we call the integrand $d\chi$, then

$$d\chi \;=\; \frac{\gamma-1}{v^2}\,|\vec{v}\times d\vec{v}| \tag{118a}$$

$$=\; \frac{\gamma-1}{v^2}\,|\vec{v}\times\vec{a}|\,dt \tag{118b}$$

Hence,

$$\frac{d\chi}{dt} = \frac{\gamma-1}{v^2}\,|\vec{v}\times\vec{a}|\,. \tag{119}$$

Equation (118a) is the expression for the Thomas-Wigner rotation angle given in Ref. 5, p. 289 and Ref. 6, p. 178.

Several interesting physical properties can be deduced from Eq. (117). First, the right-hand side tends to zero in the nonrelativistic limit, showing that in this limit the Thomas-Wigner rotation vanishes. Second, as $v \to c$, the Thomas-Wigner rotation angle increases without bound, as we deduced in Sec. VII A using the geometry of rapidity space. Third, the Thomas-Wigner rotation is a purely kinematic effect because it is independent of the dynamics causing the acceleration. In other words, it not only occurs for charged particles

moving in electromagnetic fields, but also can occur for elementary particles accelerated by nuclear forces,[38] and for masses accelerated by gravitational fields.

If we multiply the right-hand side of Eq. (119) by $(\gamma+1)/(\gamma+1)$ and use the identity given in Eq. (65), we find that

$$\omega = \frac{d\chi}{dt} = \frac{\gamma^2}{\gamma+1}\left|\vec{\beta}\times\frac{d\vec{\beta}}{dt}\right| = \frac{(\gamma-1)}{\beta^2}\left|\vec{\beta}\times\frac{d\vec{\beta}}{dt}\right|, \tag{120}$$

which is the expression for the angular speed of the Thomas-Wigner rotation given in Ref. 5, p. 290, Ref. 40, p. 554, and Ref. 6, p. 179. Because it is the angular velocity of the Thomas-Wigner rotation that enters into the calculation of the Thomas precession, it is this quantity that appears in the relativistic correction to the spin-orbit term in the Hamiltonian for a hydrogen atom.

D. Applying the Thomas Precession in Quantum Theory

Most derivations of the relativistic correction to the spin-orbit term in the Hamiltonian for a hydrogen atom relate the time rate of change of the electron's spin vector in its instantaneous rest frame to the corresponding rate in the lab (or proton's rest) frame.[40] Because any instantaneous rest frame of the electron is obtained from the previous instantaneous rest frame by a non-colinear Lorentz boost, the transformation back to the lab frame will include Thomas-Wigner rotations. The rate at which the Thomas-Wigner rotations occur is the rate given in Eq. (119). There are several excellent derivations of the correct form of the relativistic correction to the spin-orbit term (see, for example, Refs. 3, 4, or Ref. 42), and Eq. (119) is used in all of them.

Some authors[43] claim the factor of two that comes from including the Thomas precession in the spin-orbit term results from the electron's rest frame precessing through one complete cycle each time it completes one revolution around the proton. However, as Eqs. (120) and (124) show, this interpretation is incorrect because the number of rotations completed during each revolution is variable, and tends to infinity as $v \to c$.

E. A Special Case of Thomas Precession

The Thomas precession of an object is the sum of all the Thomas-Wigner rotations it undergoes when it completes one closed planar orbit. To see this explicitly, consider the

simple example of an object moving in a circle with a constant speed. If we use the area element given in Eq. (99), the expression for the magnitude of the Thomas precession for this case is

$$\big|\mathrm{TP}\big| = \int_0^{2\pi}\int_0^R \frac{4}{(1-r^2)^2} r\,dr\,d\theta \tag{121a}$$

$$= 2\pi \int_0^R \frac{4r}{(1-r^2)^2}\,dr. \tag{121b}$$

We can evaluate this integral by changing the integration variable to $u = (1-r^2)$. After some algebra, we find

$$\big|\mathrm{TP}\big| = 4\pi\Big(\frac{R^2}{1-R^2}\Big). \tag{122}$$

By using Eq. (115), we see that

$$\Big(\frac{R^2}{1-R^2}\Big) = \frac{\gamma-1}{2}, \tag{123}$$

which means that

$$\big|\mathrm{TP}\big| = 2\pi(\gamma-1). \tag{124}$$

Note that if the object moves around the circle in the clockwise (negative) direction, then the Thomas precession is in the opposite (positive) direction after one revolution around the circular path. This result also is derived in Ref. 6, p. 179.

