6.6 EXERCISES

1) How was latitude and longitude determined in the 1800s.
2) Describe three ways that differential GPS correction data can be obtained.
3) Find a current differential GPS coverage map for the North America on the web.
4) Investigate whether GPS differential correction beacons are available in your area.

6.7 REFERENCES

DYE, STEVE and BAYLIN, FRANK (1997) The GPS Manual: Principles and Applications. Baylin/Gale Productions.

KAPLAN, ELLIOTT (1996) Understanding GPS: Principles and Applications. Artech House.

TRIMBLE NAVIGATION (1989) GPS: A Guide to the Next Utility.

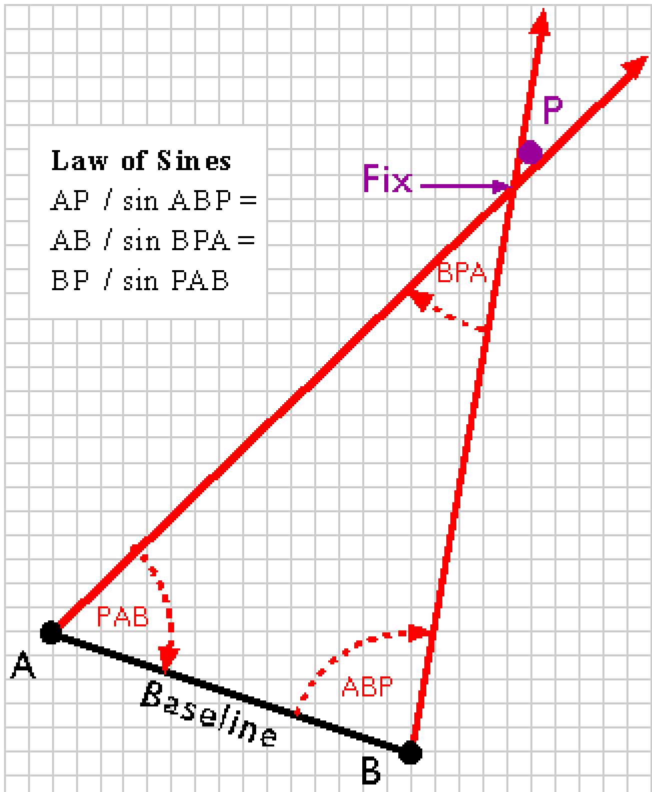

Figure 6.14 The method of triangles in surveying. Surveyors first measure the baseline distance AB between the two known points. They then measure the angles PAB and ABP. The trigonometric law of sines is then used to calculate the lengths of any other side or interior angle of the triangle.

6.5 SUMMARY

From earliest times humans needed a way to determine their position. The human struggle to 'find' themselves in the world reflects the conquest of the world beyond what they could see from any one point. They could travel and be confident that they could return to the same place. They could determine in which direction they needed to travel and they could travel great distances across oceans without getting lost. In a very real sense, humans fought a war with the space around them. Their conquest of this space was as simple as determining where they were located.

Figure 6.13 The Duke of Cumberland's theodolite. Benjamin Cole, a prominent London instrument maker, made this theodolite for the Duke of Cumberland, George III's uncle, in the 1740s or 1750s. A theodolite is a surveying instrument that measures both horizontal and vertical angles. A new invention in the eighteenth century, the theodolite was not in fact widely used until the nineteenth. Science Museum/Science & Society Picture Library.

could accurately measure long distances. The first practical use of triangulation in mapping a country began in 1668, when Jean Picard and Jean Dominique Cassini, two French astronomers and mathematicians, began the first scientific survey of France.

In the 1970s, electronic distance measurers (EDMs) were introduced that measured the distance between the device and a reflector using a special laser beam. Most surveying still takes place with these electronic measuring devices.

There are two main types of surveying. *Geodetic surveying* takes into account the theoretical shape of the earth. It is generally high in accuracy, and covers large areas (greater than 300 sq. mi.). Plane Surveying assumes that the survey area is a flat plane and covers small areas of less than 300 sq. mi. This is the most common method of surveying.

Figure 6.12 Differential GPS receiver. The device simultaneously acquires data from GPS satellites and from ground control stations called beacons. The beacon data is transmitted by radio waves and contains the current error in the GPS signal. With the known error, the differential GPS unit can calculate the position of a point to within a meter or less.

6.4.7 Surveying Applications

Differential GPS can supply accurate latitude and longitude measurements down to the centimeter and is opening up new applications in surveying. Surveying is defined as the science and art of determining the relative positions of points above, on, or beneath the surface of the earth and has a long history. There is some evidence that Egyptians first used surveying techniques to accurately divide land into plots for the purpose of taxation. The Greeks developed the science of geometry and were using it for precise land division, developed surveying equipment and standardized methods of conducting surverys.

Military leaders recognized the importance of accurate maps and this led to the development of a variety of surveying methods. In the 1600s, the French refined the methods of geometry used in surveying, particularly triangulation. Triangulation revolutionized mapmaking, producing land measurements far more accurate than earlier estimates. In the example illustrated in Figure 6.13, surveyors first measure the baseline distance AB between the two known points. They then measure the angles PAB and ABP. The trigonometric law of sines is then used to calculate the lengths of any other side or interior angle of the triangle. Knowing these dimensions, surveyors can plot point P in a plane coordinate system grid. Note that measurement error causes the fix at the intersection of the AP and BP to be slightly inaccurate.

To begin practical triangulation, a 17th-century surveyor laid out an initial triangle by plotting three points: two at the ends of a base line and one at a distant landmark. Using a theodolite, the surveyor measured the angles towards the landmark at both ends of the base line. The baseline itself was measured by counting revolutions of a wheel or by using measuring chains or rods laid end to end. Once one triangle had been measured, further triangles could be constructed from its sides, producing chains of triangles that

Figure 6.10 GPSDifferential Beacon Coverage in North America in 1999. Beacons along the coast areas are operated by the U.S. Coast Guard. Beacons along navigable rivers have belong to the U.S. Army Corps of Engineers. Many inland states have also constructed beacons.

