What is GPS?

The Global Positioning System (GPS) is a worldwide radio-navigation system formed from a constellation of 24 satellites and their ground stations. GPS uses these "man-made stars" as reference points to calculate positions accurate to a matter of meters. In fact, with advanced forms of GPS you can make measurements to better than a centimeter!

GPS receivers have been miniaturized to just a few integrated circuits and so are becoming very economical. And that makes the technology accessible to virtually everyone. These days GPS is finding its way into cars, boats, planes, construction equipment, movie making gear, farm machinery, even laptop computers. Soon GPS will become almost as basic as the telephone. Indeed, at Trimble, we think it just may become a universal utility.

How GPS Works: Five Logical Steps

- The basis of GPS is "triangulation" from satellites.
- To "triangulate," a GPS receiver measures distance using the travel time of radio signals.
- To measure travel time, GPS needs very accurate timing which it achieves with some tricks.
- Along with distance, you need to know exactly where the satellites are in space. High orbits and careful monitoring are the secret.
- Finally you must correct for any delays the signal experiences as it travels through the atmosphere.

1) Triangulating from Satellites

Improbable as it may seem, the whole idea behind GPS is to use satellites in space as reference points for locations here on earth. That's right, by very, very accurately measuring our distance from three satellites we can "triangulate" our position anywhere on earth.

Forget for a moment how our receiver measures this distance. We'll get to that later. First consider how distance measurements from three satellites can pinpoint you in space.

The Big Idea Geometrically:

Suppose we measure our distance from a satellite and find it to be 11,000 miles. Knowing that we're 11,000 miles from a particular satellite narrows down all the possible locations we could be in the whole universe to the surface of a sphere that is centered on this satellite and has a radius of 11,000 miles.

Next, say we measure our distance to a second satellite and find out that it's 12,000 miles away. That tells us that we're not only on the first sphere but we're also on a sphere that's 12,000 miles from the second satellite. Or in other words, we're somewhere on the circle where these two spheres intersect.

If we then make a measurement from a third satellite and find that we're 13,000
miles from that one, that narrows our position down even further, to the two points where the 13,000 mile sphere cuts through the circle that's the intersection of the first two spheres. So by ranging from three satellites we can narrow our position to just two points in space.

To decide which one is our true location we could make a fourth measurement. But usually one of the two points is a ridiculous answer (either too far from Earth or moving at an impossible velocity) and can be rejected without a measurement. A fourth measurement does come in very handy for another reason however, but we'll tell you about that later.

In Review: **Triangulating**

- Position is calculated from distance measurements (ranges) to satellites.
- Mathematically we need four satellite ranges to determine exact position.
- Three ranges are enough if we reject ridiculous answers or use other tricks.
- Another range is required for technical reasons to be discussed later.

Next we'll see how the system measures distances to satellites:

2) **Measuring Distance from a Satellite**

We saw in the last section that a position is calculated from distance measurements to at least three satellites. But how can you measure the distance to something that's floating around in space? We do it by timing how long it takes for a signal sent from the satellite to arrive at our receiver.

**The Big Idea Mathematically**

In a sense, the whole thing boils down to those "velocity times travel time" math problems we did in high school. Remember the old: "If a car goes 60 miles per hour for two hours, how far does it travel?"

**Velocity (60 mph) x Time (2 hours) = Distance (120 miles)**

In the case of GPS we're measuring a radio signal so the velocity is going to be the speed of light or roughly 186,000 miles per second. The problem is measuring the travel time.

The timing problem is tricky. First, the times are going to be awfully short. If a satellite were right overhead the travel time would be something like 0.06 seconds. So we're going to need some really precise clocks. We'll talk about those soon.

But assuming we have precise clocks, how do we measure travel time? To explain it let's use a goofy analogy: Suppose there was a way to get both the satellite and the receiver to start playing "The Star Spangled Banner" at precisely 12 noon. If sound could reach us from space (which, of course, is ridiculous) then standing at the receiver we'd hear two versions of the Star Spangled Banner, one from our receiver and one from the satellite.

These two versions would be out of sync. The version coming from the satellite would be a little delayed because it had to travel more than 11,000 miles. If we wanted to see just how delayed the satellite's version was, we could start delaying
the receiver's version until they fell into perfect sync.

The amount we have to shift back the receiver's version is equal to the travel time of the satellite's version. So we just multiply that time times the speed of light and BINGO! we've got our distance to the satellite.

