Astronomy's much more fun when you're not an astronomer.

Brian May

Arizona State University
AST 112

Introduction to Stars, Galaxies and Cosmology

Frank Timmes

francis.timmes@asu.edu

www.cococubed.com/class_pages/class_astro.shtml

Image of the Day

First picture of a planet around a normal star similar to the Sun.
Young star 1RXS J160929.1-210524, lies about 500 light-years away.

The basic principles behind the Global Positioning System (GPS) are really quite simple - even though the system itself employs some "high-tech" equipment.

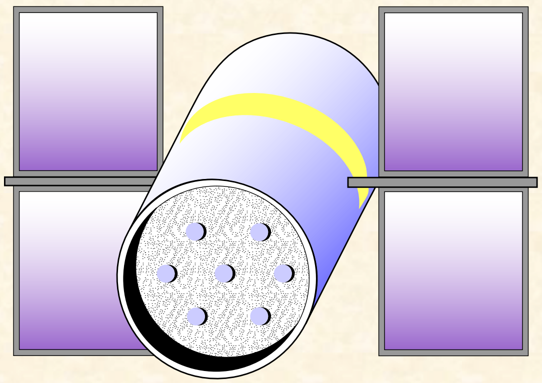
Let's break the system into 4 pieces and take those pieces one step at a time.

We'll start with the "big" idea and ignore some of the details. Later we'll fill in those fine points.

Modern GPS IIR-M satellite

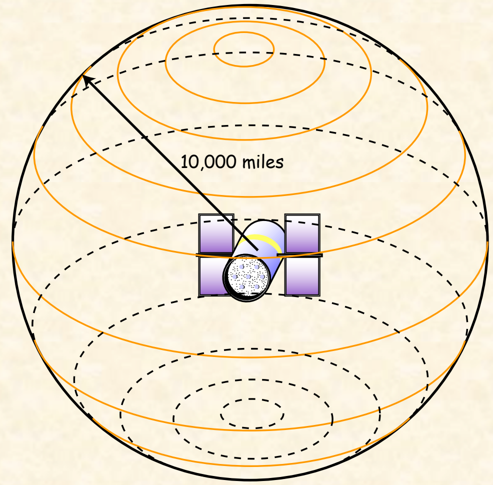
Step 1 - The basic idea is satellite ranging.

That means we figure our position on earth by measuring our distance from a group of satellites in space. The satellites act as precise reference points for us.



Let's say we're lost and were trying to locate ourselves.

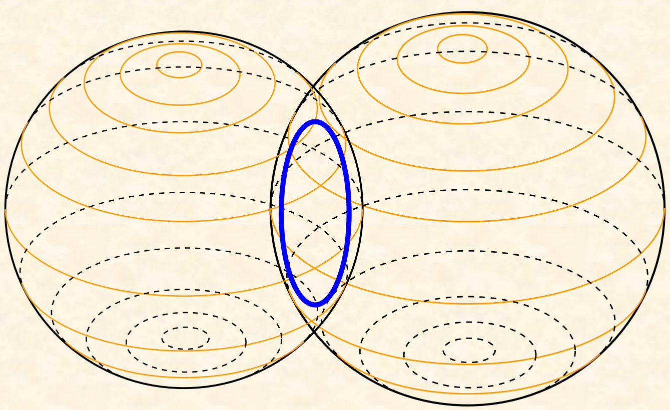
If we know we are a certain distance from satellite A, say 10,000 miles, that really narrows where in the whole universe we can be.



It tells us we must be somewhere on an imaginary sphere that is centered on the satellite and that has a radius of 10,000 miles.

