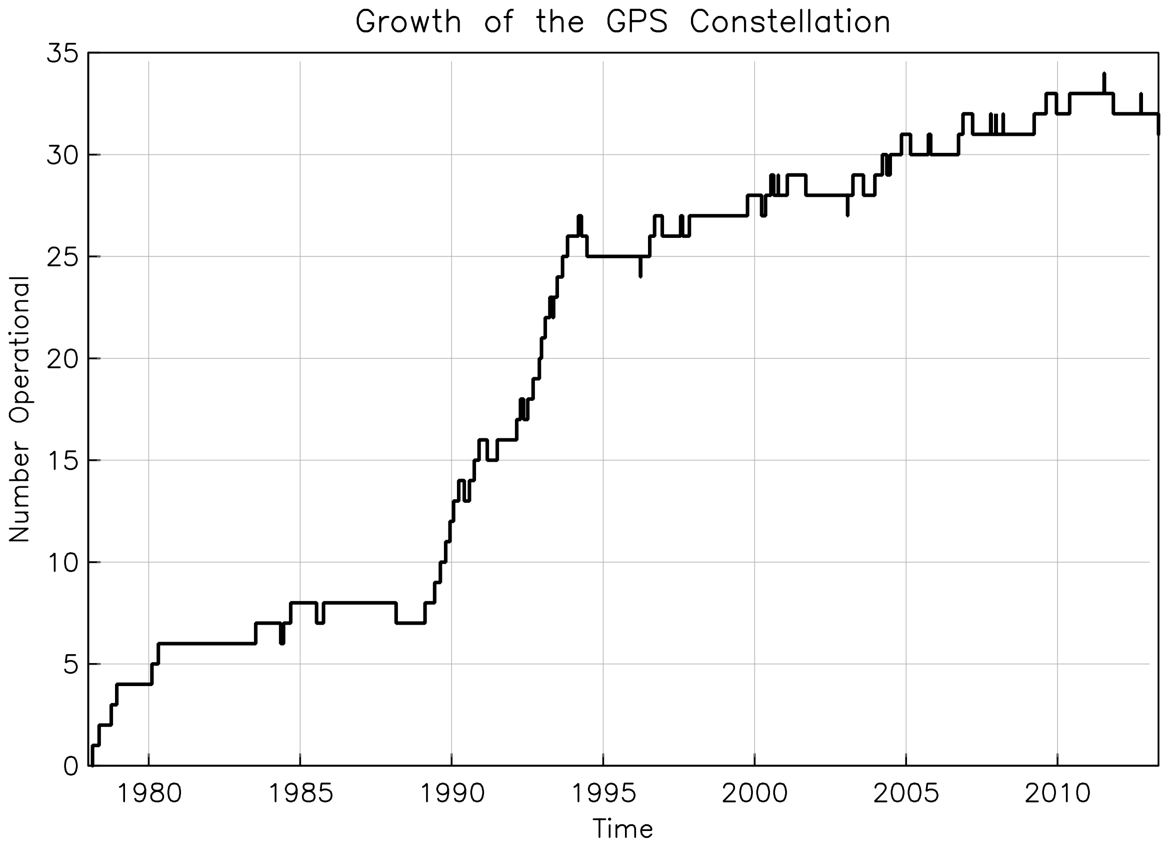
GPS: History, Operation, Processing

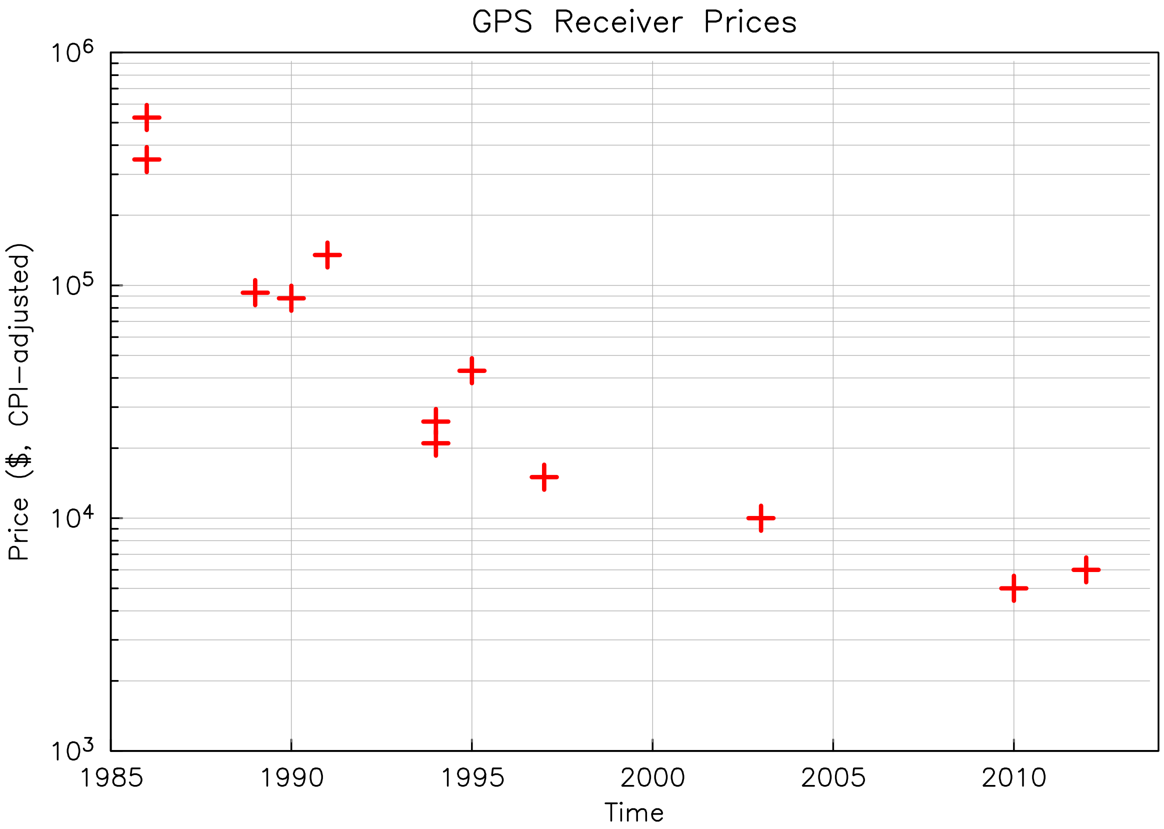
Important Dates

- **1970’s:** conceived as radionavigation system for the US military: real-time locations with few-meter accuracy.
- **1978:** first satellite launched
- **1983:** Declared to be a “dual-use” system, civilian and military.
- **Early 1980’s:** Academics realized it could be used for geodesy.
- **1986:** Initial measurements by academic institutions; start of UNAVCO.
- **Mid-1990’s:** Many more satellites; receiver costs drop substantially.
- **1994:** System becomes “operational”; International GPS Service produces high-quality orbits.
- **2000:** Selective Availability turned off – much better civil navigation.

GPS History: Celestial



GPS History: Mundane



GPS Operations: Orbits

- Near-circular orbits with 55° inclination, altitude of about 20,200 km, period of 43077.2 s.

GPS Operations: Orbits

- Near-circular orbits with 55° inclination, altitude of about 20,200 km, period of 43077.2 s.
- Period means that every two revolutions the satellite is over the same place: the **ground track** repeats exactly in space, with a period of 86154.4 s, or 4^{m} 5.6^{s} less than a day.

