Global Navigation Satellite Systems: Genesis, State of the Art, and Future Directions

Mark L. Psiaki*

206 Upson Hall, Cornell University, Ithaca, N.Y. 14853-7501

Abstract

The Global Positioning System (GPS) has risen from a mere paper design in the early 1970s to become a global utility of the 21st century. The system is used for diverse applications that include navigation of ground vehicles, ships, aircraft, and spacecraft, monitoring of millimeter-level shifts of the Earth's tectonic plates, and the precise timing of financial transactions and electrical power transmission. As a testimony to the system's success, the European Union is developing the competing Galileo system, partly to avoid reliance on the U.S. military, which runs the GPS, and partly as a means of increasing its market share in the rapidly growing ~ $10B/year Positioning, Navigation, and Time industry. The introduction of Galileo has forced a change of the system acronym from GPS to GNSS (Global Navigation Satellite Systems) in order to denote both systems without giving precedence to either.

This paper focuses on 3 aspects of GNSS: the historical development of the GPS, current applications/research, and future directions. A selection of current applications and research topics illustrates the system's range of capabilities and can inspire the conception and development of new applications. Future directions will be driven by the development of competing and augmenting systems. One of the main future issues is the development of Galileo and its prospects for competition/cooperation with the GPS.

I. Introduction

The Global Positioning System (GPS) is being used in a variety of important and unusual applications. One example is the Joint Precision Approach and Landing System (JPALS) that is being developed as a joint program of the U.S. Army, Navy, and Air Force. One JPALS requirement is to enable automatic landings on an aircraft carrier in zero visibility conditions. In order to do this, the system must reliably maintain a relative position accuracy of 0.3 meters between the moving ship and the approaching aircraft. This level of accuracy can be achieved and even exceeded by using GPS in an ultra-precise mode of operation known as Carrier-Phase Differential GPS (CDGPS), possibly with aiding from an inertial measurement unit 1. Other applications of CDGPS achieve cm-level accuracy for real-time guidance of farm machinery that automatically cultivates fields 2 or mm-level accuracy in post-processed analysis of shifts of the Earth's crust 3.

New and unusual applications of GPS are being proposed and implemented on a regular basis. GPS monitors the timing of signal phases in electrical power generation and distribution grids. Its micro-second accuracy helps control algorithms to avert system failures 4, and similar GPS-based timing techniques are regularly used to secure large electronic financial transactions. Novel applications include the tracking of ski jumpers to help in their training 5 and the processing of reflected GPS signals as a means of monitoring the terrain below an aircraft 6. The GPS has even been used to solve a murder: The guilty party left the crime scene in a rented Lincoln Navigator automobile that was GPS-equipped. It recorded the

* Associate Professor, Sibley School of Mechanical and Aerospace Engineering.
exact time and place of his car when he started it, and that information placed him at the crime scene minutes before firefighters discovered his victim inside a burning apartment.

The basic function of the GPS is to provide position determination, navigation, and time synchronization (PNT) services on a global scale. A stand-alone civilian receiver can determine its absolute position -- latitude, longitude, and altitude -- anywhere on the Earth to an accuracy of about 10 m. The system even functions in Low Earth Orbit (LEO) up to an altitude of about 3000 km. It can determine its absolute time to an accuracy of about 30 nsec and its velocity to an accuracy of about 0.01 m/sec. The PNT errors of stand-alone military receivers are about half as large. Additional advantages of military receivers are their ability to avoid being spoofed and their lower susceptibility to jamming. The receiver must have access to signals from 4 or more GPS satellites in order to function normally. If the receiver has a clear sky view and if its antenna is properly oriented, then there are enough GPS satellites in the constellation to provide access to 5 or more signals on a global scale. The requirement to receive 4 or more signals presents a problem only if the lines of sight to the GPS satellites are obstructed. Typical obstructions include tall buildings in a city, mountains around a narrow valley, or thick foliage above the floor or a jungle or a dense forest.

There are several enhanced versions of the system that yield improved accuracy through differential corrections. The basic idea of differential corrections is to monitor and calibrate error sources using a receiver or a set of receivers at known surveyed locations and to transmit the error corrections to mobile receivers using a radio data link. Differential correction systems include the Wide-Area Augmentation System (WAAS) over the U.S.A. and the European Geostationary Navigation Overlay System (EGNOS) over Europe. Both systems transmit their correction signals from geostationary satellites on the GPS L1 frequency, \( f_{L1} = 1575.42 \text{ MHz} \). A GPS receiver can be modified to process these correction signals with only minimal changes to its hardware. A GPS receiver that uses WAAS or EGNOS corrections can deliver absolute accuracies of about 2 m in the zones where the corrections apply.