VIII. MATHEMATICAL CONNECTIONS AND ALTERNATIVE EQUATIONS FOR THE THOMAS-WIGNER ROTATION

The purpose of this section is to give a brief discussion of the relation between the results presented in this paper and Möbius transformations, spinors, the group $SL_2(\mathbb{C})$, and models of the hyperbolic plane. The only new physical result is Eq. (146), which expresses the Thomas-Wigner rotation angle in terms of the rapidities that give rise to it and the angle between their corresponding boosts. Equation (146) also can be obtained geometrically in rapidity space; it is just easier to derive in the present context.

We have shown that rapidity space and the actions of Lorentz transformations on it provide valuable insight into the Thomas-Wigner rotation and the Thomas precession. Although our presentation has not required an actual algebraic expression for the action of Lorentz transformations on rapidity space, it is natural to ask for one. Once we have this

expression, it will be easy to relate Lorentz transformations to certain Möbius transformations (linear fractional transformations), and then to the spinor map between $SL_2(\mathbb{C})$ and the Lorentz group.

A. Lorentz Transformations of Rapidity Space

To see how the Lorentz transformation of Eq.(10) (a boost in the positive x-direction with rapidity ϕ) acts on a point (x, y) in the Poincaré disk, let $(x,\ y,\ (x^2+y^2+1)/2)$ denote the point on the paraboloid that projects to this point in the disk (see Fig. 3). If we apply the boost to this vector, we obtain

$$\begin{pmatrix} x \\ y \\ \frac{x^2+y^2+1}{2} \end{pmatrix} \mapsto \begin{pmatrix} (\cosh\phi)x - \sinh\phi\big(\frac{x^2+y^2+1}{2}\big) \\ y \\ -(\sinh\phi)x + \cosh\phi\big(\frac{x^2+y^2+1}{2}\big) \end{pmatrix}, \tag{125}$$

which we then need to rescale so that it lies on the paraboloid. Some rather messy algebra shows the correct scaling factor is

$$\lambda = \Big(\frac{\cosh\phi+1}{2} - (\sinh\phi)x + \frac{\cosh\phi-1}{2}(x^2+y^2)\Big)^{-1}, \tag{126}$$

and thus the boost maps points in the Poincaré disk by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \lambda\big((\cosh\phi)x - \sinh\phi(\frac{x^2+y^2+1}{2})\big) \\ \lambda y \end{pmatrix}. \tag{127}$$

Equation (127) can be expressed in a surprisingly simple way if we use complex notation to denote points in the disk. If we let $z = x + iy$, and set

$$a \ = \ \sqrt{\frac{\cosh\phi+1}{2}} = \cosh\frac{\phi}{2} \tag{128}$$

$$b \ = \ -\sqrt{\frac{\cosh\phi-1}{2}} = -\sinh\frac{\phi}{2}, \tag{129}$$

the action of the boost becomes

$$z \mapsto z' = \frac{az+b}{bz+a}. \tag{130}$$

Thus, when we use complex notation to label points on the Poincaré disk, the action of the boost can be expressed in a particularly simple way as a Möbius transformation.

Another special type of Lorentz transformation that is easy to analyze is a spatial rotation. It is not difficult to see that a counterclockwise spatial rotation by an angle of θ produces the map of the disk

$$z \mapsto z' = e^{i\theta} z = \frac{e^{i\theta/2} z + 0}{0z + e^{-i\theta/2}}, \tag{131}$$

which is again a Möbius transformation.