GPS Operations: Orbits

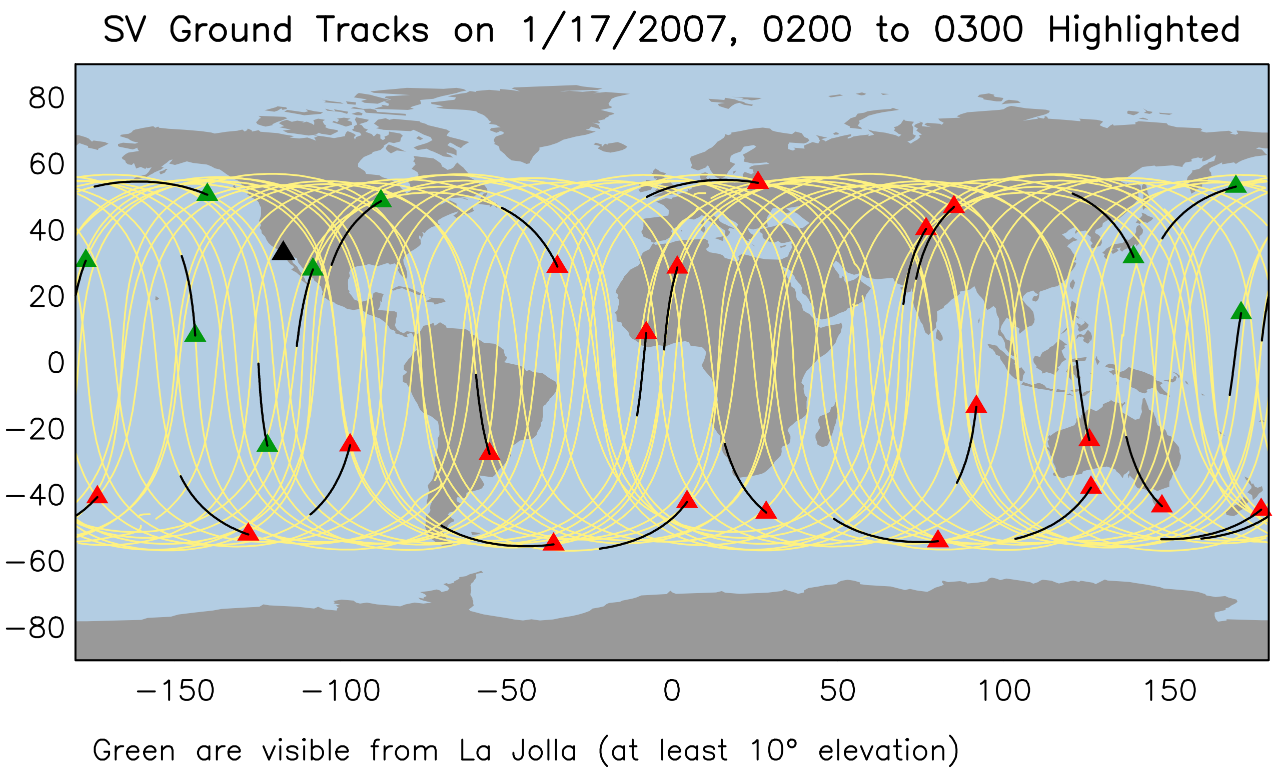
- Near-circular orbits with 55° inclination, altitude of about 20,200 km, period of 43077.2 s.
- Period means that every two revolutions the satellite is over the same place: the **ground track** repeats exactly in space, with a period of 86154.4 s, or 4^{m} 5.6^{s} less than a day.
- Satellites move in "inertial space"; largest non-gravitational acceleration is 10^{-8} g, giving a nearly perfect reference frame, for times of a few days or less.

GPS Operations: Orbits

- Near-circular orbits with 55° inclination, altitude of about 20,200 km, period of 43077.2 s.
- Period means that every two revolutions the satellite is over the same place: the **ground track** repeats exactly in space, with a period of 86154.4 s, or 4^{m} 5.6^{s} less than a day.
- Satellites move in “inertial space”; largest non-gravitational acceleration is 10^{-8} g, giving a nearly perfect reference frame, for times of a few days or less.

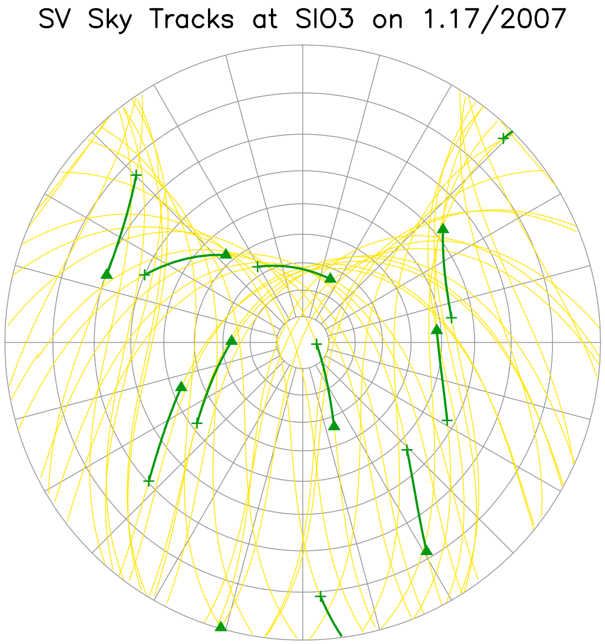
The processing packages all retain orbit determination, but we can usually assume that someone else (eg SOPAC) has found a precise orbit.

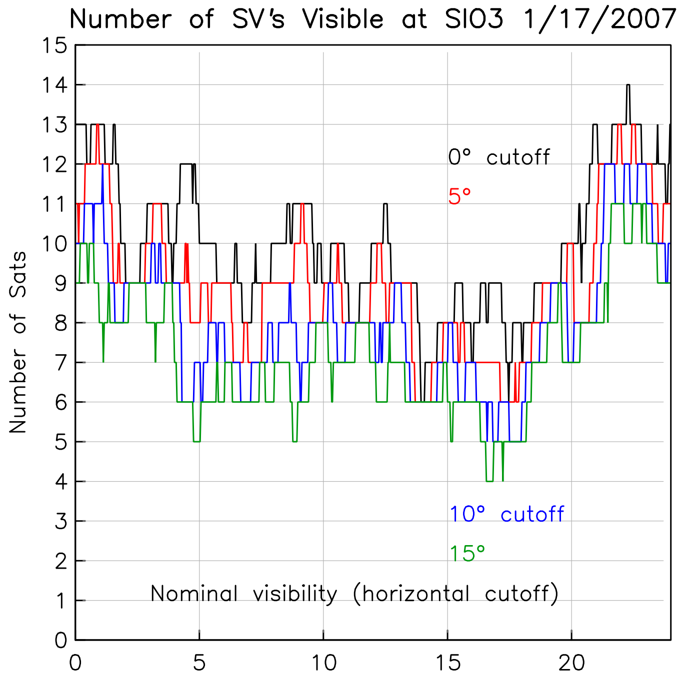
Where the Satellites Are (Looking Down)



Black lines show the movement over an hour.

Where the Satellites Are (Looking Up)





Viewed from near the Scripps Aquarium. Yellow is sky tracks over a day (notice the hole to the N), green shows the them over an hour. The number of satellites visible depends on the **elevation cutoff**. Fewer satellites means worse estimates.

What do the Satellites Transmit? (I)

All the radio signals are "L-band": frequencies about 1.5 GHz, wavelengths about 0.2 m. The two frequencies used for positioning are:

- **L1**: Frequency 1575.42 MHz, wavelength 190.3 mm.

What do the Satellites Transmit? (I)

All the radio signals are “L-band”: frequencies about 1.5 GHz, wavelengths about 0.2 m. The two frequencies used for positioning are:

- **L1**: Frequency 1575.42 MHz, wavelength 190.3 mm.
- **L2**: Frequency 1227.60 MHz, wavelength 244.2 mm.

What do the Satellites Transmit? (I)

All the radio signals are “L-band”: frequencies about 1.5 GHz, wavelengths about 0.2 m. The two frequencies used for positioning are:

- **L1**: Frequency 1575.42 MHz, wavelength 190.3 mm.
- **L2**: Frequency 1227.60 MHz, wavelength 244.2 mm.