Another type of differential correction scheme, known as CDGPS, uses measurements of the phase of the carrier signal from several receivers in order to determine the relative location of one receiver with respect to another. A CDGPS system can have relative position accuracies on the order of 1 cm or better over baselines up to 10 km or more. CDGPS can operate in real-time if there is a dedicated radio link for passing the carrier phase data between the two receivers. This form of CDGPS is called the Real-Time Kinematic (RTK) mode of operation.

There are many common uses of GPS today in addition to those already mentioned. The basic system is used for navigation of ships and aircraft. It is also popular in automobiles and for personal navigation in the wilderness when hiking, hunting, or fishing. Many cell phones offer GPS as an additional service, though this service can fail indoors or in cities. CDGPS systems with cm-level accuracy have become standard technology for surveying and for site preparation on construction jobs. CDGPS can be used to do attitude determination by mounting an array of antennas on a vehicle. GPS is gaining a foothold in space where it can be used for absolute orbit determination of a single LEO spacecraft. Orbital accuracies on the order of several cm can be achieved if the data is post-processed in a batch filter.

GPS is used for remote sensing of the Earth's ionosphere and atmosphere. These applications prove the following proverb: "One man's noise is another man's signal." The ionosphere and the troposphere cause delays in the GPS ranging signal, and they affect the phase of its carrier wave. The disturbed ionosphere can cause rapid fluctuations of a GPS signal's power and carrier phase in a phenomena known as equatorial spread-F (also known as ionospheric scintillations). These various effects can be measured by a GPS receiver that has been designed appropriately. Its output data can be used to study the physics of the ionosphere.
and the troposphere and to monitor the space weather that occurs in the ionosphere. A common remote-sensing application of GPS is to infer the ionosphere’s electron density profile, as in Fig 1.

The GPS is funded and controlled through the U.S. military, and it has many military applications. A number of these applications are similar to civilian applications, e.g., position determination and navigation for infantry, land vehicles, ships, aircraft, and spacecraft. A unique military application is the precision delivery of weapons. These include GPS-equipped bombs and artillery shells. The latter application is particularly challenging because the receiver must survive the high-g launch and rapidly acquire GPS satellites at the beginning of its short trajectory. Figure 2 illustrates a typical benefit of munitions delivery from a GPS-equipped aircraft: the dispersion around the target is much smaller than what is achievable with a radar-based system.

The GPS is important economically. The user equipment market in the U.S. accounted for about $5 billion in sales in 2005, and that figure is expected to double in 5 years. About 92% of this market is for civilian equipment, and the remaining 8% is for military equipment. The world market may be 3 to 4 times larger than this. One optimistic prediction is that worldwide GNSS sales just for cell phones and civilian vehicle navigation will be $190 Billion by 2020.

The technical and economic success of the GPS has inspired the European Union to begin the development of a similar system that is called Galileo. The Russians have a system too, called GLONASS, but it is not fully functional and has not been well maintained. Galileo, on the other hand, is expected to equal or exceed the performance of the GPS and to be interoperable so that a given receiver can improve its performance by using signals from both satellite systems. The new acronym GNSS has been coined in order to specifically refer to the combined GPS/Galileo system, and the premier annual conference on this subject has changed its name from the "ION (Institute of Navigation) GPS" conference to the "ION GNSS" conference.

The present paper reviews the operation and history of the GNSS and discusses current important issues and projected future developments. Section II contains a overview of the operation of the GPS. Section III
reviews its history. Section IV discusses technical developments that are currently of interest
to civilian applications of the GPS, and Section V deals with important issues for the military
use of GPS. Section VI discusses the future of GNSS, and this paper's conclusions are
presented in Section VII.

