The gravitational time delay of light, also called the Shapiro time delay, is one of the four classical tests of Einstein’s theory of general relativity. This article derives the Newtonian version of the Shapiro time delay from Einstein’s principle of equivalence and the Newtonian description of gravity, in a manner that is accessible to undergraduate students and advanced high-school students. The derivation can be used as a pedagogical tool, similar to the way that simplified derivations of the gravitational deflection of light are used in teaching about general relativity without making use of the more advanced mathematical concepts. Next, we compare different general-relativistic derivations of the Shapiro time delay from the Schwarzschild metric, which leads to an instructive example for the challenges of formulating the post-Newtonian limit of Einstein’s theory. The article also describes simple applications of the time delay formula to observations within our solar system, as well as to binary pulsars.
I. INTRODUCTION

The fourth classical test of general relativity is variously known as the gravitational time delay or the Shapiro delay, after Irwin I. Shapiro, who predicted the effect in 1964\cite{1} and conducted the first direct observations a few years later, using radar echoes within our solar system\cite{2,3}. The Shapiro delay has been measured with great accuracy using space probes with onboard transponders. One such measurement, using the Cassini probe in orbit around Saturn, yields the most stringent current test confirming that the propagation of light (both light deflection and the gravitational time delay) follows the predictions of Einstein’s theory of gravity\cite{4}. More recently, the Shapiro effect has played an important role in tests of general relativity in an environment with strong gravitational fields, namely near White Dwarfs or neutron stars, making use of binary pulsars.

Several derivations of the effect at the undergraduate level exist, which are based on the Schwarzschild metric in either Schwarzschild or in isotropic coordinates\cite{5} I am not aware of attempts to derive the gravitational time delay without recourse to the metric, at the same level as numerous existing derivations for the deflection of light\cite{6,7}. The aim of this article is to fill the gap, presenting a simple derivation, accessible to undergraduates and advanced high-school students, which deduces a location-dependent speed of light from the equivalence principle, and pays particular attention to the notion of the post-Newtonian and Newtonian limits of general relativity. The result of the simplified derivation deviates from the full general-relativistic result by a factor 2; this is the same as for Newtonian calculations of the gravitational deflection of light\cite{8}. Section \textsection{V} extends the analysis to descriptions using the Schwarzschild metric, comparing different text-book derivations of the Shapiro effect and pointing out ambiguities that can be used to prepare students for the necessity of defining a rigorous framework for post-Newtonian physics. Section \textsection{V} provides the information that will enable students to understand astronomical applications of the gravitational time delay, both in the solar system and for binary pulsars.

II. EQUIVALENCE PRINCIPLE AND COORDINATE SPEED OF LIGHT

General relativity contains both special relativity and Newtonian gravity as limiting cases: Take away all masses and all gravitational waves, so that all terms containing the gravitational constant $G$ can be neglected, and you obtain special relativity. Consider only speeds that are much slower than the speed of light, $v \ll c$, and you obtain Newton’s theory of gravity (and the associated laws of classical mechanics). The challenge of describing general-relativistic effects in a weak-gravity situation is the need to combine elements from Newtonian gravity and special relativity in a suitable way, at least approximately. This is made more difficult by general relativity’s underlying covariance: Coordinates do not automatically carry physical meaning.

A simple — albeit incomplete — way is to combine gravity as described by Newton’s theory with light propagation as derived from the Einstein Equivalence Principle (abbreviated EEP in the following). The EEP says that, in a free-falling reference frame, locally, the laws of physics are the same as those of special relativity. “Locally” means that the region is small enough, both spatially and in terms of the length of time our description is meant to cover, for us to be able to neglect the influence of tidal forces: changes of gravity from location to location, or over time, are too small to significantly affect the accuracy of our description. The laws of physics to which the EEP is applicable are not taken to include the laws governing gravity (that is left to the Strong Equivalence Principle), but they do include the special-relativistic law that has light moving along straight lines at the constant speed $c_0$. In the following, we will always write $c_0$ for the special-relativistic value for the speed of light, reserving non-annotated $c$, in particular with an argument, such as $c(r)$, for the local coordinate speed of light in a specified coordinate system.

Specifically, consider a static situation, with test particles moving in a Newtonian gravitational field. We choose Cartesian spatial coordinates $\vec{x}$ in which the geometry of space is Euclidean and the space coordinates faithfully reproduce physical lengths, and we choose a time coordinate $t$ such that, in these coordinates, the Newtonian equations for gravity hold at least for test particles with $v \ll c$. We will call that combination of time coordinate and space coordinates our Newtonian coordinates.

In that setting, a test particle with trajectory $\vec{x}(t)$ will experience the gravitational acceleration

$$\frac{d^2 \vec{x}}{dt^2} = -\vec{\nabla} \Phi(\vec{x}(t)), \quad (1)$$

where $\Phi$ is the gravitational potential. For simplicity, we will also assume spherical symmetry, with a potential $\Phi(r)$ depending only on the usual radial coordinate $r = |\vec{x}| \equiv \sqrt{x^2 + y^2 + z^2}$ for some chosen origin, although all of the following local arguments are also valid when considering a direction perpendicular to the local equipotential surface. We will also assume that $\lim_{r \to \infty} \Phi(r) = 0$, so that far from the center, physics is described in good approximation by the laws of special relativity. In particular, light in such regions will move at the usual constant speed $c_0$. 