Other frequencies are used for other purposes, and some have been added on the newest systems.

What do the Satellites Transmit? (II)

On the L1 and L2 frequencies there are:

- A **carrier**, the main frequency, which is modulated in various ways to provide:

What do the Satellites Transmit? (II)

On the L1 and L2 frequencies there are:

- A **carrier**, the main frequency, which is modulated in various ways to provide:
- Positioning Codes:

What do the Satellites Transmit? (II)

On the L1 and L2 frequencies there are:

- A **carrier**, the main frequency, which is modulated in various ways to provide:
- Positioning Codes:
    - **C/A code**, on the L1 frequency only. This has a “repeat wavelength” of about 300 m, and can be decoded by anyone: this is what all civilian navigation systems use.

What do the Satellites Transmit? (II)

On the L1 and L2 frequencies there are:

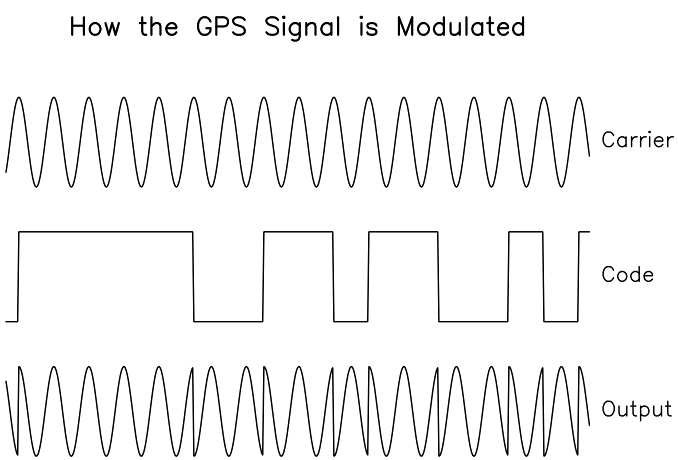
- A **carrier**, the main frequency, which is modulated in various ways to provide:
- Positioning Codes:
  - **C/A code**, on the L1 frequency only. This has a "repeat wavelength" of about 300 m, and can be decoded by anyone: this is what all civilian navigation systems use.
  - **P code**, on both the L1 and L2 frequencies. This has a "repeat wavelength" of about 30 m – but to use it unambiguously, you need a DoD decoder. But it can be used for geodesy without the decoder.

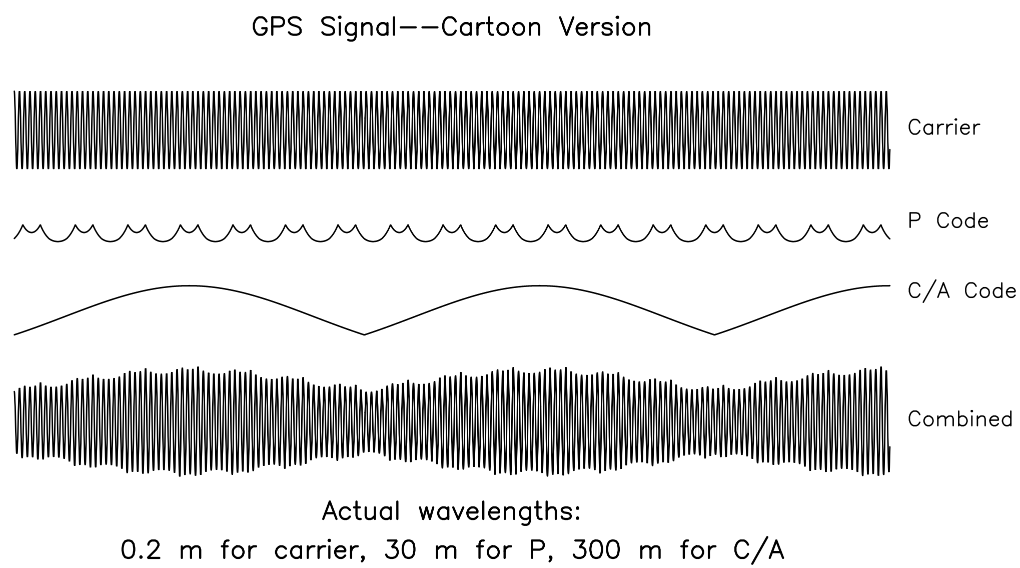
What do the Satellites Transmit? (II)

On the L1 and L2 frequencies there are:

- A **carrier**, the main frequency, which is modulated in various ways to provide:
- Positioning Codes:
  - **C/A code**, on the L1 frequency only. This has a “repeat wavelength” of about 300 m, and can be decoded by anyone: this is what all civilian navigation systems use.
  - **P code**, on both the L1 and L2 frequencies. This has a “repeat wavelength” of about 30 m – but to use it unambiguously, you need a DoD decoder. But it can be used for geodesy without the decoder.
- Satellite position information, timing information, and lots more.

Carrier Modulation





Actual modulation is done (as on left) by changing the phase of the carrier.
Right-hand plot is a cartoon of how two codes amplitude modulate carrier.

Signal Wavelengths

Fundamental limit on precision is that we can only measure to within some fraction of λ, where λ is the wavelength.

Signal Wavelengths

Fundamental limit on precision is that we can only measure to within some fraction of λ, where λ is the wavelength.

λ is 0.2 m for carrier, 30 m for P, 300 m for C/A; we can measure to about 0.01 of the wavelength, so 2 mm using the carrier, 3 m using C/A.

So for geodesy we use the carrier – ideally, after demodulating, which requires that we know the code.

Code is also useful for getting approximate positions, as a first step.

Basic Geodetic Observable: Carrier-Beat Phase

Nominally, geodetic observable is the phase of the carrier over time. If at some time this was zero, and later one is 90°, the distance has changed by $\lambda/4$.

Basic Geodetic Observable: Carrier-Beat Phase

Nominally, geodetic observable is the phase of the carrier over time. If at some time this was zero, and later one is 90°, the distance has changed by $\lambda/4$.

We actually use the **carrier-beat phase**: the phase of $\exp(2\pi i[{}^{j}f - {}_{k}f]t)$ which is the difference between the carrier frequency from satellite j and the frequency of an oscillator in receiver k. This beat frequency is at much less than 1.5 GHz.

So this involves on two frequencies:

- ${}^{j}f$ depends on the satellite clock and the velocity of the satellite relative to the receiver (Doppler shift).
- ${}_{k}f$ depends on the receiver clock.

Note that we use superscripts for satellites, subscripts for receivers.

GPS for Navigation – Oversimplified

We start with a simplified version, which is closer to how a receiver finds its position.

- The j-th satellite sends a message with the time, ${}^{j}t$, at which the message was sent.

GPS for Navigation – Oversimplified

We start with a simplified version, which is closer to how a receiver finds its position.

- The j-th satellite sends a message with the time, ${}^{j}t$, at which the message was sent.
- The k-th receiver gets this message at time ${}_{k}^{j}t$.

If these times are not in error, the distance (or "range") between the receiver and satellite is then: ${}_{k}^{j}d = c({}_{k}^{j}t - {}^{j}t)$ where c is the speed of light.

GPS for Navigation – Oversimplified

We start with a simplified version, which is closer to how a receiver finds its position.

- The j-th satellite sends a message with the time, ${}^{j}t$, at which the message was sent.
- The k-th receiver gets this message at time ${}^{j}_{k}t$.

If these times are not in error, the distance (or "range") between the receiver and satellite is then: ${}^{j}_{k}d = c({}^{j}_{k}t - {}^{j}t)$ where c is the speed of light.

But this is actually called the **pseudorange**; "pseudo-" because none of the terms in the equation are exactly right:

GPS for Navigation – Oversimplified

We start with a simplified version, which is closer to how a receiver finds its position.

