New technology and an old theory: the Global Positioning System as a test of the ether wind hypothesis

Rainer Müller
Didaktik der Physik, Universität München, Schellingstr. 4, D-80799 München, Germany

Abstract:
In this article, we give a simple illustration how and why GPS (Global Positioning System) navigation would be affected by a possible ether wind. With a qualitative analysis of the navigation process we can estimated how big the discrepancy between the GPS receiver's inferred position and its true position would be. In addition, a one-dimensional quantitative model is discussed that leads to the same conclusion.

1. Introduction

In most courses on relativity, Einstein's theory is contrasted with its historical precursor, ether theory. Historically, the ether was introduced because prior to relativity, it was generally believed that each wave needed a medium to propagate. Consequently, such a medium was also postulated for the propagation of electromagnetic waves. It was called ether.

At the end of the 19th century, ether theory was widely accepted, in particular in Lorentz's version [1]. The ether was assumed to permeate all bodies and to be immobile, not being dragged by moving bodies. According to ether theory, the vacuum velocity of light equals c only in the rest frame of the ether. In any other frame of reference, the velocity of the frame is added to the velocity of light via the Galileian addition law. This is often visualized as "ether wind": the light wave seems to be "dragged" by the motion of the medium (the ether) relative to the source.

Because of the earth's orbital motion around the sun an ether theorist could not assume that the earth was at rest with respect to the ether. Because of the ether wind it should be possible to detect the motion of the earth by optical means. The velocity of light should be different in different directions on the earth's surface. A number of experiments were carried out to test this prediction. Their negative result was a major impulse for Einstein to develop his Special Theory of Relativity (cf. the second paragraph of his seminal paper [2]).

The most celebrated of these experiments was Michelson's and Morley's (1881/87). In retrospect it appears as the one that definitely ruled out ether theory. Even today it is impressing to see the great precision of the experiment and the enormous efforts that were necessary to realize it. Almost all courses on relativity refer to it to motivate Einstein's postulate of the constancy of the velocity of light.

The difficulty of the ether wind experiments prevents a classroom refutation of ether theory. Even a century of technological advance has not done much to facilitate this task. However, today we are able to test the ether wind hypothesis in a different way. In this article I will discuss a large-scale several-billion-dollar experiment to rule out ether theory at a 1%-5% level. Fortunately, most of the money has been spent for us already by the United States' DOD when the Global Positioning System (GPS) satellites were launched. All the remaining equipment that is needed is a $150 GPS receiver (which has the additional advantage that, unlike most other lab equipment, it can be used for recreational purposes too).

Certainly the idea that the GPS can be used to test the foundations of special relativity is not new [3] (the constructors of the system had to take time dilation into account and it has even been proposed that general relativistic effects could be measured using the GPS [4]). However

I think that it is worth to explain as simply as possible how and why the GPS can be employed as an ether wind experiment. In my experience, students in relativity courses find the connection between these two seemingly unrelated topics motivating and are interested to learn how the high-tech satellite navigation system works. In Section 3 I will give a simple qualitative argument how the position determination would be affected by the ether wind and how big the discrepancy between the receiver's inferred position and its true position would be. In Section 4, a very simple quantitative model is presented to illustrate the argument. But first it is necessary to explain the functioning of the GPS in some detail [5].

2. The Global Positioning System

The Global Positioning System was designed as a navigation system for military purposes and is now widely used also for civilian navigation. It consists of 24 satellites which orbit around the earth at 20 000 km altitude. Each satellite has an atomic clock and is sending radio signals which is used to infer one's position.

Let us assume for the moment that the receiver's clock is synchronized with the satellites' clocks very well (better than a few parts of a microsecond). The receiver position is then determined as follows: Each satellite sends a signal which contains the information about its momentary position and the exact time at which the signal was sent.

The satellite signal propagates with the velocity c and reaches the receiver after some time t_1. The receiver compares the time information in the signal with its local clock and is thus able to determine the time of travel t_1. The distance to the satellite is then $s_1 = c\, t_1$. This restricts the possible receiver positions to a sphere of radius s_1 centered around the satellite (Fig. 1).

The measurement is the repeated with a second satellite which gives a second sphere with radius $s_2 = c\, t_2$. The receiver is now known to be located at the intersection of the two spheres, which is a circle. A third measurement gives another sphere and the exact position: It is the location where all three spheres intersect (Fig. 2). The position determination seems to be complete.

However, we forgot something: In order to determine the travel time of the signals, the receiver's clock must be accurately synchronized with the satellites' clocks. Remember that a deviation of 1 μs means an error of up to 300 m in the position determination. It is not possible to maintain such a precise synchronization over a longer period of time with the receiver's quartz clock. A trick is used instead: four time-of-travel measurements (instead of three) are made with the imprecise receiver clock. The resulting four spheres will not intersect at a single point. The receiver time is then varied until all spheres intersect (Fig. 3). Now the position as well as the exact time can be read off from the receiver.