Given any Möbius transformation $z \mapsto \frac{az+b}{cz+d}$, with a, b, c, d normalized so that $ad - bc = 1$, we may associate[45] with it the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ of determinant 1. The composition of two Möbius transformations corresponds to multiplication of the corresponding matrices, and an inverse transformation corresponds to the inverse matrix. Thus the boost and rotation are associated with matrices

$$B_x(\phi) = \begin{pmatrix} \cosh\frac{\phi}{2} & -\sinh\frac{\phi}{2} \\ -\sinh\frac{\phi}{2} & \cosh\frac{\phi}{2} \end{pmatrix}, \quad R(\theta) = \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}, \tag{132}$$

which both have the rather special form,

$$M(\alpha, \beta) = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix}, \tag{133}$$

for complex numbers α and β with $\alpha\bar{\alpha} - \beta\bar{\beta} = 1$. Furthermore, for any such matrix $M(\alpha, \beta)$ with $\theta_1 = \arg(\alpha) + \arg(\beta) + \pi$, $\theta_2 = \arg(\alpha) - \arg(\beta) - \pi$, and ϕ such that $\cosh\frac{\phi}{2} = |\alpha|$ and $\sinh\frac{\phi}{2} = |\beta|$, we have

$$M(\alpha, \beta) = R(\theta_1) B_x(\phi) R(\theta_2). \tag{134}$$

Thus the matrices arising from boosts and rotations generate all matrices of the form in Eq. (133).

In fact, the Möbius transformations associated with matrices of the form in Eq. (133) are known to be all the (orientation-preserving) conformal maps of the Poincaré disk to itself.[46] Because every Lorentz transformation (on (2+1)-dimensional space) must give rise to a conformal map of the disk, and every such conformal map arises from a product of two rotations and a boost B_x, then not only do all the conformal maps arise from Lorentz transformations, but also every Lorentz transformation is a product of at most two rotations and a boost in the x-direction.

B. The Upper-Half Plane Model

The transformation associated with $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$ maps the disk conformally onto the set of points $z = x + iy$ with $y > 0$, and results in the upper half-plane model. The conformal transformations of this model are the Möbius transformations corresponding to 2×2 real matrices of determinant 1, that is, to the group $SL_2(\mathbb{R})$, since

$$\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}^{-1} \tag{135}$$

ranges through $SL_2(\mathbb{R})$ as α, β range through all complex numbers with $\alpha\bar{\alpha} - \beta\bar{\beta} = 1$.

Thus Lorentz transformations (on (2+1)-dimensional space) correspond to elements of $SL_2(\mathbb{R})$, and the action of a Lorentz transformation on rapidity space is simply the action of the corresponding Möbius transformation on the upper half-plane model.

C. Extension to Three Spatial Dimensions and the Spinor Map

Although our discussion has been limited to (2+1)-dimensions for ease of exposition, all the work carries over in a fairly straightforward way to (3+1)-dimensions (or more). A higher dimensional paraboloid within the light cone leads to a conformal model of rapidity space, which is now the interior of a unit ball. The metric is given by[47]

$$ds^2 = \frac{4}{(1 - x^2 - y^2 - z^2)^2}(dx^2 + dy^2 + dz^2), \tag{136}$$

and the geodesics are arcs of circles that intersect the bounding sphere orthogonally. Within the ball, the surfaces formed by pieces of spheres centered outside the unit ball which intersect the unit sphere orthogonally should be though of as "planar," because geodesics remain inside them. Any of these surfaces can be mapped (by a conformal transformation of the ball to itself) to a disk bounded by the equator of the ball. The geometry of such a disk arising from its embedding in the ball is the same as the geometry developed here for the Poincaré disk.