- The j-th satellite sends a message with the time, ${}^{j}t$, at which the message was sent.
- The k-th receiver gets this message at time ${}^{j}_{k}t$.

If these times are not in error, the distance (or "range") between the receiver and satellite is then: ${}^{j}_{k}d = c({}^{j}_{k}t - {}^{j}t)$ where c is the speed of light.

But this is actually called the **pseudorange**; "pseudo-" because none of the terms in the equation are exactly right:

- The satellite clock has error ${}^{j}\varepsilon$, so the time ${}^{j}t$ is wrong.

GPS for Navigation – Oversimplified

We start with a simplified version, which is closer to how a receiver finds its position.

- The j-th satellite sends a message with the time, ${}^{j}t$, at which the message was sent.
- The k-th receiver gets this message at time ${}^{j}_{k}t$.

If these times are not in error, the distance (or “range”) between the receiver and satellite is then: ${}^{j}_{k}d = c({}^{j}_{k}t - {}^{j}t)$ where c is the speed of light.

But this is actually called the **pseudorange**; “pseudo-” because none of the terms in the equation are exactly right:

- The satellite clock has error ${}^{j}\varepsilon$, so the time ${}^{j}t$ is wrong.
- The receiver clock has error ${}_{k}\varepsilon$, so the time ${}^{j}_{k}t$ is wrong (by much more).

GPS for Navigation – Oversimplified

We start with a simplified version, which is closer to how a receiver finds its position.

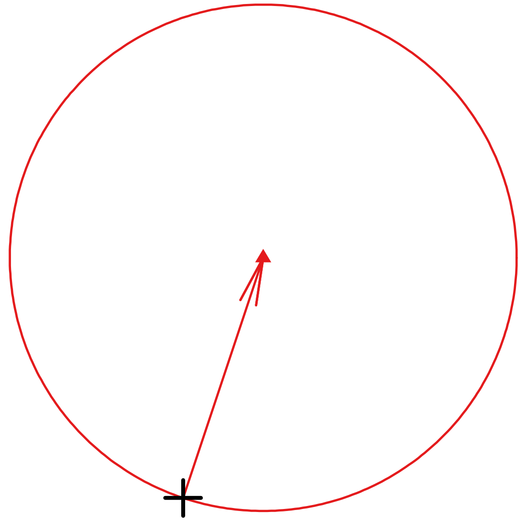
- The j-th satellite sends a message with the time, ${}^{j}t$, at which the message was sent.
- The k-th receiver gets this message at time ${}^{j}_{k}t$.

If these times are not in error, the distance (or "range") between the receiver and satellite is then: ${}^{j}_{k}d = c({}^{j}_{k}t - {}^{j}t)$ where c is the speed of light.

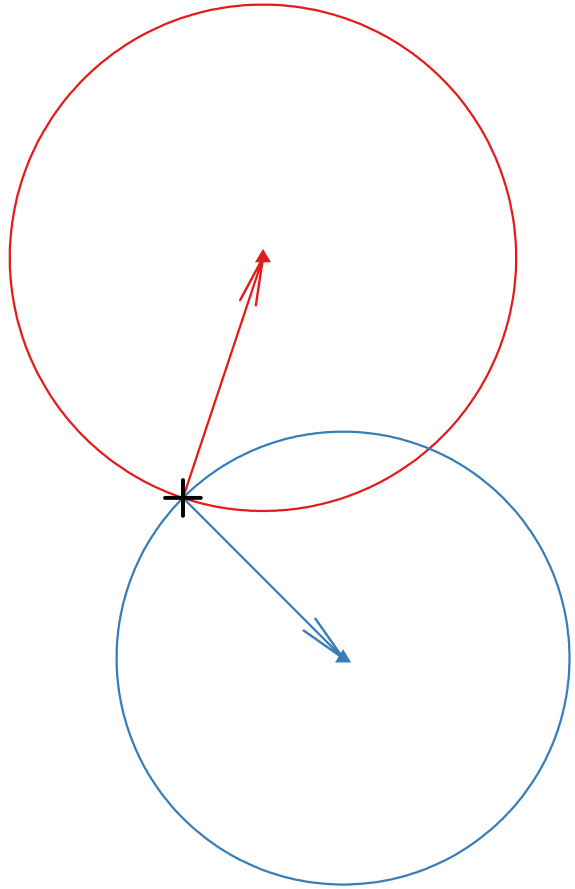
But this is actually called the **pseudorange**; "pseudo-" because none of the terms in the equation are exactly right:

- The satellite clock has error ${}^{j}\varepsilon$, so the time ${}^{j}t$ is wrong.
- The receiver clock has error ${}_{k}\varepsilon$, so the time ${}^{j}_{k}t$ is wrong (by much more).
- The radio waves not travel at c in the troposphere and ionosphere.