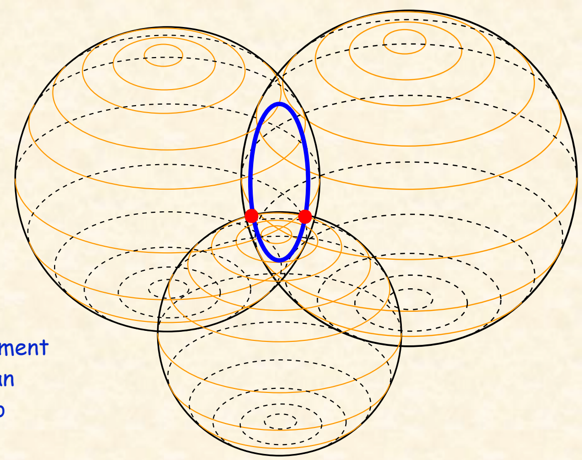
Now if at the same time we also know that we're 12,000 miles from another satellite, satellite B, that narrows down where we can be even more.

The only place in the universe where we can be 10,000 miles from satellite A and 12,000 miles from satellite B is on the circle where those two spheres intersect.



If we know that at the same time we're 14,000 miles from satellite C, there are only two points in space where that can be true.

Those two points are where the 14,000 mile sphere cuts through the circle that's the intersection of the 10,000 mile sphere and the 12,000 mile sphere.



Finally, if we make a measurement from a fourth satellite, we can determine which of those two points is our true location.

And that's it! The basic principle behind GPS: using satellites as reference points for determining your position somewhere on earth.

Everything else about the system is technical detail designed to help carry out this ranging process - to make it more accurate or easier to do.

Step 1 Summary:

Position is calculated from distance measurements to satellites.

Mathematically we need four measurements to determine an exact position.

Step 2 - Measuring your distance from a satellite

Since GPS is based on knowing your distance to satellites in space, we need some method for figuring out how far we are from those satellites.

Surprisingly, the basic idea behind measuring a distance to a satellite is just our old friend distance = speed • time.

If your car goes 60 miles an hour for two hours, how far have you gone?
Your speed (60 miles/hr) times time (2 hr) equals distance (120 miles).

The GPS system works by timing how long it takes a radio signal to reach us from a satellite and then calculating the distance from that time.

Radio waves travel at the speed of light: 186,000 miles per second.

So if we can figure out exactly when the GPS satellite started sending its radio message and when we received it, we'll know how long it took to reach us.

We just multiply that time in seconds by 186,000 miles per second and that's our range to the satellite.

distance = speed of light * time

Now, our clocks have to be pretty good with short times because light moves fast.

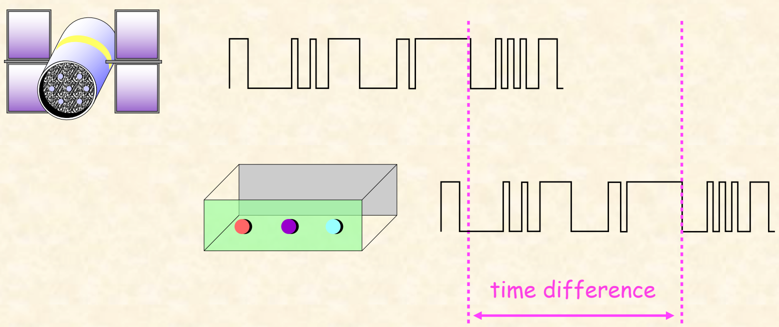
For example, if a GPS satellite were directly overhead it would only take about 6/100ths of a second for the radio message to get to us.

Most GPS receivers can measure time with nanosecond accuracy.
That's 0.000000001 second. We'll talk more about how they do that in a moment.

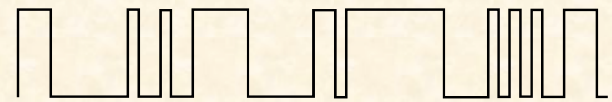
The trick to measuring the travel time of the radio signal is to figure out exactly when the signal left the satellite.

To do that we synchronize the satellites and receivers so they're generating the same pattern or code at exactly the same time.

We receive the code from a satellite and see how long ago our receiver generated the same code. The time difference is how long the signal took to get down to us.



Both satellites and receivers actually generate a complicated set of digital codes.