The simplest case that complies with all our assumptions is that of a point mass with mass $M$ located at the origin $\vec{x} = 0$, which corresponds to a gravitational potential

$$\Phi(r) = -\frac{GM}{|\vec{x}|} = -\frac{GM}{r}. \quad (2)$$

Given our assumptions, there are two sets of properties that are not yet fixed: Consider a set $S$ of clocks, each at rest at a fixed location in the Newtonian coordinate system. We will refer to observers who are at rest relative to one of those clocks, who make use of the clock’s proper time for their time measurements and measure spatial distances using the local Newtonian space coordinates, as stationary observers. The proper time rates for the $S$ clocks are not determined by our Newtonian framework, since purely Newtonian physics only knows about coordinate time, taken to represent absolute time. Also, within Newtonian physics alone, limited as it is to $v \ll c$, we cannot make any deductions about the propagation of light.

We will now derive the proper times of clocks at rest in our gravitational field, and the laws of light propagation in the shape of a location-dependent coordinate speed of light $c(\vec{x})$ in the Newtonian coordinates, with the following strategy: Whenever we want to calculate effects of gravity in the regime $v \ll c$, we will use the Newtonian formulas. Whenever we want to study the propagation of light, we will introduce suitable local inertial frames in free fall, and use the EEP to describe how light propagates in those frames. We will go back and forth between the two kinds of description — Newtonian coordinates vs. the coordinates of the local inertial frames — as appropriate, linking their results in order to derive the local proper time rate and the local coordinate speed of light at each location.

To this end, consider two locations in our scenario, one at radial coordinate value $r$, the other infinitesimally higher at $r + dr$, as shown in Fig. 1. At each location, we have written down the value of the potential $\Phi$ and the rate

$$\frac{d\tau}{dt}(r)$$

relating the Newtonian time coordinate $t$ to the proper time $\tau$ of the $S$ clock at rest at fixed $r$. In order to see that the coordinate speed of light, which we have denoted $c(r)$, only depends on $r$, consider an inertial system $I$ in free fall that, at the event $E$ when the light passes a specific clock, is at rest in the Newtonian coordinate system. At $E$, length measurements in $I$ coincide with length measurements by an observer attached to one of the $S$ clocks, since both frames are at relative rest, and the length of an infinitesimal time interval as measured in $I$ is the same as the proper time $\tau(r)$ of the $S$ clock. In the free-fall inertial frame, light moves at the speed $c_0$; thus, as measured by the $S$ clock and the local space coordinates, the observer at rest relative to the $S$ clock, using the spatial coordinates and its own proper time, will find the light to be moving at the same speed,

$$\left| \frac{d\vec{x}}{d\tau} \right| = c_0. \quad (4)$$

With our assumption of spherical symmetry, the rates of our stationary $S$ clocks can only depend on $r$. In consequence, the coordinate speed of light in our Newtonian coordinates must be

$$c(r) = \left| \frac{d\vec{x}}{dt} \right| = \left| \frac{d\vec{x}}{d\tau}(r) \right| \cdot \frac{d\tau}{dt}(r) = c_0 \cdot \frac{d\tau}{dt}(r). \quad (5)$$

Now assume that a light signal is travelling from an observer at rest at $r + dr$ to an observer at rest at $r$. If the light wave leaves the upper observer at the time $t$, it will arrive at the lower observer at the time

$$t + dt = t + \frac{dr}{c(r)}. \quad (6)$$
Introduce two inertial frames in free fall: \( \mathcal{I}_1 \) is at rest relative to the upper stationary observer at \( r + dr \), at the exact event when the light is emitted, and \( \mathcal{I}_2 \) is at rest relative to the lower stationary observer at \( r \) at the event where the light is received. By this definition, an observer in \( \mathcal{I}_1 \) will measure the same wavelength \( \lambda_E \) for the light as the upper stationary observer at the time of emission, and \( \mathcal{I}_2 \) the same wavelength \( \lambda_R \) as the lower stationary observer at the time the light is received.

Between the time of emission and the time of reception, the free-falling frame \( \mathcal{I}_1 \) has gained a speed \( dv \) in our Newtonian \( r-t \) coordinate system. According to (10), applied to purely radial motion, and inserting the time interval \( \Delta t \), we have

\[
dv = \frac{d^2r}{dt^2} \cdot dt = - \frac{d\Phi}{dr} \cdot dt = \frac{d\Phi}{dr} \cdot \frac{dr}{c(r)} = - \frac{d\Phi}{c(r)}.
\]

(7)

This is only the coordinate speed, however. The observer in \( \mathcal{I}_2 \), whose time is momentarily the same as the proper time \( \tau \) of the stationary observer at position \( r \), has a measure of time that differs from the Newtonian coordinate time \( t \), and will instead measure a small velocity \( dv_I \) for the inertial system \( \mathcal{I}_1 \) that is related to \( dv \) by

\[
dv = dv_I \cdot \frac{d\tau}{dt} = dv_I \cdot \frac{c(r)}{c_0}.
\]

(8)

By the EEP, we are allowed to use the laws of special relativity to express the relation of a wavelength for light as measured in \( \mathcal{I}_1 \) and the same wavelength as measured in \( \mathcal{I}_2 \): the two are linked by the Doppler formula, and since \( dv_I \) is infinitesimally small, we are allowed to apply the classical Doppler formula, which yields

\[
z \equiv \frac{\lambda_R - \lambda_E}{\lambda_E} = \frac{dv_I}{c_0} = - \frac{d\Phi}{c(r)^2}.
\]

(9)