Mathematically, the four time-of-travel measurements correspond to a set of four equations for the three spatial coordinates and for τ, the deviation of the receiver's time from the satellites' time. In practice, more than four satellites can be used if visible (standard receivers have 12 channels). The additional information is used to compensate for other sources of error and bad satellite geometry (satellites directly above you don't help much to determine your lateral position). The claimed accuracy of the GPS position determination is 100 m. If the satellite signals were not intentionally deteriorated in the selective availability mode the accuracy would be about 10 m.

It is worth noting that, from a relativistic point of view, GPS navigation involves a (quite complicated) way of clock synchronization – a key notion in relativity.

3. Satellite navigation and ether wind

How can satellite navigation help to test the ether wind hypothesis? Consider a signal (say "emission time = 12 o'clock") that emanates as a pulse of radio waves from a satellite. According to ether theory, the ether wind velocity $\vec{v}$ is added to the velocity of light. If the pulse is assumed to expand spherically it propagates according to the equation (the origin is at the satellite's position)

$$\begin{aligned}\vec{x}_{\text{ether}}(t) &= (c \cdot \hat{\text{u}}_{\text{r}} + \vec{v})t \\ &= \vec{x}_{\text{non-ether}}(t) + \vec{v}t.\end{aligned} \tag{1}$$

Compare this with Fig. 1. The surface at which the signal is after a time t_1 is a sphere whose origin is shifted by $\vec{v}t_1$ (Fig. 4). A receiver registering a time-of-travel t_1 will infer falsely that it is located somewhere on the surface of the shifted sphere.

The same is true for the remaining time-of-travel measurements. For the sake of simplicity, we assume synchronized clocks, i. e. only three measurements are necessary (cf. Fig. 2). The corresponding spheres is shifted by $\vec{v}t_2$ and $\vec{v}t_3$, respectively, relative to the original ones. Now, terrestrial distances are fairly small compared with the satellite distance of 20 000 km. Therefore, the travel times are nearly equal: $t_1 \approx t_2 \approx t_3 \equiv t$. That means the whole pattern of spheres is shifted uniformly by $\vec{v}t$ (Fig. 5). The result is clear: If ether wind was real a real phenomenon, a receiver would infer a position that is shifted by $\vec{v}t$ with respect to its true position. In this way, the ether wind's effect on the propagation of electromagnetic waves would be detectable.

Let us consider the magnitude of the effect. The earth orbits around the sun with 30 km/s. The time-of-travel of a satellite signal is t ≈ 20 000 km / (300 000 km/s) = 0.067 s. Therefore, the position inferred from the satellite determination would deviate from our actual position by $v\,t$ = 2 km. But in reality, a GPS measurement reveals the true position with a much better precision (at least 100m = 5%). Thus, the mere fact that the GPS works shows that there is something wrong with the ether wind hypothesis. Each day, thousands of GPS measurements disprove ether theory.

One important remark must be made: In order for the argument not to be void, the map with which we determine our actual position must have been produced without the help of the GPS. Fortunately, generations of land surveyors have produced accurate maps with conventional methods.

4. A simple quantitative model

In order to illustrate the above considerations more quantitatively let us investigate a simple one-dimensional model of GPS navigation. Two satellites A and B are located to the left and to the right from the center of the earth in a distance of 20 000 km (Fig. 6). Their positions are denoted x_A and x_B. Simultaneously, they send a signal that is received by the receiver after a time $c\,t_1$ and $c\,t_2$, respectively. The receiver can determine its unknown position x_0 with the following equations:

$$\begin{aligned}x_0 &= x_A + c\,t_1, \quad \text{(signal from satellite A)} \\ x_0 &= x_B - c\,t_2. \quad \text{(signal from satellite B)}\end{aligned} \tag{2}$$

Adding these two equations gives:

$$x_0 = \tfrac{1}{2}\left(x_A + x_B\right) + \tfrac{1}{2} c\,\Delta t. \tag{3}$$

The position of the receiver can thus be inferred from the known satellite positions x_A and x_B and the measured difference $\Delta t = t_1 - t_2$ of the arrival times of the two signals.