The codes are made complicated on purpose so that they can be compared easily and unambiguously. The codes almost look like a long string of random pulses.

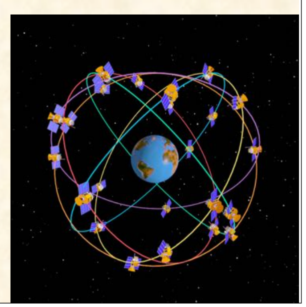
They're not really random though, as the sequences repeat every millisecond. So they are often referred to as pseudo-random codes.

Step 2 Summary:

The distance to a satellite is determined by measuring how long a radio signal takes to reach us from that satellite.

We assume that both the satellite and our receiver are generating the same pseudo-random code at exactly the same time.

We know how long it took for the satellite's signal to get to us by comparing how late its pseudo-random code is relative to our code.



Step 3 - Getting perfect timing

Light travels at 186,000 miles/sec. If the satellite and our receiver were out of synch by 1/100th of a second, our distance would be off by 1,860 miles!

How do we know both our receiver and the satellite are really generating their codes at exactly the same time??

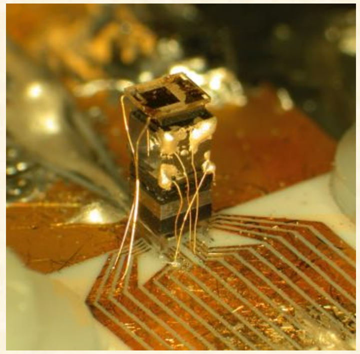
Well, at least one side of the clock synch problem is easy to explain: the satellites have atomic clocks on board.

GPS atomic clock

Atomic clocks are extremely precise and expensive. They cost about $250,000 apiece and each satellite has four, just to be sure one is always working.

Atomic clocks don't run on atomic energy. They use the oscillations of a particular atom as their "metronome".

It's the most stable and accurate time reference humans have ever developed.

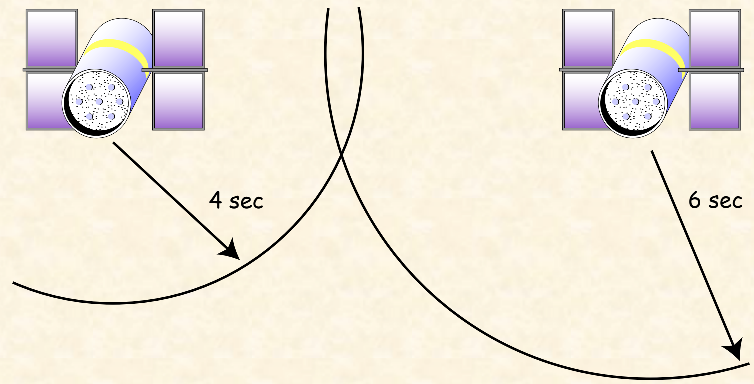


NIST atomic clock on a chip

That's fine for the satellites, but what about us mortals down here on earth? We can't have a$250,000 atomic clock in every portable GPS receiver.

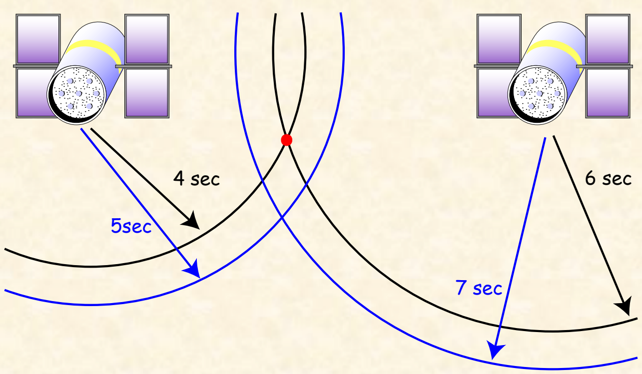
Fortunately, we can get by with only moderately accurate clocks in our receivers.

Let's say, in reality, we're 4 sec from satellite A and 6 sec from satellite B. In two dimensions, these two ranges would locate us at a point.