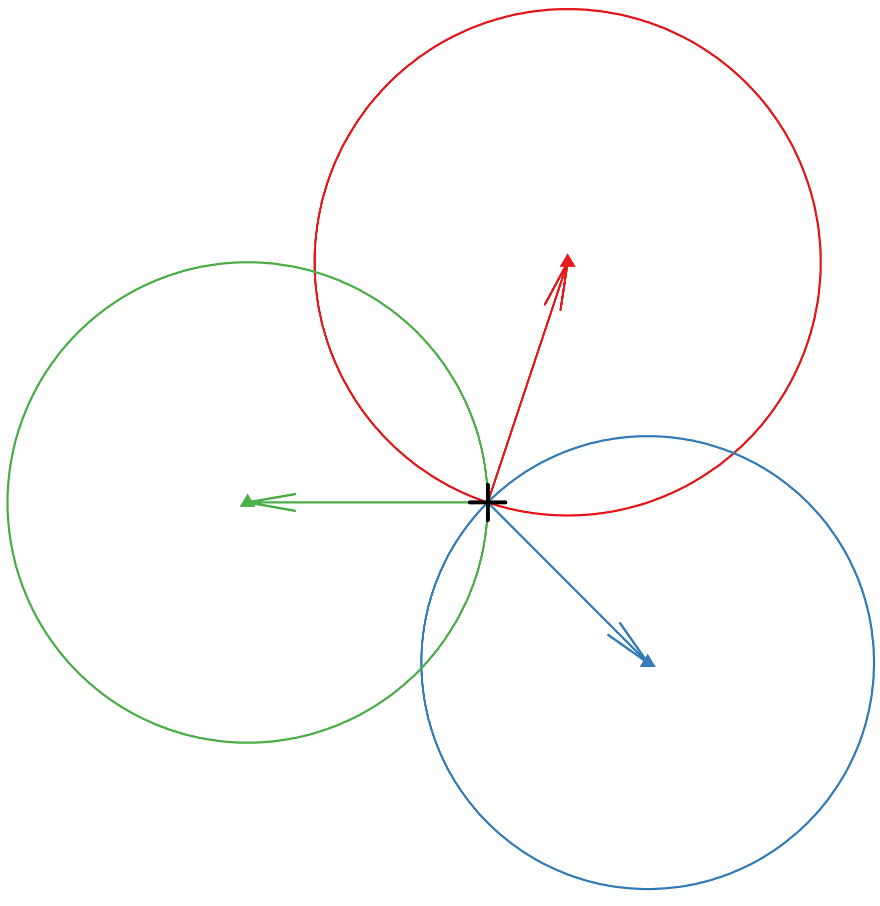
Locating a Single Source I



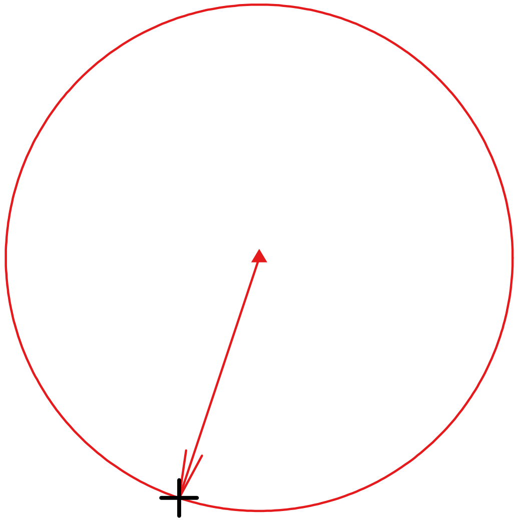
Locating a Single Source II



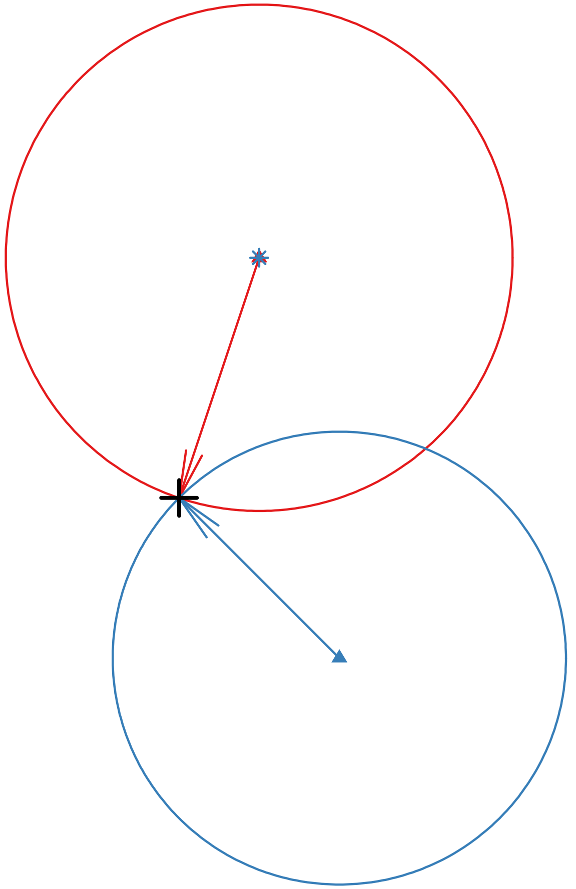
Locating a Single Source III



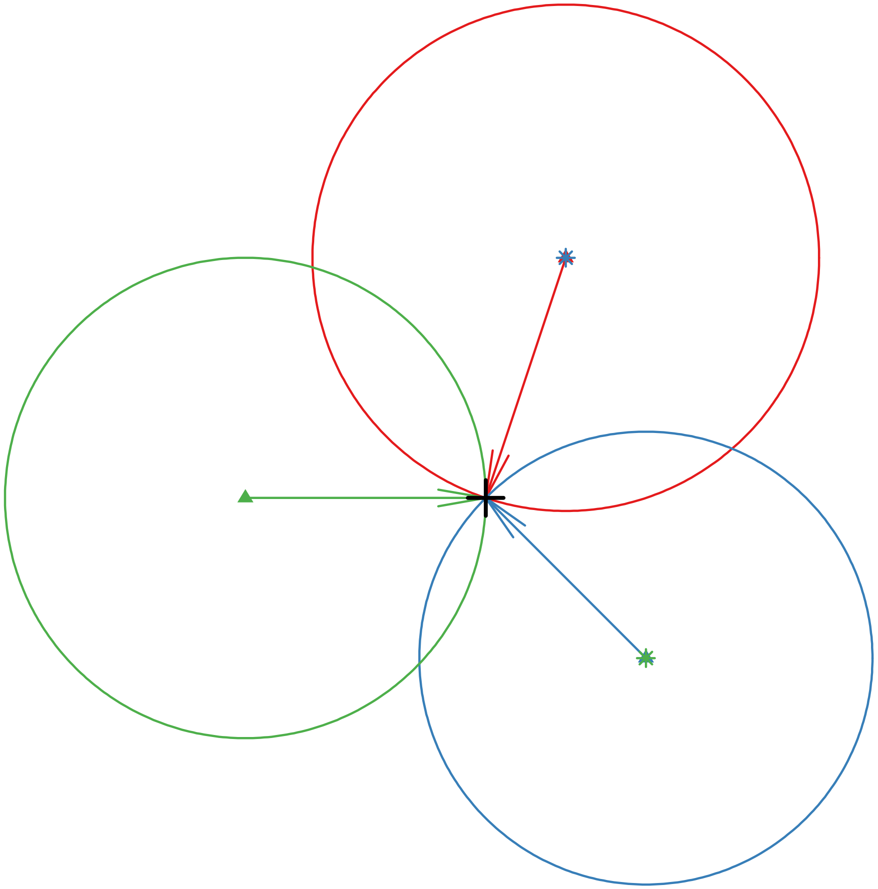
Locating a Single Receiver I



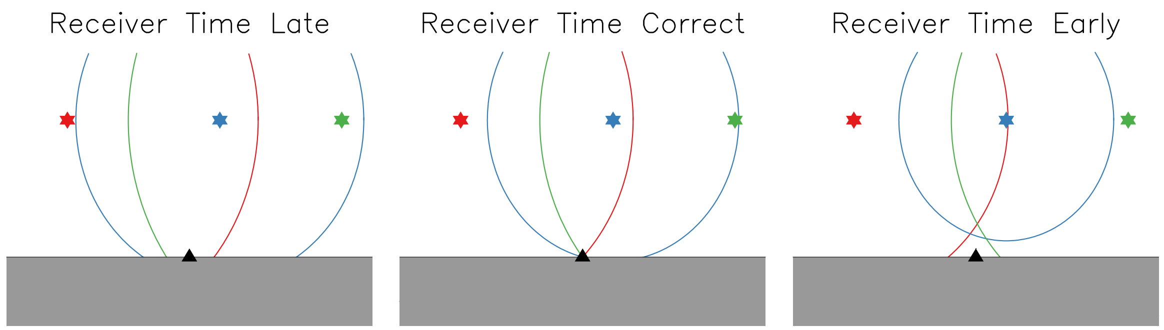
Locating a Single Receiver II



Locating a Single Receiver III



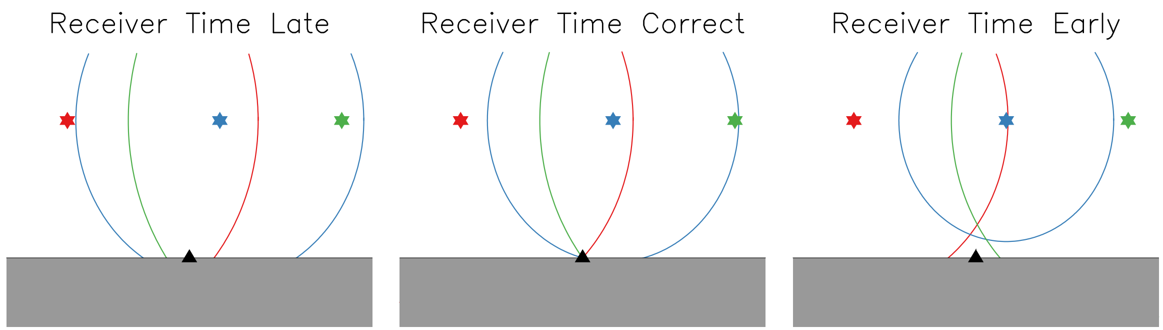
Solving the Clock Problem



We show a GPS-like system, but in a flat and 2-D world, with three satellites.

Given two pseudoranges, we have to be at one of two points where the circles intersect (since we know where the satellite are).