Finally, in addition to the ball model, there is an upper half-space model composed of points in $\mathbb{R}^3$ where the third coordinate is positive. Although points in it cannot be naturally identified by complex numbers — it is after all three-dimensional — they can be

identified with certain quaternions $x + iy + jz$, where $z > 0$. The (orientation-preserving) conformal transformations of this space are identified with matrices in $SL_2(\mathbb{C})$, where the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ acts by[48]

$$x + iy + jz \mapsto (a(x + iy + jz) + b)(c(x + iy + jz) + d)^{-1}. \tag{137}$$

The correspondence of Lorentz transformations, which give rise to conformal transformations of the model, to elements of $SL_2(\mathbb{C})$ is usually called the spinor map.[49]

D. More Useful Forms of the Thomas-Wigner Rotation

Because we have identified in the matrix $B_x(\phi)$ with a boost of rapidity ϕ in the x-direction and the matrix $R(\theta)$ with a spatial rotation through an angle θ (see Sec. VIII A), we can derive a relatively simple equation for the Thomas-Wigner rotation produced by two successive, non-colinear, boosts.

As is easily proved, a pure boost with rapidity ϕ in the direction of θ can be obtained by first rotating through $-\theta$, then applying an x-boost of ϕ, and then rotating back by θ. Expressing these three operations with matrices, we have

$$R(\theta)B_x(\phi)R(-\theta) = \begin{pmatrix} \cosh\frac{\phi}{2} & -\sinh\frac{\phi}{2}e^{i\theta} \\ -\sinh\frac{\phi}{2}e^{-i\theta} & \cosh\frac{\phi}{2} \end{pmatrix}. \tag{138}$$

Therefore, as shown in Figs. 7 and 8, a boost with a rapidity of ϕ_1 in the x-direction, followed by a boost of rapidity ϕ_2 in the $\theta = \pi - \alpha_1$ direction, corresponds to

$$R(\theta)B_x(\phi_2)R(-\theta)B_x(\phi_1) = \begin{pmatrix} \cosh\frac{\phi_2}{2} & -\sinh\frac{\phi_2}{2}e^{i\theta} \\ -\sinh\frac{\phi_2}{2}e^{-i\theta} & \cosh\frac{\phi_2}{2} \end{pmatrix} \begin{pmatrix} \cosh\frac{\phi_1}{2} & -\sinh\frac{\phi_1}{2} \\ -\sinh\frac{\phi_1}{2} & \cosh\frac{\phi_1}{2} \end{pmatrix}. \tag{139}$$

On the other hand, any Lorentz transformation in the direction of ω_1 can be expressed as the product of a boost $R(\omega_1)B_x(\phi_3)R(-\omega_1)$ in the ω_1 direction followed by a rotation through an angle ω_2. Expressing these operations with matrices, we have

$$R(\omega_2)\big(R(\omega_1)B_x(\phi_3)R(-\omega_1)\big). \tag{140}$$

In the specific case shown in Figs. 6, 7, and 8, ω_2 is the Thomas-Wigner rotation angle and ω_1 is the angle α_3. Thus, the product in Eq. (139) must equal the product in Eq. (140)

which, when expressed in matrix form, is

$$\begin{pmatrix} e^{i\omega_2/2} & 0 \\ 0 & e^{-i\omega_2/2} \end{pmatrix} \begin{pmatrix} \cosh\frac{\phi_3}{2} & -\sinh\frac{\phi_3}{2}e^{i\omega_1} \\ -\sinh\frac{\phi_3}{2}e^{-i\omega_1} & \cosh\frac{\phi_3}{2} \end{pmatrix}. \tag{141}$$

If we solve for ω_2 by equating the upper left entries of Eqs. (139) and (141), we find

$$\omega_2 = 2\arg\left(\cosh\frac{\phi_1}{2}\cosh\frac{\phi_2}{2} + \sinh\frac{\phi_1}{2}\sinh\frac{\phi_2}{2}e^{i\theta}\right) \tag{142a}$$

$$\omega_2 = 2\arg\left(1 + \tanh\frac{\phi_1}{2}\tanh\frac{\phi_2}{2}e^{i\theta}\right). \tag{142b}$$