But what if we used our "imperfect" receiver, which is, say, 1 sec too fast?

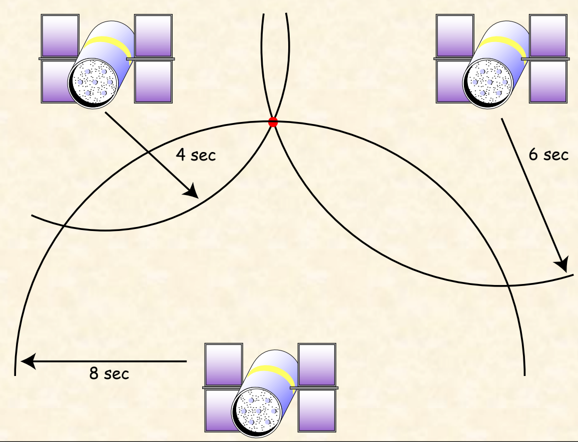
It would call the time 4 sec to satellite A and the time 7 sec to satellite B. That causes the two circles to intersect at a different point.



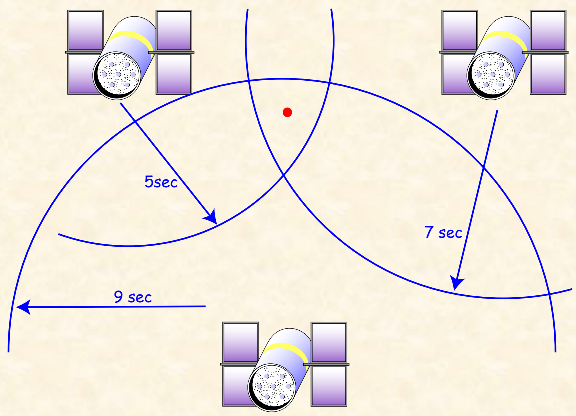
This would seem like a perfectly correct answer to us, since we'd have no way of knowing our receiver was a little fast.

Now let's add another measurement. Let's say in reality (if we had perfect clocks) satellite C is eight seconds from our true position.

All three circles intersect at the red dot because those circles represent the true ranges to the three satellites.



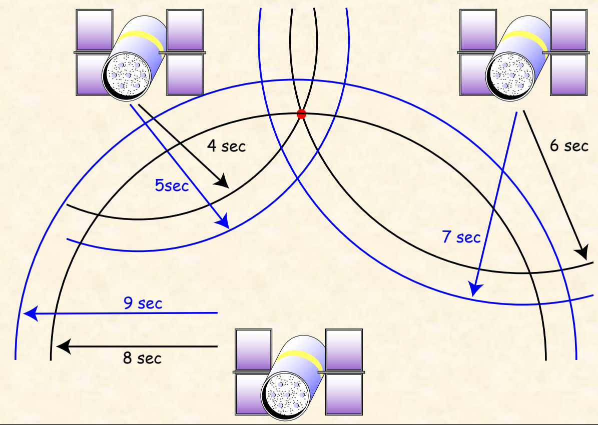
Now let's add 1 sec of timing error. There is no point that can be 5 sec from A, 7 sec from B, and 9 sec from C.



When our receivers get a series of measurements that cannot intersect at a single point, they assume their internal clock is off - that it has some offset.

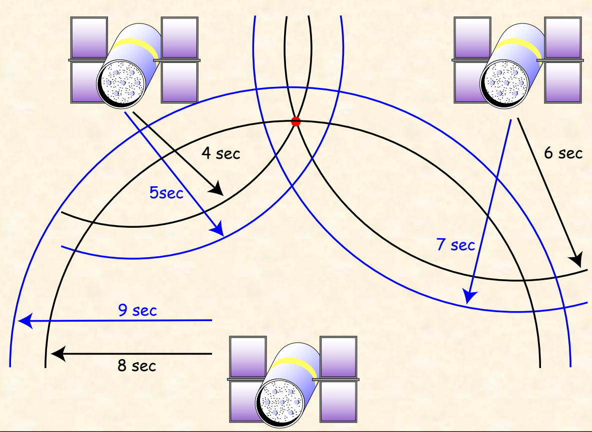
So the receiver starts subtracting (or adding) time, the same amount of time from all the measurements.