Solving the Clock Problem

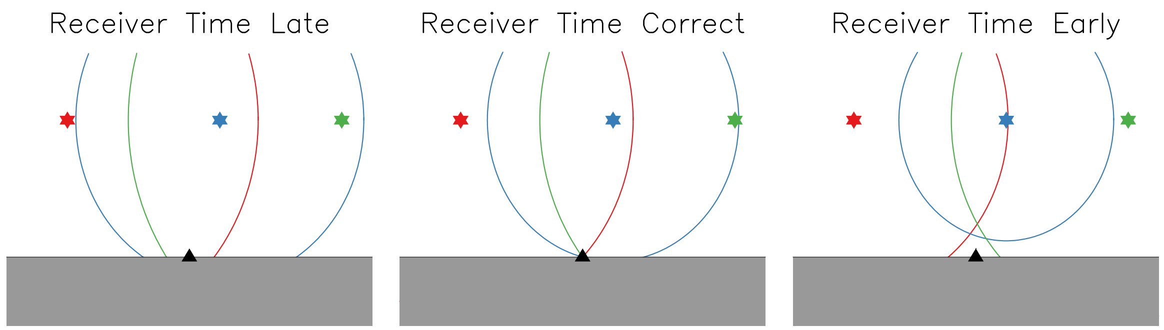


We show a GPS-like system, but in a flat and 2-D world, with three satellites.

Given two pseudoranges, we have to be at one of two points where the circles intersect (since we know where the satellite are).

If the receiver time is off, we get the wrong answer.

Solving the Clock Problem



We show a GPS-like system, but in a flat and 2-D world, with three satellites.

Given two pseudoranges, we have to be at one of two points where the circles intersect (since we know where the satellite are).

If the receiver time is off, we get the wrong answer.

But with *three* satellites, the wrong time will not fit any position – so we adjust the receiver time until it does. We then know both where we are, and what time it is (from the satellite clocks).

Differencing (I)

Algebraically, we create **combinations of observables** that remove the effects of clock errors. Consider:

$${}^{i}_{k}d - {}^{j}_{k}d = c({}^{i}_{k}t - {}^{i}t) - c({}^{j}_{k}t - {}^{j}t)$$

If the signals are received at the same time, the combination

$${}^{i}_{k}d - {}^{j}_{k}d = c({}^{j}t - {}^{i}t)$$

is *independent* of the receiver clock; we have subtracted it, and its errors, out. This is called a **between-satellites single difference**.

Differencing (I)

Algebraically, we create **combinations of observables** that remove the effects of clock errors. Consider:

$${}^{i}_{k}d - {}^{j}_{k}d = c({}^{i}_{k}t - {}^{i}t) - c({}^{j}_{k}t - {}^{j}t)$$

If the signals are received at the same time, the combination

$${}^{i}_{k}d - {}^{j}_{k}d = c({}^{j}t - {}^{i}t)$$

is *independent* of the receiver clock; we have subtracted it, and its errors, out. This is called a **between-satellites single difference**.

Likewise, differencing between receivers for the same satellite gives ${}^{j}_{k}d - {}^{j}_{l}d = c({}^{j}_{k}t - {}^{j}_{l}t)$ and the satellite clock drops out.

Differencing (II)

The next step is to form the difference of the between-receivers single difference, which is called a **double difference**:

$${}^{i}_{k}d - {}^{i}_{l}d - ({}^{j}_{k}d - {}^{j}_{l}d) = c({}^{i}_{k}t - {}^{i}_{l}t - {}^{j}_{k}t + {}^{j}_{l}t)$$

in which the errors ${}^{i}\varepsilon$, ${}^{j}\varepsilon$, ${}_{k}\varepsilon$, and ${}_{l}\varepsilon$ all cancel out.

Differencing (II)

The next step is to form the difference of the between-receivers single difference, which is called a **double difference**:

$$ {}^{i}_{k}d - {}^{i}_{l}d - ({}^{j}_{k}d - {}^{j}_{l}d) = c({}^{i}_{k}t - {}^{i}_{l}t - {}^{j}_{k}t + {}^{j}_{l}t) $$

in which the errors ${}^{i}\varepsilon$, ${}^{j}\varepsilon$, ${}_{k}\varepsilon$, and ${}_{l}\varepsilon$ all cancel out.

However, this double-difference depends on the **relative position** of the receivers: the **baseline** between them.

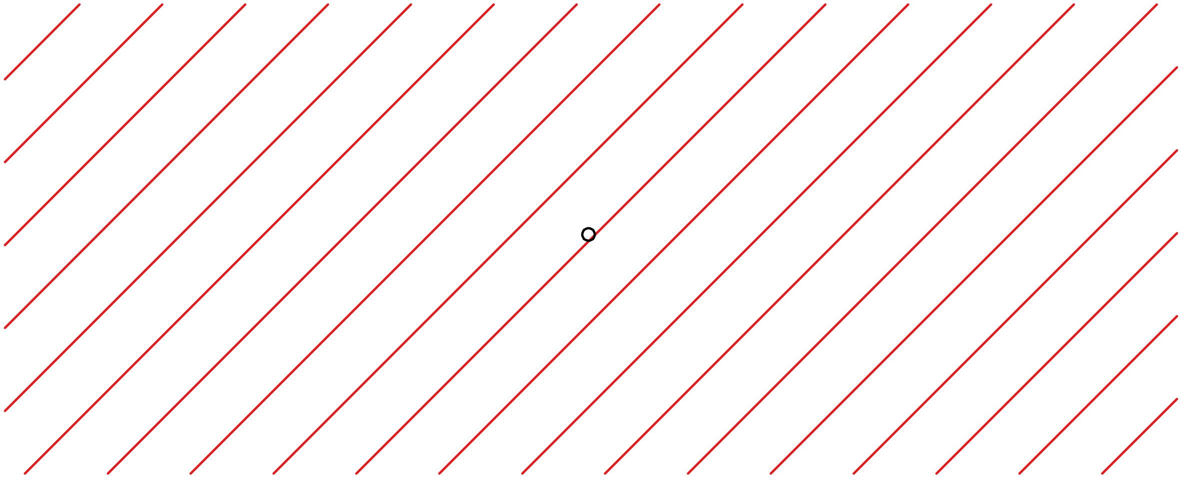
Ambiguity Resolution

Using the carrier-best-phase introduces **ambiguities**: since a sine wave repeats with wavelength λ, the distance to a satellite is (initially) uncertain to within a multiple of λ. Getting finding the correct multiple is **ambiguity resolution**.

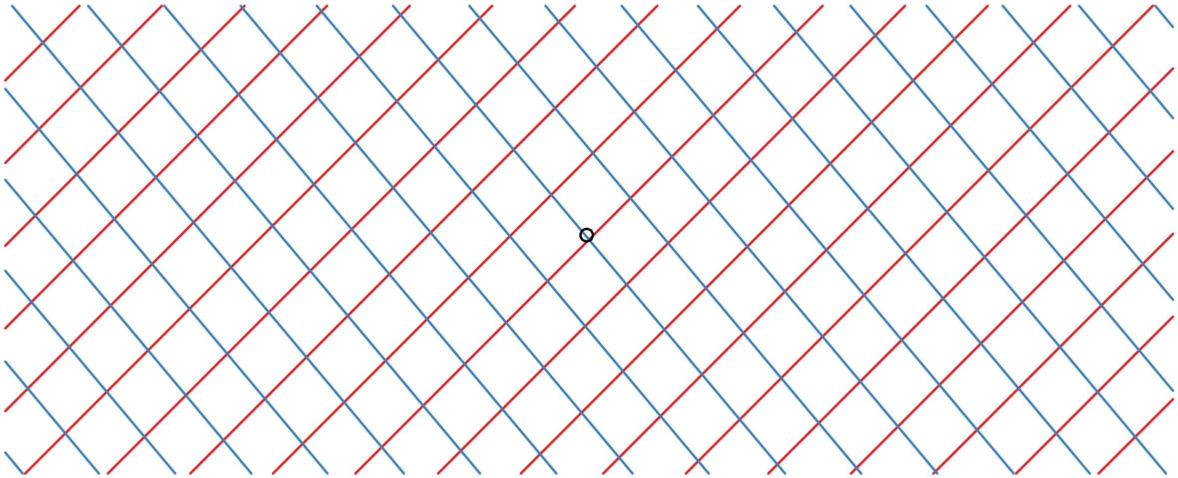
- Resolving the ambiguities removes unknowns from the solution and improves it.
- Because of error, our initial estimate of the ambiguity factor is never exactly an integer.
- Getting the wrong integer gives a very incorrect answer: unresolved is better than incorrect.
- Resolving ambiguities depends on having waves with different directions of arrival: multiple satellites, or the same satellite over a long time.
- **Single-epoch** estimates (needed for GPS seismology) involve ambiguity resolution from one time sample from multiple satellites.