Equation (142b) is an algebraic formula for the Thomas-Wigner rotation ω_2 resulting from a boost with rapidity ϕ_1 in the x-direction followed by a boost with rapidity ϕ_2 in the $\theta = \pi - \alpha_1$ direction (as shown in Figs. 7 and 8). Note that Eq. (142b) readily produces the qualitative results we derived in Sec. VII A. For example, it shows that the Thomas-Wigner rotation will take on values between $-\pi$ and π, and will approach its largest value when both velocities are near c and θ is near π. Equation (142b) also shows that the magnitude of the Thomas-Wigner rotation is the same regardless of the order in which the boosts ϕ_1 and ϕ_2 are applied.

We end this section by noting that the method used to derive Eq. (142b) also can be used to find the rapidity ϕ_3 and the angle α_3, and to derive the equations given in Ref. 24 for $\tan(\omega_2/2)$, $\cosh\phi_3$, and $\tan\alpha_3$. For example, if we equate the real and imaginary parts of the upper left entries of Eqs. (139) and (141), we find

$$\cos\frac{\omega_2}{2}\cosh\frac{\phi_2}{2} = \cosh\frac{\phi_1}{2}\cosh\frac{\phi_2}{2} + \sinh\frac{\phi_1}{2}\sinh\frac{\phi_2}{2}\cos\theta, \tag{143}$$

and

$$\sin\frac{\omega_2}{2}\cosh\frac{\phi_2}{2} = \sinh\frac{\phi_1}{2}\sinh\frac{\phi_2}{2}\sin\theta. \tag{144}$$

If we divide Eq. (144) by Eq. (143), we obtain

$$\tan\frac{\omega_2}{2} = \frac{\sinh(\phi_1/2)\sinh(\phi_2/2)\sin\theta}{\cosh(\phi_1/2)\cosh(\phi_2/2) + \sinh(\phi_1/2)\sinh(\phi_2/2)\cos\theta}, \tag{145}$$

which is Eq. (2) in Ref. 24. If we divide the numerator and denominator of Eq. (145) by $\sinh(\phi_1/2)\sinh(\phi_2/2)$, we obtain the simpler expression[29]

$$\tan\frac{\omega_2}{2} = \frac{\sin\theta}{\cos\theta + D}. \tag{146}$$

The coefficient D can be written as

$$D = \left(\frac{\cosh\phi_1/2}{\sinh\phi_1/2}\right)\left(\frac{\cosh\phi_2/2}{\sinh\phi_2/2}\right) \tag{147}$$

$$= \left(\frac{e^{\phi_1}+1}{e^{\phi_1}-1}\right)\left(\frac{e^{\phi_2}+1}{e^{\phi_2}-1}\right), \tag{148}$$

which, from Eq. (20), is simply a ratio involving Doppler blueshift factors. Alternatively, if we use Eqs. (128), (129), and (9b) in Eq. (147), we see that

$$D = \sqrt{\left(\frac{\gamma_1+1}{\gamma_1-1}\right)\left(\frac{\gamma_2+1}{\gamma_2-1}\right)}. \tag{149}$$

Equation (146), together with either Eq. (148) or (149), is the simplest expression we have seen for the Thomas-Wigner rotation angle ω_2.

IX. SUMMARY

We have presented a self-contained derivation of a relativistic velocity space called rapidity space. We then demonstrated how this space can be used to visualize and calculate various effects resulting from the successive application of non-colinear Lorentz boosts and the relativistic addition of non-colinear velocities. In particular, we showed how rapidity space provides a geometric approach to the Thomas-Wigner rotation and the Thomas precession, and how it offers both qualitative and quantitative insight into these (and other) effects.

Acknowledgements

We would like to thank Professors Elizabeth Allman and Matthew Côté for their generous and expert help with the figures. We also would like to thank the referees for bringing several interesting articles to our attention, and for pointing out that Eq. (142b) can be rewritten in the simpler form of Eq. (146).