This expression is related to the time it takes light to move past a clock: In \( \mathcal{I}_1 \), one wavelength of light will take the time \( \Delta \tau = \lambda/c_0 \) to move past at the speed \( c_0 \), and correspondingly for \( \mathcal{I}_2 \). But in our static Newtonian coordinates, one peak of the wave will take the same coordinate time to move from \( r + dr \) down to \( r \) as the peak following it; in consequence, wave peaks that have been sent by the upper stationary observer a coordinate time interval \( \Delta \tau \) apart will arrive at the location of the lower stationary observer the same coordinate time interval \( \Delta t \) apart. Thus, the two proper time intervals in \( \mathcal{I}_1 \) and in \( \mathcal{I}_2 \) are proper time intervals on the local \( \mathcal{S} \) clocks of the upper and the lower observer that correspond to the same coordinate time interval \( \Delta t \), so that

\[
\frac{d\tau(r + dr)}{d\tau(r)} = \frac{\lambda_E}{\lambda_R} = \frac{1}{1 + z} \approx 1 - z = 1 - \frac{d\Phi}{c(r)^2},
\]

(10)

where in the last step we have used eq. (9). With these results, and taking the presence of infinitesimal entities such as \( dr \) to imply the taking of appropriate limits, we can compute

\[
\frac{dc(r)}{dr} = \frac{c(r + dr) - c(r)}{dr} = \frac{d\tau(r + dr) - d\tau(r)}{dr \cdot dt} \cdot \frac{c_0 \cdot d\tau(r + dr) - c_0 \cdot d\tau(r)}{c_0 \cdot d\tau(r)} \\
= \left[ \frac{d\tau(r + dr)}{d\tau(r)} - 1 \right] \cdot c_0 \cdot \frac{d\tau(r)}{dt} \cdot \frac{d\tau(r + dr)}{d\tau(r)} - 1 \right] \cdot c(r) \cdot \frac{d\Phi}{c(r)^2} \cdot \frac{dr}{dt}.
\]

(11)

Let us integrate up both sides from a specific \( r \) value to infinity, remembering that \( \Phi(\infty) = 0 \) and \( c(\infty) = c_0 \). As a result, we obtain the equation

\[
c(r) = c_0 \sqrt{1 + \frac{2\Phi(r)}{c_0^2}}.
\]

(12)

For a single spherically-symmetric mass, this comes out as

\[
c(r) = c_0 \sqrt{1 - \frac{2GM}{rc_0^2}} = c_0 \sqrt{1 - \frac{R_{SSR}}{r}},
\]

(13)
where in the last equation, we have introduced the Schwarzschild radius corresponding to our mass $M$, defined as \( R_{SSR} \equiv 2GM/c_0^2 \). The Schwarzschild radius characterises the size of a spherical black hole of mass $M$, and more generally represents the typical length scale at which strong gravitational effects become important in a given situation.

Note that this result is in direct contradiction to what we would have expected, had we treated light like a Newtonian particle, governed by classical mechanics. The classical conservation of energy for light moving at the speed $c_0$ when infinitely far from a central mass would yield

\[
c(r) = c_0 \cdot \sqrt{1 + \frac{2GM}{r_0^2}}, \tag{14}
\]

which is larger than $c_0$ as the light approaches the mass. As we would expect from special relativity, Newtonian calculations give incorrect results when applied to entities travelling near or at the speed of light. This raises the question of how Newtonian calculations, such those of Soldner or Cavendish, can yield results that are at least qualitatively (that is, up to an overall factor 2) correct. An important part of the answer is likely to be that, in order to reconstruct the shape of an orbit in the x-y plane, all that is needed is a valid expression for the ratio of the velocity components

\[
\frac{v_x}{v_y}, \tag{15}
\]

Re-scaling both velocity components with the same location-dependent factor $c(r)$ does not change that ratio. In calculating the gravitational time delay, on the other hand, the correct form of $c(r)$ plays the central role.

### III. NEWTONIAN GRAVITATIONAL TIME DELAY

With the result (13) for $c(r)$, let us calculate the gravitational time delay near a point mass, with Newtonian potential (2). The simplest situation is that shown in Fig. 2 where we have light from a planet $P$ pass the Sun $S$ on its way to Earth $E$. We have approximated the trajectory of the light as a straight line from $P$ to $E$, and placed the origin of our x-y coordinate system at the point $O$ of the light’s closest approach to the Sun. In these coordinates, the planet is at $(x_P, 0)$, the Earth at $(x_E, 0)$ and the Sun at $(0, -\eta)$, where $\eta$ is called the impact parameter. The radius values are $r_E = \sqrt{x_E^2 + \eta^2}$ and $r_P = \sqrt{x_P^2 + \eta^2}$.

When light moving in the x direction at $y = 0$ is at the coordinate position $x$, its distance from the Sun is $r = \sqrt{x^2 + \eta^2}$, and from the expression (13) for $c(r)$, we know that the light will be moving at the coordinate speed

\[
\frac{dx}{dt} = c_0 \left[ 1 - \frac{GM}{c_0^2 \sqrt{x^2 + \eta^2}} \right], \tag{15}
\]

where we have used the approximation $\sqrt{1 + \xi} \approx 1 + \xi/2$. In order to integrate this expression, we separate the variables $x$ and $t$ and, after having made use of the identity $(1 - \xi)^{-1} \approx 1 + \xi$, we obtain

\[
\frac{dt}{c_0} = \frac{dx}{c_0} \left[ 1 + \frac{GM}{c_0^2 \sqrt{x^2 + \eta^2}} \right]. \tag{16}
\]