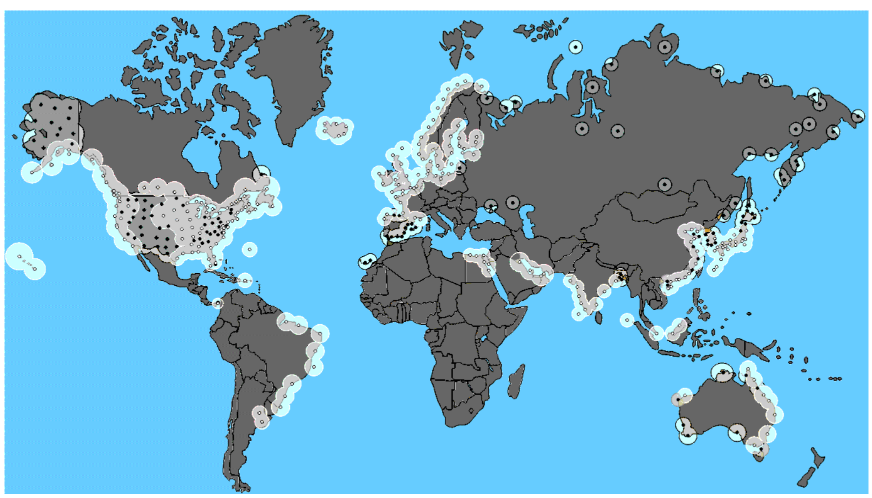

Figure 6.11 World Differential GPS coverage. Most coastal areas in the Northern Hemisphere have access to differential GPS data from beacons. In most of the world, inland areas do not have access to this type of data.

Figure 6.9 World Differential GPS Coverage by privately owned satellites. Omnistar provides differential GPS data from ground stations via satellite. A special receiver is needed to download the differential data from the private satellite. This data is then compared to that received by the GPS receiver.

State governments and foreign governments have added beacons in inland areas (see U.S. beacon coverage map in Figure 6.10). These beacons now represent the major means of supplying differential GPS data for most of the world (Figure 6.11). The more expensive GPS receivers all integrate beacon radio receivers so that they can simultaneously receive data from the GPS satellites and ground-based beacons (Figure 6.12).

6.4.6 Applications of GPS

A variety of applications have developed for differential GPS where sub-meter accuracy is required. GPS is used in airplane, car, and marine navigation, land survey, mapping & GIS, marine survey, mining, public safety, and timing and synchronization applications. In 1998 the cost of a car navigation that includes non-differential GPS was less than $2000. The system consists of a small, color screen that mounts in the passenger compartment; a computer, a direction sensor and the GPS unit. Dead-reckoning and map-matching are used to precisely locate the vehicle and visual and voice prompts to guide the driver turn-by-turn to the destination.

One of the most intriguing uses of GPS is in agriculture. The method, called precision agriculture, involves gathering data on yields as the field is being harvested. This is done by attaching a differential GPS unit to the harvester. The following Spring, fertilizer is applied at a different rate across the field according to the yields of the previous year. The purpose of the procedure is to apply fertilizer only to those parts of the field where it would actually increase the yield. In many cases, the savings in fertilizer in a single year outweigh the cost of the GPS receiver and associated computers.

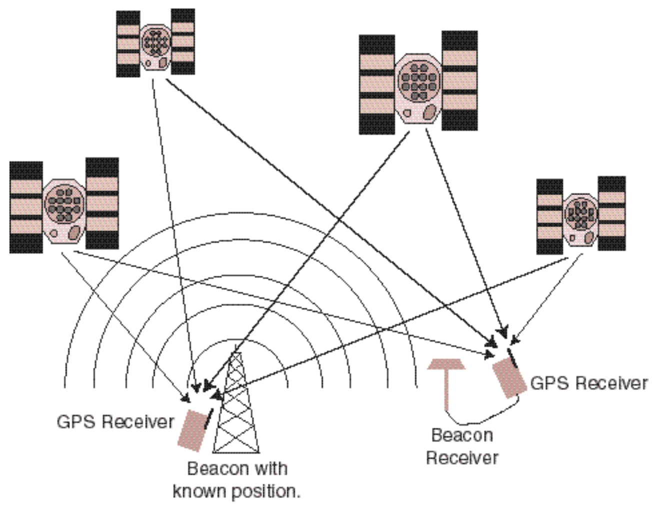

Figure 6.8 Differential GPS. To obtain more accurate measurements than is possible from a single GPS unit, a GPS receiver broadcasts the signal it receives from a know position. The GPS unit in the field simultaneously receives data from the GPS satellites and the other GPS receiver on the ground through a radio signal. The GPS error from the known position is compared to that of the GPS receiver in the unknown location.

GPS measurements can be made much more accurate using two receivers. One of the receivers is placed in a known location and calculates the error in the satellite range data. It acts as a static reference point because its position is precisely known. The correction data that it calculates can then be applied to all other receivers in the same area, thereby eliminating most of the error in their measurements (see Figure 6.8). The concept works because the satellites are so far away that any errors measured by one receiver will be the same for all other receivers in the area. This differential correction factor simultaneously resolves all errors in the GPS signal, whether the error is associated with the receiver clocks, satellite clocks, satellite position, or ionospheric or atmospheric delays.

A variety of ways have developed to supply the GPS differential correction data. Private companies have established GPS ground control stations that transmit their data to their own satellites. Special receivers are then used to receive this data and this is compared to the GPS measurements made on the ground. This differential data can be obtained in most of the world (see Figure 6.9). A second method makes use of FM radio stations. Each FM station has a certain part of its frequency – called the sub-carrier – that can transmit data to a radio that is capable of receiving this signal. Although the data rate is slow, it can provide the needed differential data from a GPS ground control station in the area. The price charged by the companies that offer the differential correction data by satellite or FM subcarrier varies by the desired level of accuracy.

The third method of transmitting differential GPS data is through ‘marine’ beacons. These beacons operate in the 283.5 - 325.0 kHz frequency range and the signal is free of charge. The U.S. Coast Guard and the U.S. Army Corps of Engineers have erected a series of GPS beacons along the coastal areas and internal waterways.