That's basically how GPS works. Only instead of the Star Spangled Banner the satellites and receivers use something called a "Pseudo Random Code" - which is probably easier to sing than the Star Spangled Banner.

**A Random Code?**

The Pseudo Random Code is a fundamental part of GPS. Physically it's just a very complicated digital code, or in other words, a complicated sequence of "on" and "off" pulses.

The signal is so complicated that it almost looks like random electrical noise. Hence the name "Pseudo-Random." There are several good reasons for that complexity: First, the complex pattern helps make sure that the receiver doesn't accidentally sync up to some other signal. The patterns are so complex that it's highly unlikely that a stray signal will have exactly the same shape.

Since each satellite has its own unique Pseudo-Random Code this complexity also guarantees that the receiver won't accidentally pick up another satellite's signal. So all the satellites can use the same frequency without jamming each other. And it makes it more difficult for a hostile force to jam the system. In fact the Pseudo Random Code gives the DoD a way to control access to the system.

But there's another reason for the complexity of the Pseudo Random Code, a reason that's crucial to making GPS economical. The codes make it possible to use "information theory" to "amplify" the GPS signal. And that's why GPS receivers don't need big satellite dishes to receive the GPS signals.

We glossed over one point in our goofy Star-Spangled Banner analogy. It assumes that we can guarantee that both the satellite and the receiver start generating their codes at exactly the same time. But how do we make sure everybody is perfectly synced? Stay tuned and see.

**In Review: Measuring Distance**

1) Distance to a satellite is determined by measuring how long a radio signal takes to reach us from that satellite.

2) To make the measurement we assume that both the satellite and our receiver are generating the same pseudo-random codes at exactly the same time.

3) By comparing how late the satellite's pseudo-random code appears compared to our receiver's code, we determine how long it took to reach us.

4) Multiply that travel time by the speed of light and you've got distance.

**3) Getting the Perfect Timing**
If measuring the travel time of a radio signal is the key to GPS, then our stop watches had better be damn good, because if their timing is off by just a thousandth of a second, at the speed of light, that translates into almost 200 miles of error!

On the satellite side, timing is almost perfect because they have incredibly precise atomic clocks on board. But what about our receivers here on the ground? Remembe that both the satellite and the receiver need to be able to precisely synchronize their pseudo-random codes to make the system work.

If our receivers needed atomic clocks (which cost upwards of $50K to $100K) GPS would be a lame duck technology. Nobody could afford it. Luckily the designers of GPS came up with a brilliant little trick that lets us get by with much less accurate clocks in our receivers. This trick is one of the key elements of GPS and as an added side benefit it means that every GPS receiver is essentially an atomic-accuracy clock.

The secret to perfect timing is to make an extra satellite measurement. That's right, if three perfect measurements can locate a point in 3-dimensional space, then four imperfect measurements can do the same thing. This idea is very fundamental to the working of GPS...here's a quick summary:

**Extra Measurement Cures Timing Offset**

If our receiver's clocks were perfect, then all our satellite ranges would intersect at a single point (which is our position). But with imperfect clocks, a fourth measurement, done as a cross-check, will NOT intersect with the first three. So the receiver's computer says "Uh-oh! there is a discrepancy in my measurements. I must not be perfectly synced with universal time." Since any offset from universal time will affect all of our measurements, the receiver looks for a single correction factor that it can subtract from all its timing measurements that would cause them all to intersect at a single point.

That correction brings the receiver's clock back into sync with universal time, and bingo! - you've got atomic accuracy time right in the palm of your hand. Once it has that correction it applies to all the rest of its measurements and now we've got precise positioning.

One consequence of this principle is that any decent GPS receiver will need to have at least four channels so that it can make the four measurements simultaneously. With the pseudo-random code as a rock solid timing sync pulse, and this extra measurement trick to get us perfectly synced to universal time, we have got everything we need to measure our distance to a satellite in space. But for the triangulation to work we not only need to know distance, we also need to know exactly where the satellites are. In the next section we'll see how we accomplish that.

In Review: **Getting Perfect Timing**

1) Accurate timing is the key to measuring distance to satellites.

2) Satellites are accurate because they have atomic clocks on board.
3) Receiver clocks don't have to be too accurate because an extra satellite range measurement can remove errors.

4) Satellite Positions - Knowing Where a Satellite is Located in Space

In this tutorial we've been assuming that we know where the GPS satellites are so we can use them as reference points. But how do we know exactly where they are? After all they're floating around 11,000 miles up in space.