II. Overview of the GPS

A. Components, Concepts, and Operation of the Basic System

The GPS consists of three segments, the space segment, the ground control segment, and
the user segment, as depicted in Fig. 3. The space segment, shown in Fig. 4, consists of a
constellation of 29 satellites in 6 orbital planes that are inclined by 55 deg and that are equally
spaced in longitude. There are nominally 4 satellites per plane, but orbiting spares bring the
total number of satellites to 29. They orbit at about 20,000 km altitude above the earth with
periods of 12 hours. Each satellite broadcasts ranging signals that are received by the user
segment. The ground control segment operates a set of tracking stations around the Earth that
are used to determine the precise orbits of the GPS spacecraft along with calibration
corrections to their precise atomic clocks. The ground control segment uploads these
corrections to the GPS satellites so that they can broadcast this information to the user
segment. The ground control segment also performs normal satellite "housekeeping" functions
such as commanding orbital correction maneuvers and monitoring spacecraft
health. The user segment consists of radio receivers that passively receive signals directly
from the space segment. These signals provide the necessary information for the user
receivers to accurately determine their position, velocity, and time. The space segment and the ground control segment are paid for, owned,
and operated by the U.S. military. A user segment receiver is available to virtually anyone
who can pay the $100 or more that it costs to buy a civilian GPS receiver.

The satellites of the space segment broadcast signals on the same two nominal carrier
frequencies, and these signals are encoded with unique pseudo-random number (PRN) codes
that allow each user receiver to distinguish between different satellites, in a code-division
multiple-access (CDMA) spread-spectrum signaling scheme that works like some cell phone
systems. The two nominal broadcast frequencies are \( f_{L_1} = 1575.42 \text{ MHz} \) and \( f_{L_2} = 1227.6 \text{ MHz} \). Each PRN code is broadcast as a binary phase-shift keyed (BPSK) pseudo-random
The transmitted civilian signal on the $f_{L1}$ frequency takes the form:

$$y_{trns}(t) = AC_{PRN}(t)D(t)\cos[2\pi f_{L1}t + \theta_0]$$  \hspace{1cm} (1)$$

where $y_{trns}(t)$ is the transmitted signal at time $t$, $A$ is its amplitude, $C_{PRN}(t)$ is the PRN code that takes on $+1/-1$ values, $D(t)$ is a navigation data message that is encoded as $+1/-1$ bit values, and $\theta_0$ is the initial carrier phase. The $+1/-1$ values of the $C_{PRN}(t)$ PRN code are called chips. They undergo pseudo-random chip transitions at a nominal frequency of 1.023 MHz, and the pseudo-random code repeats itself every 1023 chips, that is, every 1 msec. The data bits in $D(t)$ are transmitted at a slow 50 Hz rate, and the system requires 30 seconds to transmit a full 1500-bit navigation data message, with a single data bit being transmitted once every 20 periods of the PRN code.

The GPS uses the PRN codes to make absolute range measurements. Neglecting the effect of various error sources, the received navigation signal is

$$y_{rcvd}(t) = A_{rcvd}C_{PRN}(t-\rho/c)D(t-\rho/c)\cos[2\pi f_{L1}(t-\rho/c) + \theta_0]$$  \hspace{1cm} (2)$$

where $A_{rcvd}$ is the received amplitude, $c$ is the speed of light in vacuum, and $\rho$ is the distance from the point of transmission to the point of reception. The receiver uses signal processing techniques to measure the time of reception of a particular feature of the PRN code. Referring to Fig. 5, which illustrates sections of a typical transmitted and received PRN code, suppose that the feature in question is the last $+1$ to $-1$ transition on the figure, which was transmitted at time $t_k$. The receiver measures the reception time of this feature, $t_k + \rho/c$. The transmission time $t_k$ is known to the receiver because the navigation data stream in $D(t)$ can be used to compute the transmission time of any PRN code feature. The receiver subtracts the computed transmission time from the measured reception time and multiplies the result by $c$ in order to derive its absolute range measurement.

A GPS measurement of the range $\rho$ is called a pseudorange because it includes the effects of receiver clock error.

A GPS receiver determines its clock correction and position by solving the following system of pseudorange measurement equations:

$$p^i = \sqrt{(r^i - r_{user})^T (r^i - r_{user}) + c\delta t + v^i} \quad \text{for } i = 1,\ldots, n_{sats}$$  \hspace{1cm} (3)$$

for the unknown 3-dimensional user position vector $r_{user}$ and the unknown user receiver clock correction $\delta t$. The quantity $p^i$ is the measured pseudorange for the $i^{th}$ GPS satellite signal. The number of available satellite signals is $n_{sats}$. The position vector of the $i^{th}$ GPS satellite is $r^i$, and it is determined from orbital information for the satellite that is contained in the
of the user receiver relative to the GPS satellites. The errors in \( r' \) result from errors in the satellite ephemerides that are broadcast in the \( D(t) \) navigation data stream, and these errors can be as large as 1-2 m. A related error occurs in the computed transmission time \( t_k \), and it can have a similar magnitude; this error can be modeled as part of \( \nu^i \). These errors are caused by the ground control segment’s imprecise determination of the GPS satellites’ orbits and atomic clock calibration parameters. These errors are slowly varying bias terms. The ionosphere and the troposphere contribute error components to \( \nu^i \) by delaying the signal so that it does not travel at the speed \( c \) when it traverses these regions. These errors can be on the order of 4 and 0.5 m RMS, respectively. Additional error sources are multipath and receiver thermal noise. Multipath refers to reflected signals that get processed by the receiver along with the direct line-of-sight signal. These