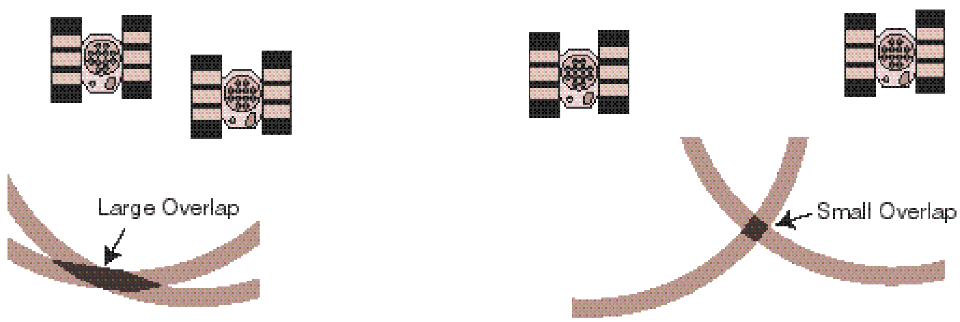

Figure 6.7 Geometric Dilution of Precision (GDOP). When the satellites are close together, the overlap between the estimated distances is larger than when the satellites are close together. More accurate GPS measurements are possible when the available satellites are further apart.

these distance measurements intersect is fairly large. This is called Geometric Dilution of Precision or GDOP (see Figure 6.7). Even if the GPS receiver does provide a latitude and longitude measurement, the error can be as much as 300 to 500 feet (~90 - 152 meters). In most cases, the particular configuration of the satellites does not lead to this much error.

A third factor in GPS accuracy are obstacles to the GPS signal such as tall buildings or mountains. These obstacles can block the signal from some or all of the satellites. No measurement will be possible if too many signals are blocked. Obstructing some of the signals may cause a loss of accuracy, depending upon the satellite geometry.

A fourth source of error is called multipath. This results when a radio signal is reflected by an object. It's the same phenomenon that causes ghost images on television reception with an antenna. Multipath errors occur with GPS when the signal bounces off a building or terrain before reaching the GPS receiver's antenna. The signal takes longer to reach the receiver than if it travelled a direct path. As a result, the GPS receiver calculates a longer distance to the satellite. Multipath errors are typically under 15 feet (~5 meters).

Further sources of error are propagation delay due to atmospheric effects and internal clock errors. GPS receivers are designed to compensate for these effects but very small positional errors still occur. Propagation delay is the 'slowing down' of the GPS signal as it passes through Earth's ionosphere and troposphere. In space, radio signals travel at the speed of light (186,000 mph), but they are significantly slower once they enter our atmosphere. Clock errors can affect the measurement of the amount of time it takes for the GPS signal to reach the receiver. All of the sources of error add up to an accuracy of 60 to 225 feet (18 to 69 meters) depending on current status of selective availability, number of satellites available, and the geometry of those satellites.

6.4.6 Differential GPS

are transmitted to the satellite and these are then made available in the data message that is sent by the satellite.

6.4.4 GPS Communications

A GPS satellite transmits in two microwave frequencies (1575.42 MHz and 1227.60 MHz). The first frequency carries the navigation message and the Standard Positioning Service (SPS) code signals. The second frequency is used to measure the delay in the signal caused by the ionosphere. This frequency can only be used by authorized users with cryptographic equipment, electronic data keys, and specially equipped receivers that use the Precise Positioning System (PPS). The United States and allied military, certain U.S. Government agencies, and selected civil users approved by the U.S. Government can use the PPS signal.

The GPS Navigation Message transmitted by the satellite is a time-tagged binary transmission. A single message or data frame consists of 1500 bits and is transmitted by each satellite every 30 seconds. The message consists of orbital and clock data. The orbital or ephemeris data describe the satellite orbit and is updated every hour.

The need for at least four simultaneous measurements affects how GPS receivers are designed. A receiver with four channels will be able to read the data from four GPS satellites simultaneously and give an instantaneous position. A receiver with fewer channels will need to sequence through satellite measurements and this can take up to 30 seconds. Newer receivers have the ability to receive more channels simultaneously.

6.4.5 GPS Accuracy

Several factors affect the accuracy of GPS measurements. Up until the year 2000, the most significant source of inaccurate data was a system known as Selective Availability (or S/A). S/A was an intentionally-imposed degradation in the accuracy of civilian GPS by the U.S. Department of Defense. S/A degraded the accuracy of GPS to a maximum of 100 meters (328 feet). The signal wasn’t usually degraded to this level but S/A induced errors of 30 meters (109 feet) were typical. The purpose of S/A was to deny hostile military or terrorist organizations the maximum accuracy benefits of GPS. Of course, the civilian portion of the signal could been deactivated very quickly in time of war. Through an executive order by the President of the United States, S/A was turned off at midnight on May 1, 2000 and this major source of GPS error was eliminated. President Clinton cited the need to make the GPS system more responsive to civil and commercial users worldwide. Additionally, the capability had been demonstrated to selectively deny GPS signals on a regional basis when US national security is threatened.

The second factor affecting GPS accuracy is satellite geometry. If the GPS satellites are all in a similar position relative to the receiver, then the satellite geometry is poor and the receiver may be unable to provide a position reading. Because the satellites in a similar position, the triangulation between them is poor and the common area where the time it takes for the signal to travel from a satellite that is directly overhead is 6/100th of a second.

6.4.3 Calculation of Latitude and Longitude

The distances to at least four satellites are used to calculate the position of the GPS receiver with what is called satellite ranging. It is essentially a form of triangulation involving trigonometric calculations. Imagine that each satellite is in the middle of a sphere. Each distance to a satellite defines the periphery of a sphere. If you are 13,000 miles from satellite "A" then you are somewhere on the surface of a sphere with a 13,000 mile radius. At the same time, you also determine that you are 15,000 miles from the second satellite. The imaginary sphere that is formed by this satellite intersects with the first 13,000 mile sphere. The distance measurement from the third satellite narrows down the position even more because the only place that three spheres can come together is in two points (see Figure 6.6). One of these two points could not possibly be correct and the receivers incorporate various techniques to determine the correct position with just three satellites. The distance to the fourth satellite can be used to distinguish between these two points and also eliminates certain errors in the GPS receiver's clock.