Integrating up from the time $t_P$ the light left the planet to the time $t_E$ the light reached Earth, we find that

\[
t_E - t_P = \int_{t_P}^{t_E} dt = \int_{x_P}^{x_E} \frac{dx}{c_0} \left[ 1 + \frac{GM}{c_0^2 \sqrt{x^2 + \eta^2}} \right] = \frac{x_E - x_P}{c_0} + \frac{GM}{c_0^2 \eta} \ln \left[ \frac{x_E + r_E}{x_P + r_P} \right], \tag{17}
\]
where we have made use of the integral formula
\[
\int \frac{\, \, \, \, \mathrm{d}x}{\sqrt{x^2 + a^2}} = \ln \left[ \sqrt{x^2 + a^2} + x \right]
\]  
which is straightforward to verify, albeit not so straightforward to derive from scratch. The linear term in (17) is what we would have expected for light travelling at the constant speed \( c_0 \). The logarithmic term is the additional contribution of the gravitational time delay.

While the coordinate choice we made was particularly well-suited for the calculation, there is a way of re-writing our result in a somewhat different form, which is closer to the conventions in the literature, using the quantities defined in Fig. 3. This time, we describe positions with the help of position vectors anchored at the Sun \( S \). Introducing the unit vector \( \vec{n} \) pointing from \( P \) to \( E \) along the direction of travel of the light, we have \( x_P = \vec{n} \cdot \vec{x}_P \) and \( x_E = \vec{n} \cdot \vec{x}_E \). Multiply numerator and denominator in the argument of the logarithm in (17) by \( r_p - x_P \) and substitute the new position vectors, and the travel time becomes
\[
t_E - t_P = \frac{|\vec{x}_E - \vec{x}_P|}{c_0} + \frac{GM}{c_0^2} \ln \left[ (r_E + \vec{n} \cdot \vec{x}_E)(r_P - \vec{n} \cdot \vec{x}_P) \right].
\]  
Comparing this expression with a proper calculation in the framework of the post-Newtonian formalism for general relativity, one finds that the Shapiro delay term is off by a factor 2: The general-relativistic effect is twice as large as the Newtonian effect calculated here. This is exactly the same as for simplified Newtonian derivations of the gravitational deflection of light; there, too, the general-relativistic deflection angle is twice as large as the angle predicted from a Newtonian perspective. We will follow up on that factor in section IV.

Taken together, the results of this section suggest one possible way of teaching about the Shapiro delay in a setting where direct derivations from the Schwarzschild metric are not an option: Begin by presenting the simplified derivation developed in this section. This will yield a result that has the correct functional dependence on the geometry, but is off by an overall factor 2. Give the students the additional information that a more thorough derivation, which includes the curvature of space, will yield a result that includes the additional factor 2. After that statement, you can use the corrected formula, with the extra factor of 2, to consider applications such as the ones presented in section V, where the Shapiro time delay formula is used to compare predictions with data.

IV. GRAVITATIONAL TIME DELAY, SCHWARZSCHILD METRIC AND POST-NEWTONIAN AMBIGUITIES

Before we move on to applications, let us consider the origin of the factor 2 that comes out of more rigorous calculations. The results will also be of interest for teachers who derive the Shapiro time delay formula directly from the Schwarzschild metric, as it will highlight some of the challenges of teaching about post-Newtonian approximations in general.

Consider a light-like geodesic in the Schwarzschild metric, written in Schwarzschild’s original coordinates,
\[
ds^2 = - \left[1 - \frac{2GM}{c_0^2 r} \right] c_0^2 \mathrm{d}t^2 + \left[1 - \frac{2GM}{c_0^2 r} \right]^{-1} \mathrm{d}r^2 + r^2 (\mathrm{d}\theta^2 + \sin^2 \theta \cdot \mathrm{d}\varphi^2).
\]  
In these coordinates, the (spherical) mass \( M \) is located at the origin of the coordinate system. Let us choose the orientation of our angular coordinates so that the positions of the two planets involved, as well as the trajectory of the light, are all in the equatorial plane \( \theta = \pi/2 \). Fig. 4 shows part of that plane, where the origin is now at the...
center of the Sun $S$. Along the line $EP$, we have $x = r \sin \varphi$. Changing the $x$ coordinate on this line by $dx$, we must change $r$ by $dr = x/\sqrt{x^2 + \eta^2} \cdot dx$ and $\varphi$ by $d\varphi = dx[\eta/(\eta^2 + x^2)]$. Inserting those two infinitesimal spatial coordinate changes into the metric \eqref{20} and setting $ds^2 = 0$ as the condition for light-like geodesics, we obtain

$$\frac{dt}{c_0} = \frac{dx}{c_0} \left[ 1 + \frac{2GM}{c^2_0 r^2} - \frac{GM\eta^2}{c^2_0 r^3} \right], \quad \text{(21)}$$

where we have already made use of $2GM/c^2_0 \ll r$, which is valid in weak gravitational fields, to simplify the right-hand side. Compared with \eqref{16}, we find that there is indeed a new factor of 2 in the $1/r$ term. But there is also an extra term, which is proportional to $1/r^3$.