Ambiguity Resolution: Point with Errors

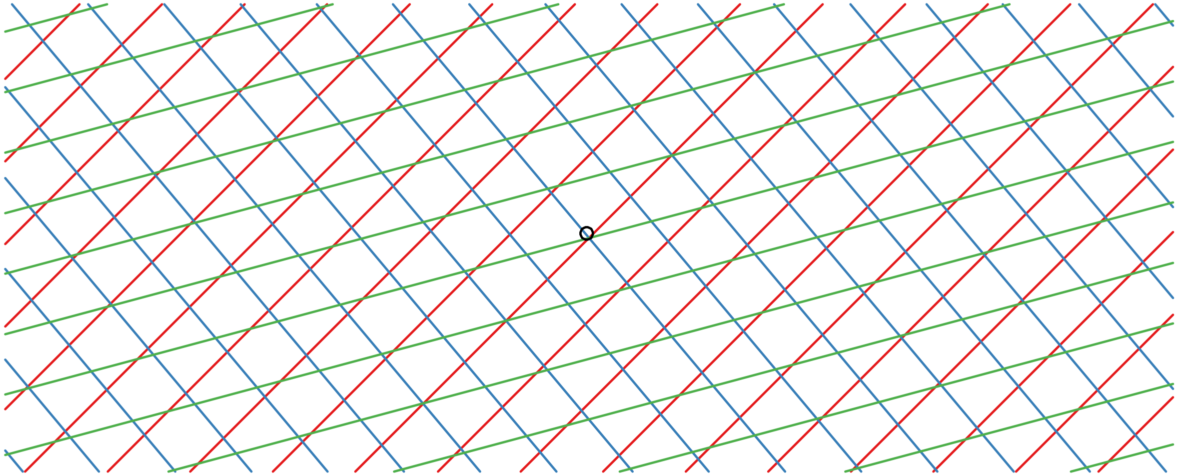
Ambiguity Resolution: First Satellite



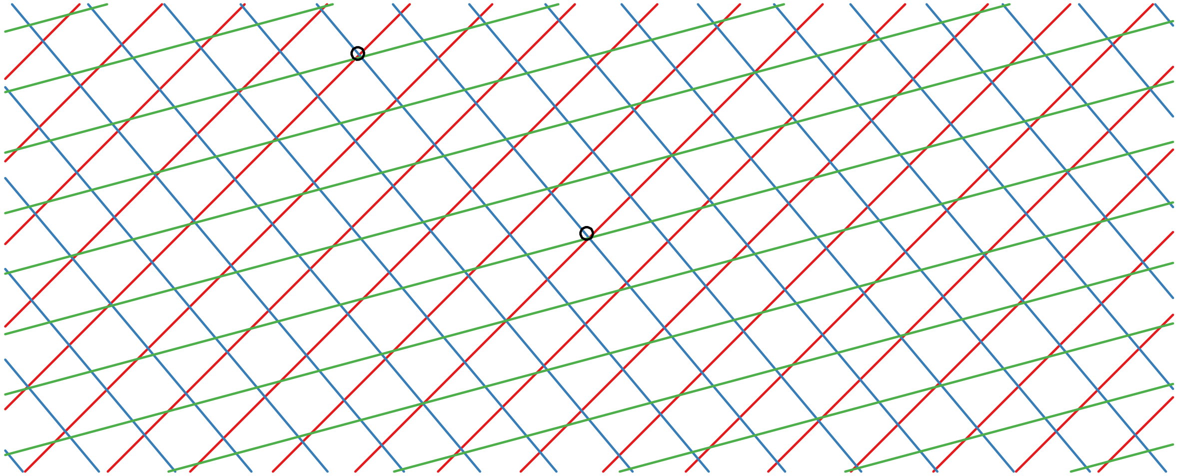
Ambiguity Resolution: First Two Satellites



Ambiguity Resolution: First Three Satellites



Ambiguity Resolution: Which Is It?



If we choose the wrong value for the ambiguity, we can be far from the correct value.

Propagation Delay

The velocity *c* varies along the path, because of:

- The ionosphere, from 400 to 60 km, which contains charged particles.

Propagation Delay

The velocity *c* varies along the path, because of:

- The ionosphere, from 400 to 60 km, which contains charged particles.
- The atmosphere, from 20 km to the surface; mostly, the troposphere, from 6 km to the surface, which contains

Propagation Delay

The velocity c varies along the path, because of:

- The ionosphere, from 400 to 60 km, which contains charged particles.
- The atmosphere, from 20 km to the surface; mostly, the troposphere, from 6 km to the surface, which contains
  - Nitrogen, oxygen, carbon dioxide etc. all well-mixed.

Propagation Delay

The velocity *c* varies along the path, because of:

- The ionosphere, from 400 to 60 km, which contains charged particles.
- The atmosphere, from 20 km to the surface; mostly, the troposphere, from 6 km to the surface, which contains
  - Nitrogen, oxygen, carbon dioxide etc. all well-mixed.
  - **Precipitable water vapor**, not well mixed.

Propagation Delay

The velocity c varies along the path, because of:

- The ionosphere, from 400 to 60 km, which contains charged particles.
- The atmosphere, from 20 km to the surface; mostly, the troposphere, from 6 km to the surface, which contains
    - Nitrogen, oxygen, carbon dioxide etc. all well-mixed.
    - **Precipitable water vapor**, not well mixed.

These bend the wave (which we can ignore), and delay it (which we cannot).

Propagation Delay

The velocity *c* varies along the path, because of:

- The ionosphere, from 400 to 60 km, which contains charged particles.
- The atmosphere, from 20 km to the surface; mostly, the troposphere, from 6 km to the surface, which contains
    - Nitrogen, oxygen, carbon dioxide etc. all well-mixed.
    - **Precipitable water vapor**, not well mixed.

These bend the wave (which we can ignore), and delay it (which we cannot).

Ionospheric delays can be 10 m or more, and can change very rapidly.

Propagation Delay

The velocity *c* varies along the path, because of:

- The ionosphere, from 400 to 60 km, which contains charged particles.
- The atmosphere, from 20 km to the surface; mostly, the troposphere, from 6 km to the surface, which contains
    - Nitrogen, oxygen, carbon dioxide etc. all well-mixed.
    - **Precipitable water vapor**, not well mixed.

These bend the wave (which we can ignore), and delay it (which we cannot).

Ionospheric delays can be 10 m or more, and can change very rapidly.

Tropospheric delays can be up to 2.5 m, and change more slowly.

Reducing Propagation Delay Effects

Ionosphere: delay depends on frequency; with L1 and L2 we can form an **ionosphere-free** observable.

Reducing Propagation Delay Effects

Ionosphere: delay depends on frequency; with L1 and L2 we can form an **ionosphere-free** observable.

We can model the **dry delay** (well-mixed gases) using a hydrostatic atmosphere; local meteorological data adds little to this.

Reducing Propagation Delay Effects

Ionosphere: delay depends on frequency; with L1 and L2 we can form an **ionosphere-free** observable.