Each GPS receiver cannot include an atomic clock at $100,000 a piece. The use of the fourth satellite is critical in compensating for the slight time variations in the much less expensive clock in the receiver. These clocks are consistent, like a quartz watch, but they aren't synched with universal time. A variation of only a second, however, from universal time will cause a distance calculation error of 186,000 miles. With measurements from only two satellites, such an error would not be noticable. But, three erroneous distances from three separate satellites does not lead to a viable solution as the distances will not intersect at a point. GPS receivers are programmed in such a way that when they get a series of measurements to satellites that cannot intersect at a single point, they realize that something is wrong and it suspects the internal clock is not correct. The computer in the GPS receiver then begins adding and subtracting a constant amount of time from all the measurements until it determines an answer that intersects at a point. This compensation is actually done through an algebraic expression and requires distances to all four satellites. The fourth satellite is also needed to determine the elevation of the point which is also provided by the GPS receiver.

To calculate position on the ground, we not only need to know the distance to the satellites, we also need to know the position of the satellites in space. Fortunately, at nearly 11,000 miles (17,500 KM), the satellites are not affected by the earth's atmosphere and their orbits are very predictable. The orbits change so little that GPS receivers on the ground can calculate their positions from an internal "almanac." This almanac tells the receiver where each satellite is located at any given moment. The GPS satellites are monitored twice a day by ground stations that measure their altitude, position and speed. The variations in the orbit are "ephemeris" errors. The variations are very minor and are usually caused by the gravitational pulls from the moon or the sun and by the pressure of solar radiation on the satellite. The slight variations in the orbit

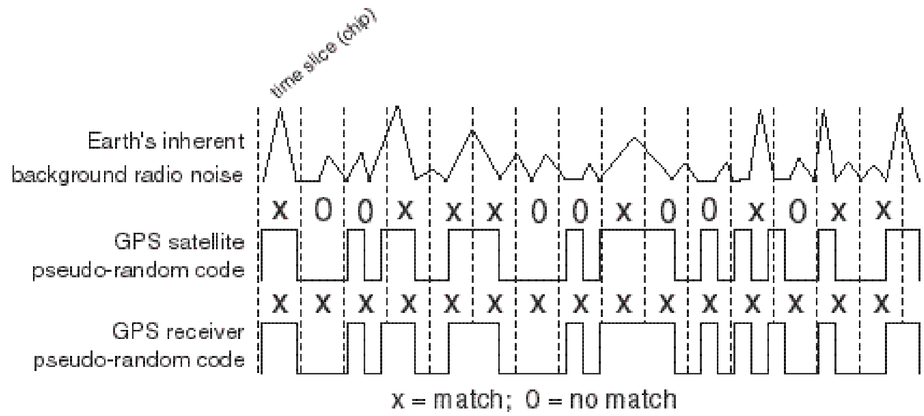

Figure 6.6 The GPS satellites transmit a very weak signal, about the same as the earth's inherent background radio noise. Both the GPS signal and the background noise are random so that when we divide the signal up into time slices or chips, the number of signal matches (X's) will equal the number of non-matches (0's). If we slide the the receiver's pseudo-random code back and forth until it lines up with the satellites, there will be more matches and we will be able to distinguish the signal from the earth's background noise.

advantages. The first is that the code is transmitted with a very weak signal using only a very small amount of power from the satellite. In fact, a GPS satellite uses less electricity than a 50 watt light bulb and the pseudo-random signal that it generates is so faint that it doesn't even register above the earth's inherent background radio noise. However, the GPS receiver can distinguish between the pseudo-random code and the background noise because it knows the pattern of the fluctuations (see Figure 6.6).

The second advantage of this system is that access to the satellites can be very easily controlled by the U.S. government. In time of war, the code can be easily changed to prevent the enemy from using the system. In fact, there are two separate forms of the pseudo-random code – the C/A and the P code. The C/A or civilian code is that one that civilian receivers use. The P code can be encrypted so that only military users are guaranteed access to it. In addition, the P code is almost impossible for a foreign government to jam. It was thought that the P code is inherently more accurate than the C/A code because it uses a more powerful and higher frequency signal. New GPS receiver designs are proving that there is practically no difference in the accuracy between the C/A and P codes.

A third benefit of the pseudo-random code is that the satellites can share the same frequency without interfering with each other. Each satellite has its own specific pseudo-random code so it is easy to identify their signal. Because all of the signals are faint, none overpowers the others.

The sending of the pseudo-random codes is controlled by very precise atomic clocks that maintain accurate time to within three nanoseconds, or three billionth of a doesn't use atomic energy but simply uses the oscillations of a particular atom to maintain time. Each clock costs about $100,000 and every satellite has four clocks to be sure that one is always operational. This precision is important because the signal travels so fast and slight errors in the timing can cause huge errors in the distance. For example, corrections in the orbits of the satellites are small because their high altitude ensures that their orbits are very stable and precise.

Unmanned ground stations that track and monitor the GPS satellites are located around the world – Hawaii and Kwajelein islands in the Pacific Ocean; Diego Garcia in the Indian Ocean; and the Ascension Island in the Atlantic Ocean. The master ground control station at Schriever Air Force Base (formerly Falcon Air Force Base) in Colorado Springs, Colorado, consists of four large ground antenna stations that broadcast signals to the satellites.

To determine latitude and longitude, the distance between a GPS receiver on the ground to at least four orbiting satellites is calculated. The system works by timing how long it takes a radio signal transmitted from the satellite to reach a GPS receiver. The speed of the radio signal is equal to the speed of light and is precisely known (186,000 miles per second). The receivers multiply the speed of light by the travel time to get distance.