Before following up on this extra term, consider an alternative calculation. Define the new radial coordinate $\bar{r}$ via

$$r = \left( 1 + \frac{GM}{2c^2_0 r} \right)^2 \bar{r}. \quad \text{(22)}$$

Using this coordinate plus $t$ and the angular coordinates, we can write down the Schwarzschild metric in isotropic coordinates,

$$ds^2 = -\left[ \frac{1 - GM}{2c^2_0 \bar{r}} \right]^2 c^2_0 dt^2 + \left[ 1 + \frac{GM}{2c^2_0 \bar{r}} \right]^4 \left( d\bar{r}^2 + \bar{r}^2 (d\theta^2 + \sin^2 \theta \cdot d\phi^2) \right). \quad \text{(23)}$$

Introducing Sun-centered Cartesian coordinates, with

$$d\bar{r}^2 + \bar{r}^2 (d\theta^2 + \sin^2 \theta \cdot d\phi^2) = dx^2 + dy^2 + dz^2, \quad \text{(24)}$$

and again restricting ourselves to a situation where motion is in the $x$ direction, so $dy = dz = 0$, we have

$$\frac{dt}{c_0} = \frac{dx}{c_0} \left[ 1 + \frac{GM}{2c^2_0 \bar{r}} \right]^{3/2} \approx \frac{dx}{c_0} \left[ 1 + \frac{2GM}{c^2_0 \bar{r}} \right]. \quad \text{(25)}$$

This differs from \eqref{16} only by the extra factor 2 in the $1/\bar{r}$ term. There is no extra term proportional to $1/\bar{r}^3$. Note that the differences between \eqref{21} and \eqref{25} are considerable. Fig. 5 shows plots of the quantity

$$C(x, \eta) = \left[ c_0 \frac{dt}{dx} - 1 \right]^{-1} \left[ \frac{GM}{c^2_0} \right]^{-1}. \quad \text{(26)}$$

at $x = 1$ AU for various values of $\eta$, for all three of our cases: The Newtonian expression for $dt/dx$ that follows from \eqref{16}, the version that follows in Schwarzschild’s original coordinates from \eqref{21}, and the version that follows from the calculation in isotropic coordinates, in \eqref{25}. As the figure demonstrates, for typical values within our solar system, the formulae give markedly different predictions. The maximum value for $C(x, \eta)$ is the same for the calculation in Schwarzschild coordinates and that for isotropic coordinates, but as we move to larger $\eta$, the two results differ significantly. The Schwarzschild coordinate calculation can be seen to interpolate between the isotropic result for small $\eta$ and the Newtonian further out.

This discrepancy — only one of the formulae can give the right approximation! — is an instructive example of the ambiguities of taking the Newtonian limit. At infinity, both isotropic coordinates and Schwarzschild coordinates have
FIG. 5. The quantity $C(x, \eta)$ for $x = 1$ AU and various values for $\eta$, comparing the Newtonian case, the result in Schwarzschild coordinates and the results in isotropic coordinates.

the same flat-space limit. But the ways this limit is approached varies from one case to the other, leading to different correction terms of first order in $GM/(c^2_0 L)$ with $L$ an appropriate length scale, such as $r$ or $\bar{r}$.

While physical results are not coordinate-dependent, a Taylor expansion in specific coordinates, meant to capture general-relativistic corrections to Newtonian gravity, evidently is coordinate-dependent. As a result, even measurable, physical quantities like the Shapiro delay can come out differently, depending on the coordinate system chosen to calculate them. The comparison of the two different results (21) and (25) is a simple demonstration of the necessity of an approach to post-Newtonian corrections that is more systematic than merely collecting first-order terms in the given coordinate system — a clear motivation for the development of a proper post-Newtonian expansion (whose details are, of course, too complicated to fit into an introductory undergraduate class).

Comparison with the post-Newtonian expansion (22) shows that the derivation based on isotropic coordinates is correct. It also gives an indication why, namely that the isotropic coordinates are similar to the harmonic coordinates that form the basis of the post-Newtonian formalism; in these harmonic coordinates, Einstein’s equations can be re-formulated at least approximatively as a simple wave equation, which serves as the foundation of the series approximation at the heart of the post-Newtonian formalism.

General relativity text books tend to present either the Schwarzschild-coordinate-based derivation or the one based on isotropic coordinates, but not the discrepancy between the two. Historically, Shapiro’s first derivation is based on Schwarzschild coordinates, while the formulae he later used to fit his observational results are those based on isotropic coordinates. The differences between the two have been discussed in depth in Ref. 16.

Returning to the simpler teaching strategy of deriving the Newtonian approximation, and analysing the additional information from the Schwarzschild metric, both the metric derivations show us directly where the factor 2 is coming from: By construction, our Newtonian derivation only included the relation between proper time $t$ and coordinate time $\tau$, which in both forms of the metric is encoded in the $dt^2$ term. In our metric calculations, this term only contributes half of the final result for the term that later on becomes the pre-factor of the logarithmic delay term. The other half comes from the spatial part of the metric — which our Newtonian derivation using the EEP could not capture, given its assumption that space is Euclidean. In short, half of the general-relativistic result comes from the time part, half from the space part. Newtonian gravity can be interpreted as a distortion of time only, and cannot capture the contributions from the distortion of space.

This shortcoming of the Newtonian version is a general feature of derivations based on the equivalence principle, and in complete analogy with Newtonian derivations of the deflection of light, where the neglect of non-Euclidean spatial geometry leads to the same factor 2 discrepancy.