It keeps trimming off time until it hits on an answer that lets all the ranges go through one point.



In essence, it "discovers" that by subtracting 1 sec from all the measurements that it can make the circles intersect at a point.

It doesn't really aimlessly hunt for an answer. The receivers applies algebra to the old "three equations and three unknowns", and quickly computes the clock offset.



Step 3 Summary:

Accurate timing is the key to measuring distance to satellites.

Satellites are accurate because they have atomic clocks on board.

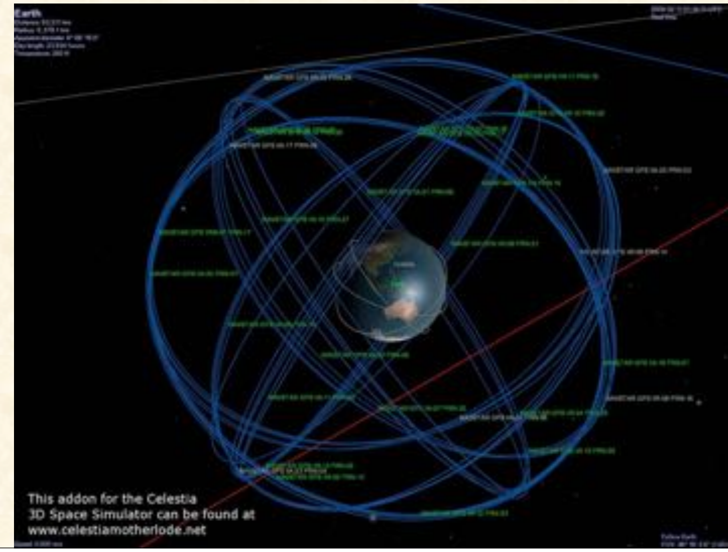
Receiver clocks don't have to be perfect because simple algebra can cancel out receiver clock errors.

Step 4 - Knowing where a satellite is in space

We've been assuming, in all our discussions so far, that we know right where all the satellites are in space so we can locate our position from theirs.

But how do we know where something that's 11,000 miles up in space is? Well, that 11,000 mile altitude is actually a benefit because its well clear of earth's atmosphere.

This means predictions of GPS satellites orbits will be very accurate.
Like the Moon, which has reliably orbited Earth for millions of years without any significant change in period, our GPS satellites are orbiting very predictably.



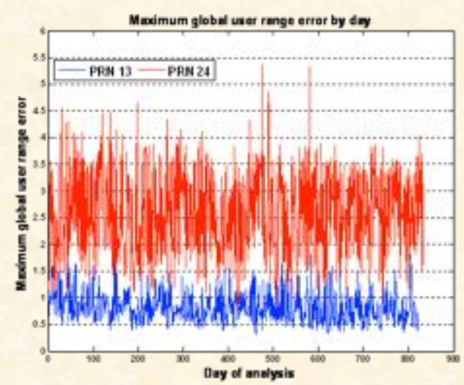
The Air Force injects each satellite into a very precise orbit.

GPS receivers on the ground have an "almanac" in their computer's memory, which tells them where in the sky each satellite will be at any given moment.

A model of the orbits would be pretty accurate by itself. To make things perfect, the GPS satellites are constantly monitored by the DoD.

Twice a day the DoD measures each satellite's position and speed. Variations are usually very minor and are caused by things like the solar radiation pressure.

Once the DoD has determined any orbit corrections, they relay that information up to the satellite. The satellite then broadcasts these corrections along with its timing information.



Step 4 Summary:

To calculate our position we need to know where our satellites are in space.

GPS satellites are so high up, their orbits are very predictable.

Minor variations in orbits are measured constantly by the DoD and that data is transmitted by the satellites.