6.4.2 The GPS Pseudo-Random Code

To determine the amount of time it takes for the signal to reach the receiver, the satellite and the receiver generate the same pseudo-random code at precisely the same time. This code is a very complicated set of digital codes that are repeated every millisecond. The code may be compared to a series of numbers that appear random but are actually carefully chosen. The codes are made purposely complicated so that they can be easily and unambiguously compared. Because the satellite and receiver synchronize the sending of these codes, the receiver can measure the time difference between when it started a particular pseudo-random code and when the same code is received from the satellite (see Figure 6.5). This time difference is multiplied by the speed of light to calculate the distance to the satellite.

The use of the pseudo-random code to determine distance has a number of

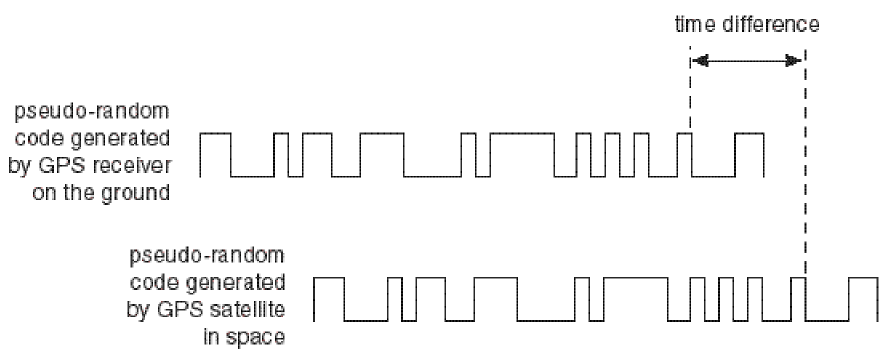

Figure 6.5 The GPS receiver and satellite generate the same pseudo-random code at exactly the same time. When the code arrives from the satellite, the time difference is compared to the same code generated by the receiver. This difference is multiplied by the speed of light (186,000 miles per second) to determine the distance to the satellite.

Figure 6.4 Examples of hand-held GPS receivers with map display. These devices display the current position on an electronic map. About 6 inches (15.2 cm) in length and weighing less than a pound, the portable devices are used for a variety of applications.

system (see Figure 6.3). Additional satellites are launched as needed to replace those that can no longer be maintained in orbit. The lifespan of each satellite is 7.5 years.

GPS spurred a multi-billion dollar business in the 1990's as a certain portion of the GPS signal was made available for civilian purposes. A variety of GPS receivers were made available for different applications (see Figure 6.4). Although the GPS signal was initially not as accurate as that used by the military and could only give position to within about 60 meters, GPS receivers became used for a variety of applications. Almost immediately, a system was devised using two GPS receivers called differential GPS that provided accuracies of under one meter, thereby effectively by-passing the military restrictions. Differential GPS is now performed routinely with a series of base stations that have been erected throughout the world. Radio signals that can be broadcast by these base stations provide the needed information to perform the differential GPS calculations.

6.4.1 GPS System Characteristics

GPS consists of a constellation of at least 24 satellites – 21 primary satellites and 3 orbiting spares – that orbit the earth at an altitude of 17,500 KM (10,900 miles) at a speed of 1.9 miles per second. They orbit the earth between 60°N and 60°S latitude. Each satellite weighs almost a ton (1900 lbs) and is 17 feet (5.81 meters) wide with solar panels extended. The satellites orbit the earth twice a day and this, along with the number of satellites, guarantees that signals from six of the satellites can be received from any point on earth at almost any time.

The orbit puts each satellite over a ground control station at least once a day where their position is precisely measured. Any discrepancy found in a satellite's position is transmitted to the satellite and this data is transmitted to GPS receivers. These information.

A system called Loran-C was developed to provide radionavigation service for U.S. coastal waters. It was later expanded to include complete coverage of the continental U.S. as well as most of Alaska. Twenty-four U.S. Loran-C stations work in partnership with Canadian and Russian stations to provide coverage in Canadian waters and in the Bering Sea. Loran-C provides better than 0.25 nautical mile absolute accuracy for suitably equipped users. It is possible to navigate to a previously determined position with an accuracy of 50 meters or better using Loran-C in the so-called time difference repeatable mode. Advances in technology have allowed greater automation of Loran-C operations. Plans to terminate Loran-C operations on December 31, 2000 were announced but, as a result of political pressure, the system will likely continue in operation through 2008.

6.4 PRINCIPLES OF THE GLOBAL POSITIONING SYSTEM

The Global Positioning System (GPS) was designed for military applications. Its primary purpose was to allow soldiers to keep track of their position and to assist in guiding weapons to their targets. The satellites were built by Rockwell International and were launched by the U.S. Air Force. The entire system is funded by the U.S. government and controlled by the U.S. Department of Defense. The total cost for initial implementation the system was over $12 billion.

The first GPS satellite was launched in 1978. Testing took place on the initial satellites in the following decade. From 1989 to 1993, 23 production satellites were put into orbit. The launch of the 24th satellite in 1994 completed the operational GPS

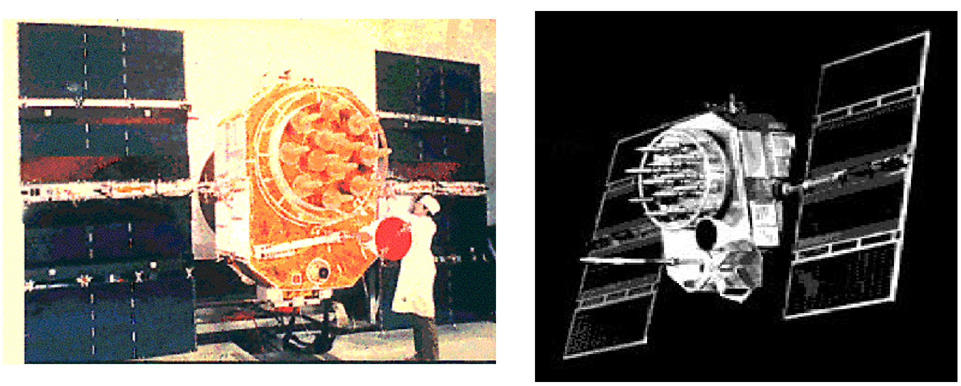

Figure 6.3 A GPS satellite. The GPS constellation of satellites consists of at least 24 satellites – 21 primary satellites and 3 orbiting spares. They orbit the earth at an altitude of 17,500 KM (10,900 miles) at a speed of 1.9 miles per second between 60°N and 60°S latitude. Each satellite weighs 1900 lbs and is 17 feet (5.81 meters) wide with solar panels extended. The satellites orbit the earth twice a day. This guarantees that signals from six of the satellites can be received from any point on earth at almost any time.