---

* Electronic address: jrhodes@bates.edu

† Electronic address: msemon@bates.edu

[1] L. H. Thomas, "Motion of the spinning electron," Nature **117**, 514, 1926; ibid., "The Kinematics of an electron with an axis," Phil. Mag. **3**, 1–23, 1927.

[2] E. P. Wigner, "On unitary representations of the inhomogeneous Lorentz group," Ann. Math. **40**, 149–204 (1939).

[3] G. P. Fisher, "The Thomas precession," Am. J. Phys. **40**, 1772–1781 (1972). This article derives the Thomas-Wigner rotation and Thomas precession in several different ways and gives an excellent summary of previous treatments.

[4] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1998), 3rd ed., pp. 548–553, 563–564, 571.

[5] R. D. Sard, *Relativistic Mechanics* (Benjamin, New York, 1970), Chap. 5.

[6] H. Arzeliès, *Relativistic Kinematics* (Pergamon, New York, 1966), pp. 173–180, 198, 201–203.

[7] H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics* (Addison-Wesley, New York, 2002), 3rd ed., pp. 282–285.

[8] Some authors derive an approximation that is valid to second order in β, which is all that is needed to calculate the relativistic correction to the spin-orbit term in hydrogen. See, for example, R. A. Muller, "Thomas precession: Where is the torque?," Am. J. Phys. **60**, 313–317 (1992) and H. Kroemer, "The Thomas precession factor in spin-orbit interaction," Am. J. Phys., to appear.

[9] S. Gasiorowicz, *Quantum Physics* (Wiley, New York, 1996), 2nd ed., p. 282.

[10] R. L. Liboff, *Introductory Quantum Mechanics* (Addison-Wesley, San Francisco, 2003), 4th ed., p. 586.

[11] R. Shankar, *Principles of Quantum Physics* (Plenum, New York, 1981), p. 477.

[12] J. J. Sakurai and S. F. Tuan, *Modern Quantum Mechanics* (Addison-Wesley-Longman, New York, 1994), p. 305.

[13] One exception, which gives an elementary and clear example of the Thomas-Wigner rotation, is the recent paper by J. P. Costella, B. H. J. McKellar, and A. A. Rawlinson, "The Thomas rotation," Am. J. Phys. **69**, 837–847 (2001).

[14] A. A. Ungar, "Thomas precession and its associated grouplike structure," Am. J. Phys. **59**, 824–834 (1991).

[15] A. A. Ungar, *Beyond the Einstein Addition Law and its Gyroscopic Thomas Precession: The Theory of Gyrogroups and Gyrovectors* (Kluwer, Dordrecht Holland, 2001).

[16] A. Ben-Menahem, "Wigner's rotation revisited," Am. J. Phys. **53**, 62–66 (1985).

[17] H. Urbantke, "Physical holonomy, Thomas precession, and Clifford algebra," Am. J. Phys. **58**, 747–750 (1990).

[18] G. H. Goedecke, "Geometry of the Thomas precession," Am. J. Phys. **46**, 1055–1056 (1978).

[19] J. D. Hamilton, "Relativistic precession," Am. J. Phys. **64**, 1197–11201 (1996).

[20] E. G. P. Rowe, "Rest frames for a point particle in special relativity," Am. J. Phys. **64**, 1184–1196 (1996).

[21] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, New York, 1975), p. 36. The first edition of this book was published by Addison-Wesley in 1951.

[22] W. Pauli, *Theory of Relativity* (Dover, New York, 1981), pp. 73–74.

[23] B. A. Rosenfeld, *History of Non-Euclidean Geometry* (Springer-Verlag, New York, 1987), pp. 270–273.

[24] P. K. Aravind,"The Wigner angle as an anholonomy in rapidity space," Am. J. Phys. **65**, 634–636 (1997).