V. APPLICATIONS: SOLAR SYSTEM AND PULSARS

Formulas (17) and (19) for one-way travel, corrected by the multiplication of the delay term with an overall factor 2 to go from the Newtonian to the general-relativistic result,

$$\Delta t = \frac{2GM}{c^2_0} \ln \left[ \frac{x_E + r_E}{x_P + r_P} \right] = \frac{2GM}{c^2_0} \ln \left[ \frac{(r_E + \hat{n} \cdot \bar{x}_E)(r_P - \hat{n} \cdot \bar{x}_P)}{\eta^2} \right],$$

(27)
can readily be applied to specific situations in which the Shapiro time delay can be measured. Let us start with the typical situation for measurements within our solar system. In that case, the mass is the Sun, and we are looking at a situation where a planet is almost directly behind the Sun, as seen from Earth — in other words: a situation close to what is known in astronomy as a superior conjunction of that planet, with the planet, the Sun and Earth all lined up, in that order. That is the situation where light travelling from us to the other planet or back will pass close to the Sun, and where the Shapiro time delay is particularly large.

Ephemeris software will typically give us the distances $r_E$ and $r_P$, plus the distance $r_{EP}$ between the planet and Earth, as well as the elongation angle $\alpha$, which is the angle $\angle SEP$ in Fig. 3. From these quantities, we can derive $\eta = r_E \sin \alpha$, $x_E = r_E \cos \alpha$, and $x_P = x_E - r_{EP}$. Fig. 6 shows the computed time delay for radar echoes sent to Venus before, during and after a particular superior conjunction. The light travel time is short compared to the rates of change in planetary positions, so the prediction for the Shapiro time delay for signals sent from Earth to Venus and reflected back to Earth is twice the delay predicted by (27), adding up the delay for the propagation from Earth to Venus and that from Venus to Earth for a single set of values for $r_P, r_E, r_{EP}$ and $\alpha$. The prediction is compared with selected data extracted from Fig. 1 in Shapiro’s 1971 article, ref. [3]. The ephemeris data was taken from the ephem Python package programmed by Brandon Rhodes. The implementation is straightforward enough that it could form part of a programming-based astronomy lab course. The data I have extracted from the Shapiro article for this purpose is available online at Ref. [24] together with the script I used to create Fig. 6.

A second interesting application is to binary pulsars, that is, to binary star systems where at least one of the stars is a pulsar. The first such system, the binary neutron star that contains the pulsar PSR B1913+16, was discovered by Joseph P. Taylor and Russell Hulse in 1974, and has enabled interesting tests of general relativity close to compact objects, including the first indirect detection of gravitational waves.

The following derivation is a pedagogical elaboration of the time-delay formula first found by Blandford and Teukolsky. We begin by describing the orbital plane of the binary, as shown in Fig. 7. The origin of the plane is the center of mass of the system, around which the two stars $S_1$ and $S_2$ revolve. Let the masses of the stars be $m_1$ and $m_2$, respectively. As shown in the diagram, the two stars are always on opposite sides of the origin. We describe the phase of their orbit by the angle $\phi$. The value of the angle for when the first star is at the periapsis point $P$, that is, in the position where the mutual distance between the two stars is minimal, is called the argument (alternatively: longitude) of pericenter, commonly denoted by $\omega$.

The mutual distance between the stars can be written as

$$r(\phi) = \frac{a(1 - e^2)}{1 + e \cos(\phi - \omega)},$$

using the standard polar coordinate formula for ellipses, where $a$ is the major half-axis and $e$ the numerical eccentricity of the ellipse that solves the reduced Kepler problem. The difference $\nu = \phi - \omega$ is called the true anomaly. Solving
FIG. 6. Comparison between the prediction of (26) and Shapiro’s radar echo data for Venus from 1970/1971

FIG. 7. Orbital plane of the binary containing a pulsar

the two-body Kepler problem is a standard exercise in celestial mechanics. By Kepler’s third law, $a$ can be expressed in terms of the orbital period $T$ and the system’s total mass as

$$a^3 = T^2 \cdot \frac{G(m_1 + m_2)}{4\pi^2}, \quad (29)$$

while there is a multi-step process to obtain the $\varphi$ as a function of time: The true anomaly $\nu$ is linked to the so-called eccentric anomaly $E$ by

$$\tan \nu = \sqrt{1 - e^2} \sin E \cos E - e, \quad (30)$$

while $E$ can be obtained from the mean anomaly $M$ by Kepler’s equation

$$M = E - e \sin E \quad (31)$$

(which cannot be solved analytically), and the mean anomaly can be written as a function of time as

$$M = \frac{2\pi}{T} (t - t_0), \quad (32)$$

with $t_0$ the time when both stars are in the pericenter position. Since our origin is at the center of mass, the distances of the two stars from the origin are given by

$$r_1(\varphi) = \frac{m_1}{m_1 + m_2} r(\varphi), \quad r_2(\varphi) = \frac{m_2}{m_1 + m_2} r(\varphi), \quad (33)$$

and as short-hand, we can introduce the parameters $p_1$ and $p_2$

$$p_{1,2} \equiv \frac{m_{1,2}}{m_1 + m_2} \cdot a(1 - e^2). \quad (34)$$

(The technical term for $p_{1,2}$ is that each is the “semi-latus rectum” of the elliptical orbit of the star in question.)