How does this result change if the effect of the ether wind is taken into account? We assume that the ether wind comes from the left so that the signal of satellite A propagates with the velocity $c + \upsilon$ while the one of satellite B propagates with $c - \upsilon$ (Fig. 7). We denote the position inferred with ether wind by x_0'. The propagation equations now take the form

$$x_0 = x_A + (c + \upsilon)\, t_1,$$
$$x_0 = x_B - (c - \upsilon)\, t_2.$$

adding the equations leads to

$$\begin{aligned} x_0' &= \tfrac{1}{2}\left(x_A + x_B\right) + \tfrac{1}{2} c\,\Delta t + \tfrac{1}{2}\upsilon\left(t_1 + t_2\right) \\ &= x_0 + \tfrac{1}{2}\upsilon\left(t_1 + t_2\right). \end{aligned} \tag{4}$$

In the last line, (3) has been used. If, as in the previous Section, we approximate $t_1 \approx t_2$ in view of the magnitudes of the numbers involved, we obtain the same result: The two position determinations x_0 and x_0' differ by an amount $\upsilon\, t$ = 2 km. The model thus confirms the qualitative considerations of the last Section.

We can gain a little more insight in the nature of the above approximation by subtracting the two equations (4). This gives

$$0 = x_A - x_B + c\left(t_1 + t_2\right) + \upsilon\left(t_1 - t_2\right). \tag{5}$$

Now we always have $|t_1 - t_2| < t_1 + t_2$, and, in addition, $\upsilon \ll$ c. It is therefore safe to neglect the last term in (5) so that the equation becomes $t_1 + t_2 \approx (x_B - x_A)/c$. If we insert this into (4) we obtain for the difference between inferred and true position

$$x_0' - x_0 = \frac{\upsilon}{c} \cdot \frac{1}{2}\left(x_B - x_A\right). \tag{6}$$

This equation shows that the GPS position mismatch predicted by the ether theory is a first-order effect in υ/c [6]. The effect (or, in fact, its absence) can be detected relatively easy because of the large scale of the experiment: the distance $(x_B - x_A)$ is 40 000 km.

Acknowledgment:
I am grateful to Jakob Reichel for the simplified model of GPS navigation.

[1] For a historical review see: J. Stachel, *History of Relativity*, in: L. M. Brown, A. Pais, B. Pippard (eds.) *Twentieth Century Physics*, IOP/AIP (Bristol, Philadelphia, New York) 1995; E. T. Whittaker, *A History of the Theories of Aether and Electricity*, Harper Torchbooks, New York, 1960.
[2] A. Einstein, *Zur Elektrodynamik bewegter Körper*, Ann. d. Phys. **17**, 891-921 (1905).
[3] The same idea is inherent, for example, in the discussion of the Universal Time Coordinated (UTC) and the LORAN-C network in R. Sexl, H. K. Schmidt: *Raum, Zeit, Relativität*, Vieweg, Braunschweig, 1978.
[4] N. Ashby, in *Matters of Gravity 9* (1997), available on the Web from preprints.cern.ch/archive/electronic/gr-qc/9702/9702010.pdf
[5] There are a number of very good resources on how GPS works can be found on the WWW. See, e. g.: www.trimble.com/gps/index.htm.
[6] The Michelson-Morley experiment is second-order in υ/c, and this had important consequences. Lorentz' original theory was able to explain the negative result of first-order

experiments such as the one considered here. However, he had to introduce complicated additional assumptions such as introducing a “local time” variable (which was not interpreted realistically) into the solutions of Maxwell’s equation. In its original form, Lorentz’ theory predicted a non-null result for second-order experiments (the observed null results could be accounted for, however, when he introduced the contraction hypothesis). I think in a classroom setting it is sufficient to disprove the simplest of all ether theories.

Figure captions:

Figure 1: The surface of the sphere gives the range of possible receiver positions after the travel time measurement.

Figure 2: The receiver is located at the intersection of the three spheres. For easier visualization, this and the subsequent figures are drawn in two dimensions.

Figure 3: Position determination (a) with improperly synchronized clocks, (b) with properly synchronized clocks.

Figure 4: Under the influence of the ether wind, a spherical pulse is shifted by $\vec{v} \cdot t$.

Figure 5: Position determination without (gray) and with (black) ether wind. All spheres are shifted by $\vec{v} \cdot t$.

Figure 6: Geometry of the simple model.

Figure 7: Position determination with ether wind.