[25] C. Criado and N. Alamo,"A link between the bounds on relativistic velocities and areas of hyperbolic triangles," Am. J. Phys. **69**, 306–310 (2000).

[26] W. Rindler, *Relativity: Special, General and Cosmological* (Oxford Univ. Press, Oxford, 2001), pp. 43–44, 46.

[27] In Sec. VIII C we discuss the more general case of the three-dimensional relativistic velocity space obtained from four-dimensional spacetime.

[28] Ref. 26, pp. 52–53.

[29] We thank one of the referees for pointing this out to us.

[30] L. Parker and G. M. Schmieg, "Special relativity and diagonal transformations," Am. J. Phys. **38**, 218–222 (1970).

[31] In this case, by "conformal," we mean that angle measurements using a natural metric induced from Minkowski space (in a manner similar to the one developed in this section) coincide with those using the metric arising from viewing the surface as embedded in Euclidean 3-space.

[32] Because our model involves only two coordinates, x_1 and x_2, by "conformal" we mean that angle measurements coincide with those using the Euclidean metric in 2-space.

[33] There are other choices of A, B and C that will solve the three equations. First, if $A = C = 0$, then $B = \pm 1$ and $g = \pm r$. These solutions, however, make the denominator on the right-hand

side of Eq. (44) vanish. Second, if $B = 0$, then A can be arbitrary as long as $C = 1/(4A)$. In this case, however, if we use any value of A other than $A = 1/2$, we end up with the same model of a disk, but one whose radius is not equal to one. In this case everything is the same, just scaled appropriately.

[34] Although the coordinates in a two- or three-dimensional velocity space are the components of the velocity, the coordinates in rapidity space are not components of the rapidity; rather, they are proportional to the coordinates of the velocity, with a proportionality factor that varies from point to point. The space is called rapidity space simply to emphasize that the distance from the origin to any point in it is the rapidity of that point.

[35] This calculation implicitly assumes that the distance s is to be evaluated along a straight line. By definition, the distance between two points is the minimum of all the path integrals connecting them, which means that distances are evaluated along geodesics (unless otherwise stated). Thus, we are assuming that any geodesic that includes the origin is a straight line. We prove this assumption in Section VI.

[36] See, for example, Ref. 16, pp. 62–63. We give an algebraic proof of this result in Sec. VIII D.

[37] M. L. Boas, *Mathematical Methods for the Physical Sciences* (Wiley, 1983), p. 433.

[38] See, for example, Ref. 4, p. 552.

[39] Y. Z. Zhang, *Special Relativity and its Experimental Foundations* (World Scientific, 1997), Chap. 13.

[40] One exception is given in Ref. 41 in which the result is derived in the lab frame.

[41] G. Muñoz, "Spin-orbit interaction and the Thomas precession: A Comment on the Lab Point of View," Am. J. Phys. **69**, 554–556 (2001).

[42] D. J. Griffiths, *Introduction to Quantum Mechanics* (Prentice-Hall, New York, 1994), pp. 240–241.

[43] H. Haken and H. C. Wolf, *The Physics of Atoms and Quanta* (Springer-Verlag, New York, 2000), p. 193.

[44] This fact is easily proved by noting that any four-sided Euclidean figure can be constructed from two triangles.

[45] There are actually two such matrices associated with any Möbius transformation, because the matrices A and $-A$ correspond to the same transformation.

[46] John G. Ratcliffe, *Foundations of Hyperbolic Manifolds* (Springer-Verlag, New York, 1994), p.

129.

[47] This is the exercise from Ref. 21 discussed in Sec. I.

[48] Note that because of the non-commutativity of quaternions, the formula for this Möbius transformation must be expressed with the "denominator" on the right.

[49] For a more algebraic approach to the spinor map, see G. L. Naber, *The Geometry of Minkowski Spacetime* (Springer-Verlag, New York, 1992), or M. Carmeli and S. Malin, *An Introduction to the Theory of Spinors* (World Scientific, New Jersey, 2000).