GPS RECEIVERS AND RELATIVITY*

Marc Weiss
National Institute of Standards and Technology
Time and Frequency Division
325 Broadway, Boulder, CO 80303

Neil Ashby
Department of Physics, Campus Box 390
University of Colorado
Boulder, CO 80309-0390

Abstract

We illustrate the general methods for applying relativistic corrections needed by a GPS receiver in providing time or position to a user. We focus on estimating the time interval it takes for GPS signals to propagate from the transmitter to the receiver, the geometric range delay. We present a few cases which apply to many common uses of GPS. The most common case for positioning is illustrated numerically.

INTRODUCTION

A GPS receiver needs to make two corrections that are related to relativity in order to provide time or position to a user. We discuss these corrections and focus mostly on estimating the geometric range delay, Δt_D, the time for GPS signals to propagate from the transmitter to the receiver. Proper estimation of Δt_D is essential for solving for position or time. This is an application of the relativistic principle of the constancy of the velocity of light which states that electromagnetic signals travel in Euclidean straight lines with velocity c relative to an inertial reference frame. We present a few cases which apply to many common uses of GPS. The case where measurements of satellite signals are time-tagged at the receiver for positioning, probably the most common GPS application, is illustrated numerically.

The theory behind corrections is presented with references given for any derivations not done here. Through our theoretical discussion we show that the Interface Control Document (ICD-GPS-200) specifications, as issued by the Joint Program Office of the Global Positioning System [1], consistently cover the requirements of relativity down to the sub-nanosecond level for time. We respond to questions in the literature [2,3] as to whether the ICD specifications include relativity corrections with enough accuracy for certain applications. In particular we discuss the relativistic Doppler effect, the formula for its instantaneous magnitude, and its relationship with typical GPS receiver operation. We also address the use of carrier-phase measurements, which is not discussed in the ICD.

THEORY

Generally, a GPS navigation user measures the arrival times, on a local clock, of timing signals from at least four different satellites, then solves for four unknowns: user position x,y,z and the receiver clock offset from GPS time t. The signals from the satellites can be thought of effectively as continuous timing signals from the satellite clocks arriving at the receiver. The receiver has its own local clock for comparison. The user measures either 1) the times in the received timing signals at a specific local clock time, or 2) the arrival times on the local clock of a specific time-tag in the received timing signals. The reception time (according to the local clock) minus the transmission time at one satellite (according to GPS time) is called the *pseudorange*. These pseudoranges are used to solve for user position and time.

For the ordinary user of broadcast ephemerides, there are two and only two relativistic effects that must be considered. First, the receiver must apply a correction to the transmitted time to account for relativistic effects arising from orbit eccentricity of the transmitting satellite. This is the Δt_r term defined in the ICD. Second, the finite and universally constant speed c of signals propagating from transmitter to receiver, relative to an inertial frame (the geometric path delay), must be accounted for.

RELEVANT RELATIVITY

Three effects in relativity are germane to GPS. Rates of clocks in GPS are adjusted (as for International Atomic Time) to match the rate that clocks would run on the geoid of the earth. This is a surface of gravitational equipotential in the rotating frame in which the effects 2) and 3) below add to a constant value. The relativity effects are as follows.

1) GPS time is defined using the principle of the constancy of c to synchronize an imagined system of clocks everywhere in space in the neighborhood of the earth (Einstein synchronization). GPS satellite clocks are in principle adjusted to agree with this imagined system of clocks. This network of synchronized GPS clocks realize a coordinate time. This definition of GPS time requires a locally inertial coordinate system. GPS time is thus defined relative to an earth-centered inertial coordinate system (an ECI), but the rate is set to match the rate at which clocks would run on the geoid. An ECI is also used to simplify the paths of signals propagating from satellites, since they move in Euclidean straight lines at the velocity c in vacuum relative to such inertial frames.

2) A clock moving with respect to an ECI runs slower relative to coordinate time than if it were at rest in the ECI. This is the time dilation effect due to the magnitude of the relative velocity, sometimes called the second-order Doppler effect. For satellites in GPS orbits, the fractional frequency offset needed to compensate for this is approximately $+8.3\cdot10^{-11}$ relative to the rate of clocks on the earth's geoid.

3) A clock in a lower gravitational potential runs slower relative to coordinate time than if it were at rest in a higher potential. This is called the gravitational red shift. Thus, standard clocks closer to the earth run slower than standard clocks farther away, since the potential becomes more negative closer to the earth. Clocks on GPS satellites need to be adjusted by about $-5.3\cdot10^{-10}$ relative to the earth's geoid, to compensate for this effect.

Atomic clocks in GPS satellites are given a fixed rate offset of $-4.4645\cdot10^{-10}$ as a consequence of the requirement that GPS satellite clocks run at the rate that a standard clock on the geoid would run, and of the relativistic effects in 2) and 3) for circular orbits. These three relativistic effects explain the reasons for the