In the right-handed coordinate system in which the orbital plane of our stars is the $y’$-$z’$ plane shown in Fig. 7, the locations of the two stars as a function of $\varphi$ are

$$\vec{x}_1'(\varphi) = \frac{p_1}{1 + e \cos(\varphi - \omega)} \begin{pmatrix} 0 \\ \cos \varphi \\ \sin \varphi \end{pmatrix}, \quad \vec{x}_2'(\varphi) = -\frac{p_2}{p_1} \cdot \vec{x}_1'(\varphi). \quad (35)$$

In summary, in order to describe the motion of the stars in the orbital plane, we need to know the epoch $t_0$, the total mass $m_1 + m_2$, the eccentricity $e$, and the argument of the pericenter $\omega$. 
As we observe the double star from afar, there is an additional parameter related to the orientation of the orbital plane relative to us. The three-dimensional set-up can be seen in Fig. 8. We choose the center of mass of the system as the origin of our coordinate system, and the x axis as our line of sight from Earth, which we take to be located at \( \bar{x}_E = (r_{CE}, 0, 0)^T \). We choose the y-z plane as shown in the figure, which amounts to choosing the longitude of the ascending node to be \( \pi/2 \). With this choice, we have one additional orbital parameter, namely the inclination \( \iota \) of the orbital plane relative to the yz plane, where \( \iota = \pi/2 \) corresponds to an edge-on and \( \iota = 0 \) to a face-on view of the system. We rotate our y'-z' system within its plane so that the y' axis coincides with the y axis. The parameters we have introduced so far are collectively known as the Keplerian orbital elements: time of epoch \( t_0 \), inclination \( \iota \), longitude of the ascending node (which we have chosen as \( \pi/2 \)), orbital eccentricity \( e \), semimajor axis \( a \) and argument of longitude \( \omega \). They are commonly used to specify orbits in classical mechanics.

From our definitions, it follows that the locations of stars 1 and 2 in our coordinate system are given by

\[
\bar{x}_1(\varphi) = \begin{pmatrix} \cos \iota & 0 & -\sin \iota \\ 0 & 1 & 0 \\ \sin \iota & 0 & \cos \iota \end{pmatrix} \cdot \bar{x}_1'(\varphi) = \frac{p_1}{1 + e \cos(\varphi - \omega)} \begin{pmatrix} -\sin \varphi \sin \iota \\ \cos \varphi \\ \sin \varphi \cos \iota \end{pmatrix}
\]

and

\[
\bar{x}_2(\varphi) = -\frac{p_2}{p_1} \bar{x}_1(\varphi).
\]

Let star 1 be the pulsar whose regular signals we can observe, and star 2 the compact mass causing the Shapiro delay. In line with our previous conventions, the location of star 2 will be the origin of our coordinate system. We will not aim at a precise description, but will freely make suitable approximations. Notably, the length scales \( p_{1,2} \) are both much smaller than the distance between the center of mass and Earth \( r_{CE} \), so we will keep only the leading-order terms in \( p_{1,2}/r_{CE} \ll 1 \). In order to be able to apply formula (27) for the Shapiro time delay in this system, we need to shift the origin of our coordinates to the location of the second star. In that new system, the position vector of the Earth is approximately

\[
\bar{x}_E''(\varphi) = \bar{x}_E - \bar{x}_2(\varphi) = \begin{pmatrix} r_{CE} \\ 0 \\ 0 \end{pmatrix},
\]

where we have only kept the leading (zeroth order) term in \( p_2/r_{CE} \ll 1 \). The position of the pulsar is

\[
\bar{x}_1'' = \bar{x}_1(\varphi) - \bar{x}_2(\varphi) = \frac{a(1 - e^2)}{1 + e \cos(\varphi - \omega)} \begin{pmatrix} -\sin \varphi \sin \iota \\ \cos \varphi \\ \sin \varphi \cos \iota \end{pmatrix}.
\]
TABLE I. Selected post-Keplerian and Keplerian orbital parameters for binary star systems containing at least one pulsar. PSR 1913+16 data from Tables 5 [DD(1) and DD(3)] from ref. 26, PSR J0737-3039A/B data from Table 2 in ref. 27, PSR J1614-2230 derived from data from Table 1 in ref. 28. The meaning of the parameters is explained in the text.

| System          | $T$ [d]   | $e$          | $\omega$ [deg] | $r_{Sh}$ [µs] | $s_{Sh}$                     |
|-----------------|-----------|--------------|----------------|----------------|----------------------------|
| PSR 1913+16     | 0.322997  | 0.6171       | 220.14         | 6.83           | 0.734                      |
| PSR J0737-3039A/B | 0.10225156248(5) | 0.0877775(9) | 87.0331(8)     | 6.21(33)       | 0.99974(-39,+16)            |
| PSR J1614-2230  | 8.6866194196(2) | 1.30(4) x 10^{-6} | 175(2)         | 2.46           | 0.999894(5)                 |

The normal vector pointing from the pulsar to Earth is approximately $\vec{n} = (1, 0, 0)^T$, again keeping only the leading (zeroth order) term in $p_1 / r_{CE} \ll 1$. From the two position vectors and the normal vector, it follows that the argument in the left-hand version of the Shapiro time delay term in equation (27) is

$$\frac{|\vec{x}'_E| + \vec{n} \cdot \vec{x}'_E}{|\vec{x}'_1| + \vec{n} \cdot \vec{x}'_1} = \frac{2r_{CE}}{a(1 - e^2)} \cdot \frac{1 + e \cos(\varphi - \omega)}{1 - \sin \varphi \sin \iota}.$$  

(40)

In the logarithm, the constant factor $2r_{CE}/a(1 - e^2)$ produces a constant time shift that is undetectable, as we can only measure how the pattern of pulsar pulses varies over time, and not their absolute timing. Inserting the time-dependent factors of (40) into (27), we find the time-dependent portion of the Shapiro delay to be

$$\Delta t(\varphi) = \frac{2Gm_2}{c_0^3} \ln \left[ \frac{1 + e \cos(\varphi - \omega)}{1 - \sin \varphi \sin \iota} \right] = 2r_{Sh} \ln \left[ \frac{1 + e \cos(\varphi - \omega)}{1 - s_{Sh} \cdot \sin \varphi} \right],$$  