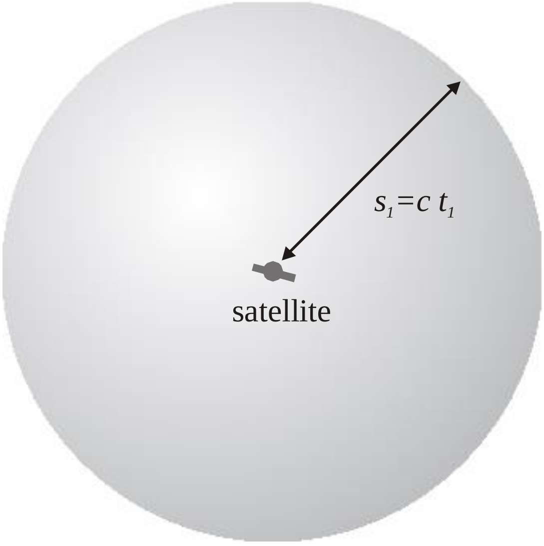

Figure 1
Müller

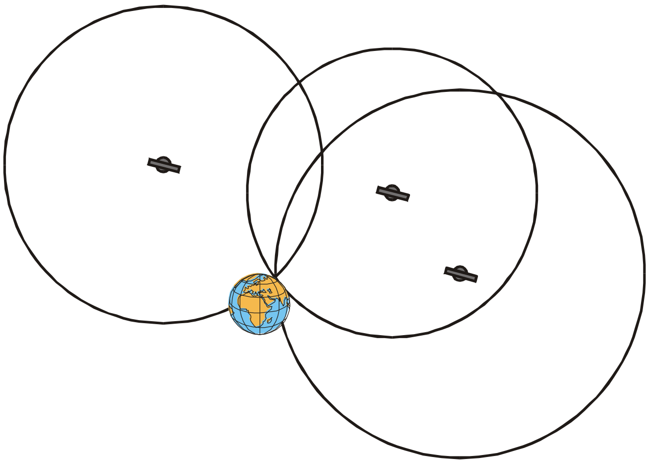

Figure 2
Müller

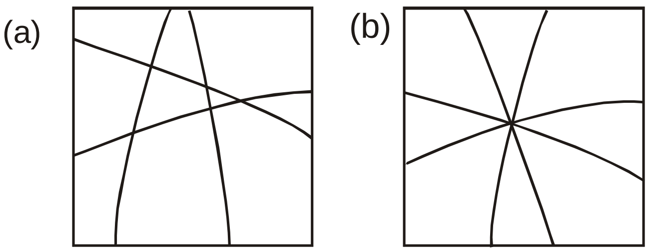

Figure 3
Müller

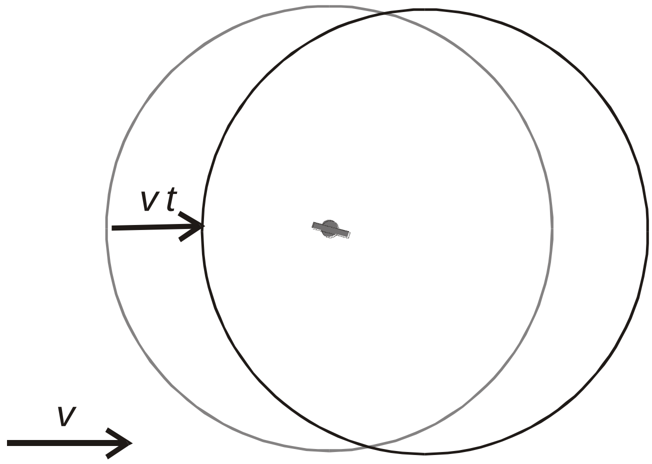

Figure 4
Müller

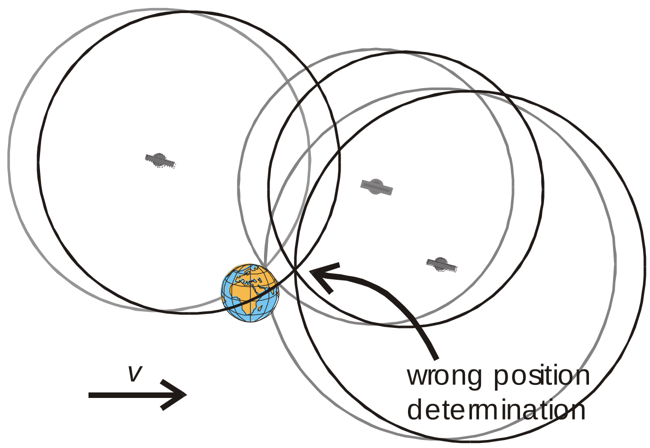

Figure 5
Müller

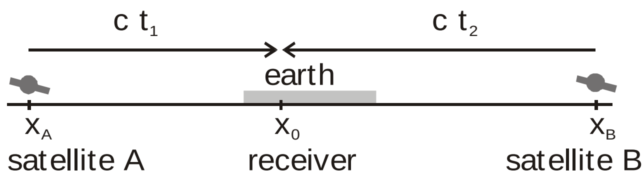

Figure 6
Müller

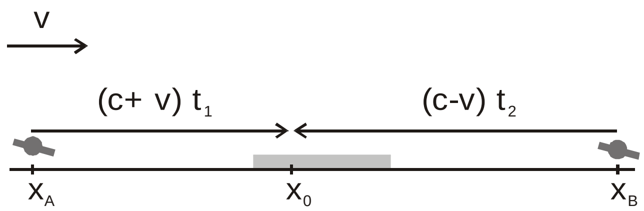

Figure 7
Müller