development of the H3 was actually used in making the H4. It was tested on a Royal Navy ship to the West Indies and it was shown to have an error of only 1 1/4 minutes. A 30 minute accuracy was sufficient for the prize. The captain of the ship gave his quadrant to Harrison's son and said that he wouldn't have much use for it anymore after using his father's chronometer. The Longitude Board, however, did not believe the results and commissioned another test, this one to Babados. The H4 passed again.

Twelve years went by before Harrison would be awarded the £20,000 prize. The newly appointed Astronomer Royal at Greenwich, Neville Maskelyne, who wanted to claim the prize based on his own astronomical work, first forced Harrison to disassemble his clock in front of other watchmakers to prove it was not some kind of "trick." Then, another clockmaker was given the designs and ask to reproduce the clock – a task that took two years. Harrison was then ordered to produce two copies of the H4 without having either the original clock or the designs.

In 1772, Harrison's first copy was finished and, at the age of 76, he was incapable of producing a second. He appealed to King George III for help. The King ordered the Board of Longitude to meet and explain why the longitude prize was not awarded to Harrison. The Board of Longitude finally awarded him the £20,000 but withheld bestowing the actual Longitude Prize. The terms for the prize were reset in such a way that it was impossible for either astronomers or clockmakers to meet the requirements and it was eventually withdrawn fifty years later, still unclaimed.

Before Harrison died in 1776, Maskelyne published Harrison's clock design effectively putting his ideas in the hands of other clockmakers who quickly seized the opportunity to become wealthy from the new maritime chronometer that Harrison had spent his life developing. Navigation by chronometer was used for the next 150 years.

6.3.3 Radio-based Navigation

Radio-based navigation was introduced in the early part of the 20th century. The LOng-Range, Accurate radio Navigation (LORAN) system was used by ships and aircraft by the middle part of the 1900s. LORAN, operating in the 1,700 kHz range, measures the time-of-arrival difference between two signals transmitted by two different ground stations. The pulse from the first station, called the master, triggers the second station, called the slave, into transmitting a similar pulse after a set time delay. A chart is then consulted that contains a series of hyperbolic curves that define constant time differences between particular station pairs. The position of the ship or airplane will be somewhere along the curve that corresponds to the measured time difference. By taking another reading from a second pair of stations whose curves intersect those of the first pair, a definite geographic location can be established.

A drawback of using radio waves generated on the ground is that you must choose between a system that is very accurate but doesn't cover a wide area, or one that covers a wide area but is not very accurate. High frequency radio waves (like UHF TV) can provide an accurate position but can only be picked up over a small area. Lower frequency radio waves (like AM radio) can cover a larger area but give poor positional a way to determine a ship's longitude within 30 nautical miles. It was left to a cabinet maker by the name of Harrison, not the astronomers at the Royal Observatory, to solve the problem of determining longitude.

6.3.2 Harrison and the Invention of the Maritime Chronometer

John Harrison was born in Yorkshire, England, in 1693, the son of a woodworker. He was apprenticed as a carpenter and cabinetmaker and had little formal education. But, he showed an aptitude for mechanical things and developed an unconventional means of solving problems. Working essentially alone on his development of the chronometer, he challenged the scientific and academic establishment and through great effort and determination, finally triumphed in winning the £20,000 prize and worldwide acclaim.

Harrison built his first clock at the age of 20, entirely out of wood. He was commissioned to build several clocks thereafter, one of which – the so-called Pelham clock, also built of wood – was still working 270 years later. In 1730 he went to London to convince the Board of Longitude that he could build a clock that would be accurate enough to determine longitude. The Board was filled with astronomers and mathematicians who would not have been sympathetic to Harrison. He was never allowed to present his ideas to the Board but one of its members, Sir Edmund Halley, known for Halley's comet, was interested enough in Harrison's ideas that he referred him to the eminent clockmaker, George Graham. Graham was impressed by Harrison and his proposed design and became his benefactor, loaning him money to be repaid "at no great haste, and at no interest."

Harrison completed the H1 in 1735, the first of five maritime chronometers he would build. Made of brass, it weighed 70 pounds and was nearly 4 feet high, wide and deep. The clock was tested at sea, and though it met the requirements for the prize, Harrison believed he could build a better clock and merely asked for £500 to continue his work. Harrison was still not satisfied with the H2, completed two years later, and again merely asked for more money to continue work on an improved design.

It was 17 years before the H3 was completed. It incorporated many new inventions and was half the size of the H1 and 15 pounds lighter. Satisfied with his accomplishment, and now 60 years of age, Harrison believed the H3 deserved the Longitude Prize. However, in the intervening years, another device called the Quadrant had been developed that made navigation by the stars more practical, if enough planets and stars could be accurately plotted. The members of the Board of Longitude were now actively engaged in finding the astronomical solution to navigation and thereby claiming the prize themselves. They were not interested in letting Harrison prove the utility of his chronometer. No ship could be found to carry it until three years later. By then, Harrison had completed the H4.