(41)

where the second formula introduces the two parameters characteristic for the Shapiro time delay, namely the range

$$r_{Sh} \equiv \frac{Gm_2}{c_0^3} \approx 5 \, \mu s \cdot \frac{m_2}{M_\odot}$$  

(42)

and the shape

$$s_{Sh} \equiv \sin \iota.$$  

(43)

The Shapiro range and shape parameters $r_{Sh}$ and $s_{Sh}$ are two of the so-called post-Keplerian parameters one can measure for general-relativistic orbits. The Shapiro contribution to the timing sequence is particularly pronounced both in the double pulsar PSR J0737-3039A/B, with two neutron stars that are both detectable as pulsars, and the pulsar PSR J1614-2230 which orbits a White Dwarf star. For PSR 1913+16, on the other hand, the marked deviation from the shape parameter from 1 smoothes out the effect, although $r_{Sh}$ and $s_{Sh}$ are fitted in at least some of the timing models. The shape parameters for those three cases, plus the Keplerian parameters necessary for calculating the Shapiro time delay, are listed in Table 1. The resulting time-variable Shapiro time delay as a function of the orbital angle $\varphi$ is plotted in Fig. 3. Given the small eccentricities for both PSR J0737-3039A/B and PSR J1614-2230, the same plot also gives a reliable impression of how the time delay changes over time in those two systems.

VI. DISCUSSION

We have shown how to derive the Shapiro time delay near a mass from the equivalence principle, at the same level of simplification as that employed in some derivations of the gravitational deflection of light. The key elements of the derivation were two sets of coordinates: one adapted to Newtonian physics (and with a non-constant speed of light), the other local free-fall coordinates in which the equivalence principle could be applied. As in the corresponding Newtonian derivations of the deflection angle of light, the overall result is too small by a factor of 2, indicating that while time distortion effects have been taken into account, the curvature of space has been neglected. Nevertheless, the result has the same functional form, with its characteristic logarithm, as more precise calculation, and as such can serve to introduce the gravitational time delay in an undergraduate or even high school setting.

While a derivation of the Shapiro time delay from the Schwarzschild metric readily showed the origins of the missing factor 2, it also demonstrated ambiguities in metric-based textbook derivations of the Shapiro time delay (that are also present in the original research articles about the effect). Arguably, those ambiguities should be mentioned explicitly.
at the introductory textbook level when a metric-based derivation of the gravitational time delay is presented, even if a rigorous treatment (that is, the proper post-Newtonian formalism) is beyond the scope of such texts.

The time delay formula itself can readily be applied to observational data. In the solar system, fitting such data can be achieved by retrieving ephemeris data from suitable software. For solar system measurements involving Venus, the article includes a link to reconstructed observational data from Shapiro et al. 1970/1971. Fitting the data for a simple test of general relativity would be a suitable exercise for an undergraduate astronomy lab. For binary stars, the conventions for Keplerian and post-Keplerian orbital elements are introduced, enabling students to check for themselves some of the basic equations that are at the foundation of today’s most precise tests of general relativity involving compact binary stars.

ACKNOWLEDGEMENTS

I would like to thank Gerhard Schäfer for helpful discussions of the post-Newtonian formalism, and Thomas Müller for his valuable comments on an earlier version of this article.

---

1. Irwin I. Shapiro, “Fourth Test of General Relativity,” PRL 13, 789–791 (1964). DOI: 10.1103/PhysRevLett.13.789
2. Irwin I. Shapiro et al., “Fourth Test of General Relativity: Preliminary Results,” PRL 20, 1265–1269 (1968). DOI: 10.1103/PhysRevLett.20.12655
3. Irwin I. Shapiro et al., “Fourth Test of General Relativity: New Radar Result,” PRL 26, 1132–1135 (1971). DOI: 10.1103/PhysRevLett.26.1132
4. Clifford M. Will, “The Confrontation between General Relativity and Experiment,” Liv. Rev. Rel. 17, 4 (2014). DOI: 10.12942/lrr-2014-4
5. William Bruckman, W. and Ernesto P. Esteban, “An alternative calculation of light bending and time delay by a gravitational field,” Am. J. Phys. 61, 750–754 (1993). DOI: 10.1119/1.17151
6. Hans-Jürgen Treder and Gerhard Jackisch, “On Soldner’s Value of Newtonian Deflection of Light,” Astron. Nachr. 302, 275–277 (1981). DOI: 10.1002/asna.2010320603
7. Daniel S. Koltun, “Gravitational deflection of fast particles and of light,” Am. J. Phys. 50, 527–532 (1982). DOI: 10.1119/1.12802
8. Clifford M. Will, “Henry Cavendish, Johann von Soldner, and the deflection of light,” Am. J. Phys. 56, 413–415 (1988). DOI: 10.1119/1.15622
9. Karl-Heinz Lotze, “Wissenschaftsdidaktische Variationen über die Lichtablenkung am Sonnenrand,” Praxis der Naturwissenschaften – Physik in der Schule 4/54, 29–37 (2005)
10. Leonid Lerner, “A simple calculation of the deflection of light in a Schwarzschild gravitational field,” Am. J. Phys. 65, 1194–1196 (1997). DOI: 10.1119/1.18757