**A high satellite gathers no moss**

That 11,000 mile altitude is actually a benefit in this case, because something that high is well clear of the atmosphere. And that means it will orbit according to very simple mathematics.

The Air Force has injected each GPS satellite into a very precise orbit, according to the GPS master plan. On the ground all GPS receivers have an almanac programmed into their computers that tells them where in the sky each satellite is, moment by moment.

The basic orbits are quite exact but just to make things perfect the GPS satellites are constantly monitored by the Department of Defense. They use very precise radar to check each satellite's exact altitude, position and speed. The errors they're checking for are called "ephemeris errors" because they affect the satellite's orbit or "ephemeris." These errors are caused by gravitational pulls from the moon and sun and by the pressure of solar radiation on the satellites. The errors are usually very slight but if you want great accuracy they must be taken into account.

4) Getting the message out

Once the DoD has measured a satellite's exact position, they relay that information back up to the satellite itself. The satellite then includes this new corrected position information in the timing signals it's broadcasting.

So a GPS signal is more than just pseudo-random code for timing purposes. It also contains a navigation message with ephemeris information as well. With perfect timing and the satellite's exact position you'd think we'd be ready to make perfect position calculations. But there's trouble afoot. Check out the next section to see what's up.

In Review: **Satellite Positions**

- To use the satellites as references for range measurements we need to know exactly where they are.
- GPS satellites are so high up their orbits are very predictable.
- Minor variations in their orbits are measured by the Department of Defense.
- The error information is sent to the satellites, to be transmitted along with the timing signals.

5) Error Correction

Up to now we've been treating the calculations that go into GPS very abstractly, as if the whole thing were happening in a vacuum. But in the real world there are lots
of things that can happen to a GPS signal that will make its life less than mathematically perfect.

To get the most out of the system, a good GPS receiver needs to take a wide variety of possible errors into account. Here's what they've got to deal with.

First, one of the basic assumptions we've been using throughout this tutorial is not exactly true. We've been saying that you calculate distance to a satellite by multiplying a signal's travel time by the speed of light. But the speed of light is only constant in a vacuum.

As a GPS signal passes through the charged particles of the ionosphere and then through the water vapor in the troposphere it gets slowed down a bit, and this creates the same kind of error as bad clocks. There are a couple of ways to minimize this kind of error. For one thing we can predict what a typical delay might be on a typical day. This is called modeling and it helps but, of course, atmospheric conditions are rarely exactly typical. Another way to get a handle on these atmosphere-induced errors is to compare the relative speeds of two different signals. This "dual frequency" measurement is very sophisticated and is only possible with advanced receivers.

Trouble for the GPS signal doesn't end when it gets down to the ground. The signal may bounce off various local obstructions before it gets to our receiver. This is called multipath error and is similar to the ghosting you might see on a TV. Good receivers use sophisticated signal rejection techniques to minimize this problem.

Problems at the satellite

Even though the satellites are very sophisticated they do account for some tiny errors in the system.

The atomic clocks they use are very, very precise but they're not perfect. Minute discrepancies can occur, and these translate into travel time measurement errors. And even though the satellites positions are constantly monitored, they can't be watched every second. So slight position or "ephemeris" errors can sneak in between monitoring times.

Basic geometry itself can magnify these other errors with a principle called "Geometric Dilution of Precision" or GDOP. It sounds complicated but the principle is quite simple. There are usually more satellites available than a receiver needs to fix a position, so the receiver picks a few and ignores the rest.

If it picks satellites that are close together in the sky the intersecting circles that define a position will cross at very shallow angles. That increases the gray area or error margin around a position.

If it picks satellites that are widely separated the circles intersect at almost right angles and that minimizes the error region. Good receivers determine which satellites will give the lowest GDOP.

In Review: Correcting Errors
1) The earth's ionosphere and atmosphere cause delays in the GPS signal that translate into position errors.

2) Some errors can be factored out using mathematics and modeling.

3) The configuration of the satellites in the sky can magnify other errors.

4) Differential GPS can eliminate almost all error.

**Putting GPS to work:**

GPS technology has matured into a resource that goes far beyond its original design goals. These days scientists, sportsmen, farmers, soldiers, pilots, surveyors, hikers, delivery drivers, sailors, dispatchers, lumberjacks, fire-fighters, and people from many other walks of life are using GPS in ways that make their work more productive, safer, and sometimes even easier.