The H4 was completely different than all previous designs. It looked like a greatly enlarged pocket watch, around six inches in diameter. Completely enclosed, it was perfect for seagoing navigation. But, how did Harrison complete such a radical new design in only three years? It is presumed that much of the 17 years used in the

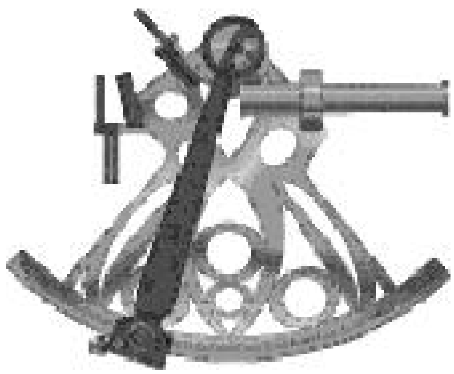

Figure 6.2 The sextant consists of a triangular frame, the bottom of which is a graduated arc of 60°. A telescope is attached horizontally to the plane of the frame. A small mirror is mounted perpendicular to the frame at the top of an index bar that swings along the arc. A half transparent, half-mirror called the horizon glass in front of the telescope reflects the image of the sun or other celestial body from the index mirror to the mirror half of the horizon glass into the telescope. When the horizon is seen through the transparent half of the horizon glass, with the reflected image of the celestial body lined up with it, the altitude of the sun or star can be read from the index arm of the arc. The latitude can then be determined with reference to navigational tables. A sextant may be used on land with an "artificial horizon" – a small, shallow receptacle containing mercury that gives a truly horizontal surface. In aerial navigation a bubble sextant is used in which a spirit level is reflected so that the center of the bubble indicates the true horizon.

Determining longitude still remained elusive. In 1675, the Royal Observatory was established in Greenwich, England, to help solve the problem of finding longitude. John Flamsteed was appointed its first astronomer and the 28-year old clergyman was instructed to "apply himself with the most exact care and diligence to rectifying the tables of the motions of the heavens, and the places of the fixed stars, so as to find out the so much-desired longitude of places for the perfecting the art of navigation." Flamsteed and astronomers that followed him at Greenwich measured the position of stars using transit telescopes. These devices were aligned in a north-south direction and were used to record the position of the stars at various times along this north-south transit line. The intent was to provide sailors with the star positions as they were viewed at Greenwich so that their position could be compared to their location in the sky as viewed from other parts of the world. Meticulous readings were taken at the observatory but initial astronomers did not want to release their data until the readings could be repeatedly verified.

Once the system of navigation by the stars was introduced it proved to be both extremely complicated and cumbersome, as there was no instrument that could make the necessary planetary sightings and measurements on board ships. The first "Nautical Almanac and Astronomical Ephemeris" was nevertheless published in 1767 that listed the position of celestial bodies for each day of the year. The ephemeris data is still being published as the Astronomical Almanac, a joint American and British publication, and contains such information as the daily ascension and declination of the sun, moon, planets, and other celestial bodies.

In 1707, more than 30 years after the Royal Observatory had been established, the British suffered a disastrous loss of a fleet of ships off the Scillies that resulted in the death of nearly 2000. Dissatisfied with the progress of the observatory in finding a practical method to determine longitude, the British formed a Board of Longitude that, in 1714, offered £20,000 (equivalent to 1 million dollars today) to anybody who could find to be a major impediment to all forms of celestial navigation.

6.3.1 Instruments of Early Navigation

The compass, consisting of a magnetic needle freely suspended so that it aligns with the earth's magnetic north and south, has been used for navigation since at least the 12th century. With the help of a map, one could reach a destination by laying a course in a certain direction. A dead-reckoning system of navigation developed that consisted of calculating the point of departure, the course as shown by a compass, the speed and distance travelled and the time elapsed.

The compass became so popular for navigation that maps were made that incorporated compass directional lines. As discussed in chapter 3, Portolan and Catalan charts, were used from approximately 1100 to 1600 AD. The problem with these charts is that they did not include lines of latitude and longitude nor was there any indication that the earth was a sphere. Indeed, the makers of these maps did know that this was the case. While the maps could be used with a compass, the lines of constant direction that they depicted – called rhumb lines – did not follow the great circle and therefore did not represent the shortest distance between the points. Moreover, the compass itself did not help the sailors determine their position enroute but only helped to follow a course between two points.

The cross-staff was used to find latitude by the early 1400s. It was constructed from two pieces of wood with the cross at right angle to, and sliding on, the staff. A sight was fixed at the end of the staff. Holes were bored at the ends of the 26 inch (66 cm) cross. The instrument was sighted in the direction of a heavenly body until the star appeared through the upper hole and the horizon through the lower. The altitude was then read on a scale marked on the staff. The device was not very accurate. Columbus navigated with a cross-staff in addition to a compass and a table of the sun's declination on his voyages to the new world in the 1490s.

The astrolabe was also used to determine latitude. It consisted of a disk of wood or metal with a circumference marked off in degrees. Pivoted at the center of the disk was a pointer called an alidade. Angular distances could be determined by sighting with the alidade and taking readings of its position on the graduated circle. More elaborate astrolabes included a star map and a zodiacal circle. Calculations of longitude were also attempted with the device. The device was used on voyages of discovery beginning in the 1500s and was still used with the advent of the sextant in the 1700s. Like the cross-staff, it was also not very precise.

The sextant used adjustable mirrors to measure the exact angle of the stars, moon or sun above the horizon. Invented independently in England and America in 1731, it is based on the principle that a reflected ray of light leaves a plane surface at the same angle at which it strikes the plane. The horizon and the reflected image of a celestial body such as the sun are aligned on the "index arm" and the altitude is determined from the reflection on this arm (see Figure 6.2). Latitude is then determined with reference to navigational tables.

Canary Islands which marked the western boundary of the known world in the 2nd century A.D. Until the 1870s, each country defined its own zero meridian. Portugal defined its zero meridian at Lisbon, Spain used its coastal city of Cadiz, Norway used Christiana, Sweden used Stockholm, Russia used Pulkowa, Germany used Ferro, and France used Paris. In 1871, at the first International Geographical Congress (IGC) in Antwerp, it was recommended that whenever ships exchanged longitudes at sea, they should be based on a prime meridian that passed through an observatory in Greenwich, England. Another meeting in Rome in 1875 failed to resolve the standardization of the prime meridian at Greenwich. France tied their acceptance of Greenwich to the British acceptance of the metric system. Finally, at the International Meridian Conference in 1884 in Washington D.C., twenty-five countries voted to adopt the meridian at Greenwich as the prime meridian for the world. The precise location of the prime meridian was defined by the cross-hairs in the eyepiece of the large “Transit Circle” telescope built by the Royal Astronomer George Biddell Airy. It was agreed that longitude would be measured in two directions from this line: “east longitude being plus and west longitude being minus.”