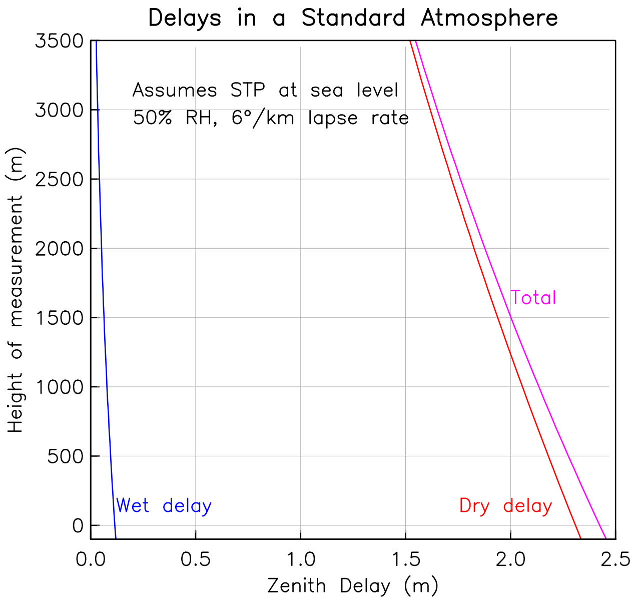
We can model the **dry delay** (well-mixed gases) using a hydrostatic atmosphere; local meteorological data adds little to this.

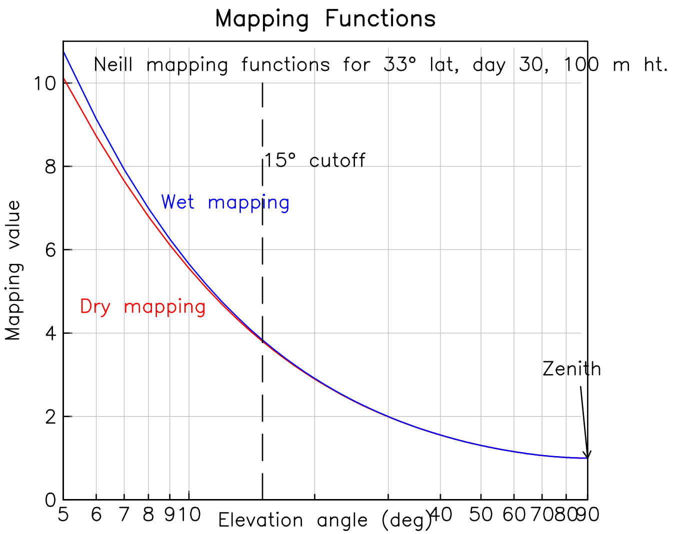
The **wet delay** cannot be modeled or be measured independently: it must be estimated using the GPS data.

We assume the delay looks like $Z(t)M(\theta)$ where

- $Z(t)$ is the **zenith delay**, varying with time.
- $M(\theta)$ is a **mapping function** of the elevation angle θ.

Propagation Delay: Models





Information Needed for Relative Positioning

- Data from the receiver at the point of interest: carrier-beat phases, and pseudoranges, to all visible satellites.
- The same for a reference receiver whose position is known.
- The positions of the satellites (“orbit information”), from the IGS or another procssing center.
- **Earth Orientation Parameters** used to connect a position on the Earth to inertial space, from the IERS,
- Satellite information (e.g. which ones are working) from the USNO.
- Antenna phase pattern (predetermined).

Information Needed for Point Positioning

- Data from the receiver at the point of interest: carrier-beat phases, and pseudoranges, to all visible satellites.
- The positions of the satellites (“orbit information”), from the IGS or another procssing center.
- A description of the satellite clock errors, from the IGS or another processing center.
- **Earth Orientation Parameters** used to connect a position on the Earth to inertial space, from the IERS,
- Satellite information (e.g. which ones are working) from the USNO.
- Antenna phase pattern (predetermined).

Steps in Relative Positioning

- Form double-difference combinations (not all, just a unique set).
- Get a preliminary position (or positions) using pseudorange information.
- Solve for:
  - Positions of unknown receivers (3 parameters)
  - Zenith delays at some time spacing, for each receiver.
  - Ambiguities at each receiver for all satellites observed there.
- Set ambiguity values to nearest integer, for all cases where this can safely be done.
- Repeat the solution for positions and zenith delays, with (we hope) many fewer parameters, since ambiguities have been resolved.

Local Effects I: Antenna Phase Delay

The ideal antenna would respond only to signal above the horizon, and not introduce any time delay: neither is realistic. In fact, if the ideal signal were $U_0 e^{2\pi i f t}$ the actual one will have two additional terms.

Local Effects I: Antenna Phase Delay

The ideal antenna would respond only to signal above the horizon, and not introduce any time delay: neither is realistic. In fact, if the ideal signal were $U_0 e^{2\pi ift}$ the actual one will have two additional terms.

First, we will have $U_0 e^{2\pi ift}[e^{i\phi_A(\theta_0,\beta_0)}]$ where ϕ_A is the phase shift introduced by the antenna itself, as a function of the elevation angle θ_0 and azimuth β_0 of the incoming signal; this shift includes any offset of the antenna "phase center" from the reference point on the antenna.

In general, this will be reduced if the same antenna types are used, or we have a model for ϕ_A: most antennas do.

Local Effects II: Multipath

In addition, we will have a term $U_0 e^{2\pi i f t}[\int_{\Omega^-} A(\theta, \beta) R(\theta, \beta) e^{i\phi_R(\theta,\beta)} d\theta\, d\beta]$ The integral term is meant to include all the "multipath" contributions, and so is an integral over Ω^-, which denotes the unit sphere excluding the direction of the direct wave. This includes

Local Effects II: Multipath

In addition, we will have a term $U_0 e^{2\pi ift}[\int_{\Omega^-} A(\theta, \beta) R(\theta, \beta) e^{i\phi_R(\theta,\beta)} d\theta\, d\beta]$ The integral term is meant to include all the "multipath" contributions, and so is an integral over Ω^-, which denotes the unit sphere excluding the direction of the direct wave. This includes

- Reflections from the ground – large at low angles.

<

Local Effects II: Multipath

In addition, we will have a term $U_0 e^{2\pi i f t}[\int_{\Omega^-} A(\theta, \beta)R(\theta, \beta)e^{i\phi_R(\theta,\beta)}d\theta\, d\beta]$ The integral term is meant to include all the "multipath" contributions, and so is an integral over Ω^-, which denotes the unit sphere excluding the direction of the direct wave. This includes

- Reflections from the ground – large at low angles.
- Reflections from other things nearby (trees, buildings).

Local Effects II: Multipath

In addition, we will have a term $U_0 e^{2\pi i f t}[\int_{\Omega^-} A(\theta, \beta) R(\theta, \beta) e^{i\phi_R(\theta,\beta)} d\theta\, d\beta]$ The integral term is meant to include all the "multipath" contributions, and so is an integral over Ω^-, which denotes the unit sphere excluding the direction of the direct wave. This includes

- Reflections from the ground – large at low angles.
- Reflections from other things nearby (trees, buildings).
- Scattering from the antenna support.

Local Effects II: Multipath

In addition, we will have a term $U_0 e^{2\pi i f t}[\int_{\Omega^-} A(\theta, \beta) R(\theta, \beta) e^{i\phi_R(\theta,\beta)} d\theta\, d\beta]$ The integral term is meant to include all the "multipath" contributions, and so is an integral over Ω^-, which denotes the unit sphere excluding the direction of the direct wave. This includes

- Reflections from the ground – large at low angles.
- Reflections from other things nearby (trees, buildings).
- Scattering from the antenna support.

None of these can be modeled well, so they are a source of noise that limits the precision of measurements made over short times; over long times this effect averages, somewhat.