Greenwich was chosen as the prime meridian for two main reasons. The first was that the United States had already chosen Greenwich as the basis for its own national time zone system. The second was that by the late 1800s, nearly 3/4 of the world’s commerce used sea charts that defined Greenwich as the prime meridian. The argument was that the acceptance of this prime meridian would benefit the largest number of people.

The system of latitude and longitude is only one of many coordinate systems that is used to define location on the earth. The other coordinate systems are used for smaller areas and consider the earth to be flat. Although this causes some error, especially further away from the origin, the error is so slight that these coordinate systems are still useful for many applications. These flat coordinates systems, such as the Universal Transverse Mercator (UTM) and State Plane Coordinates, were used because latitude and longitude were difficult to determine. With the advent of GPS, more measurements of location are now being made with latitude and longitude.

6.3 NAVIGATION

People have been trying to find a reliable way to navigate between places since prehistoric times. The term was once only used for navigation by sea. Navigation now refers to any type of movement, whether by sea, land or air. Initial forms of navigation probably relied on stones and twigs as landmarks to mark a trail. Early travel by boat stayed well within view of the coastline so as to avoid getting lost at sea. Stars were used relatively early as an aid to navigation. The Phoenicians used the north star for the journey between Egypt and Crete. In Homer, the goddess Athena tells Odysseus to “keep the Great Bear on his left” in reference to a constellation of stars in the journey to Calypso’s Island. Unfortunately, stars are only visible on clear nights. Clouds continue getting lost in the wilderness. GPS represents a revolution in how position is determined. The significance of this system can only be understood by examining how location is defined and how it was calculated before the advent of this technology.

6.2 LOCATION

A location can be given in relative or absolute terms. An example of a relative description of place is: "The house next to mine." We use this type of description in normal conversation because it can be easily understood. If the description isn't understood, we choose a different landmark until the person makes a connection with a particular feature.

Absolute location is always defined based on the origin of a coordinate system. An example would be the numbered streets of a city arranged a grid format. With this type of street layout it is possible to define an absolute position relative to an origin. For example, "42nd Street and 5th Avenue" would refer to a specific intersection that is a certain number of blocks away from 1st street and 1st avenue.

Similarly, cartesian coordinates are used to define location on a flat surface. Location is given with x and y coordinates. The horizontal distance is given with x and the horizontal with y. In the spherical coordinate system that is used for the earth, latitude can be related to the y coordinates and longitude corresponds to the x (see Figure 6.1). Lines of latitude and longitude are also referred to as parallels and meridians. The origin of this spherical coordinate system is located at the intersection of the prime meridian and the equator.

The equator is a natural origin for lines of latitude but no similar line of origin exists for longitude. The Greek cartographer Ptolemy used a zero meridian through the

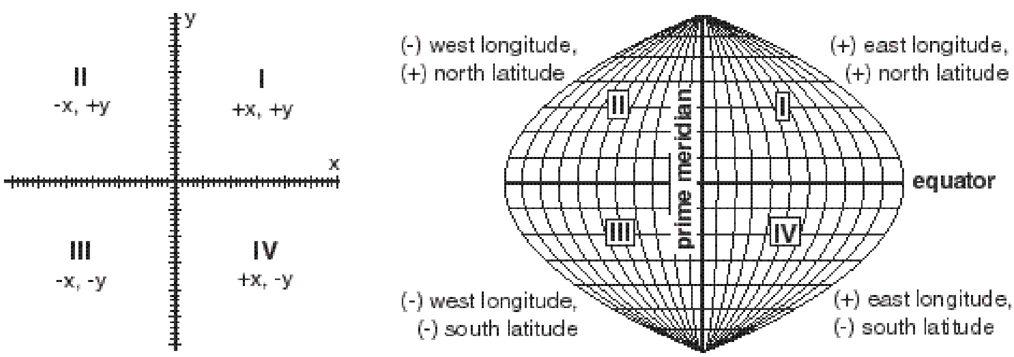

Figure 6.1 The cartesian coordinate system uses x and y coordinates to locate points on a plane. The four quadrants specify the position of points in plus and minus directions from an origin. The spherical coordinate system of latitude and longitude has its origin at the intersection of the equator and the prime meridian. It is also divided into four sectors with each being either north , south, east, or west of the origin. The x coordinates can be compared to longitude and y coordinates to longitude.

6

Global Positioning Systems

You can't get to where you want to go if you don't know where you are.
– Anonymous

6.1 INTRODUCTION

A major struggle for humans has been to know precisely where they are located on the surface of the earth. This was particularly important as people began to move large distances across the earth's surface. Sailors, for example, needed to know where they were on the open seas and how much further and in which direction they had to travel. Determining latitude was already possible at least 600 years ago with the cross-staff that measured the altitude of a celestial body such as the north star. Longitude was much more difficult to determine and was not made feasible until the mid-1700s with the advent of accurate time-keeping devices called chronometers. Before this, navigation on land or sea was a hazardous undertaking.

With the help of a small hand-held Global Positioning System (GPS) receiver for about US$100, the latitude and longitude of a point can now be determined within about 3 meters. Greater accuracy, less than a meter, can be achieved by using two GPS receivers simultaneously, one with a known position, and comparing the results between them. This technology, based on a series of earth orbiting satellites operated by the US Department of Defense, is revolutionizing navigation and the way that information is acquired for maps.

Global positioning systems are being used for a wide variety of applications. The devices help drivers navigate by displaying the current position of their vehicle on an electronic map. Trucking companies have used GPS for many years to monitor the location of their entire fleet. Small hand-held receivers are used by hikers to avoid