Fig. 5. Illustration of transmitted and received \( C_{PRN}(t) \) pseudo-random number codes and their use to measure transmission delay (Note: contrary to this figure, transmission delays are normally much longer than a single PRN code chip interval).

navigation data message \( D(t) \). The quantity \( \nu^i \) is the pseudorange measurement error. The nonlinear least-squares solution procedure determines the values \( r_{user} \) and \( \delta t \) that minimize the sum of the square errors in these equations, i.e., that minimize the least-squares cost function

\[
J(r_{user}, \delta t) = (\nu^1)^2 + \ldots + (\nu^{n_{sats}})^2.
\]

\( n_{sats} \geq 4 \) is required in order for this least-squares problem to have a unique optimal solution. The geometry of a typical GPS navigation problem is illustrated (not to scale) in Fig. 6.

The accuracy of the solution is impacted by several types of measurement errors that are present in \( r' \) and in \( \nu^i \), and it depends on the geometry of the user receiver relative to the GPS satellites. The errors in \( r' \) result from errors in the satellite ephemerides that are broadcast in the \( D(t) \) navigation data stream, and these errors can be as large as 1-2 m. A related error occurs in the computed transmission time \( t_k \), and it can have a similar magnitude; this error can be modeled as part of \( \nu^i \). These errors are caused by the ground control segment’s imprecise determination of the GPS satellites’ orbits and atomic clock calibration parameters. These errors are slowly varying bias terms. The ionosphere and the troposphere contribute error components to \( \nu^i \) by delaying the signal so that it does not travel at the speed \( c \) when it traverses these regions. These errors can be on the order of 4 and 0.5 m RMS, respectively. Additional error sources are multipath and receiver thermal noise. Multipath refers to reflected signals that get processed by the receiver along with the direct line-of-sight signal. These

\[
\sqrt{(r' - r_{user})^T(r' - r_{user})} = p' - c \delta t
\]

\[
\sqrt{(r^2 - r_{user})^T(r^2 - r_{user})} = p^2 - c \delta t
\]

\[
\sqrt{(r^4 - r_{user})^T(r^4 - r_{user})} = p^4 - c \delta t
\]

Fig. 6. Geometry and algebra of basic GPS position and time determination problem.
signals typically perturb the pseudorange measurement by 1 m RMS, but multipath errors of 15 m or larger can occur in environments with large reflective surfaces near the receiving antenna\textsuperscript{14}. Thermal noise errors are normally on the order of 0.5 m RMS\textsuperscript{14}.

The relative geometry of the GPS satellites and the user receiver affects the impact of individual pseudorange errors on the accuracy of $r_{user}$ and $\delta t$. A good geometry has the satellites well distributed over the sky, with several at high elevation and a number at medium to low elevations with a range of azimuths. If all of the satellites are concentrated in one area of the sky, then the geometry is bad. The effect of the geometry is characterized by the parameters known as the Geometric Dilution of Precision (GDOP) and the Position Dilution of Precision (PDOP). GDOP gives the mean error magnitude of the 4-dimensional vector $[r_{user}, c\delta t]$ divided by the RMS pseudorange measurement error. PDOP is similar, except that it applies to the error in the 3-dimensional vector $r_{user}$. Low values of GDOP and PDOP are preferred. The GPS constellation's orbital geometry has been designed so that there are a minimum of 5 satellites in view over the entire globe, and the average number in view is 7 or 8. Typical values of GDOP range from 2 to 3 around the globe if there is a clear view of the sky, but GDOP values of 10 or more can occur if obstructions leave only 4 visible satellites. PDOP is guaranteed to be below 6 if there are no GPS satellite outages\textsuperscript{15}.

B. Military Receivers

Military users have access to encrypted signals that have a number of advantages in comparison to the current civilian signal. The current military signal is called the P(Y) code. The P stands for precision, and the Y indicates that it is encrypted. The civilian user currently only has access to the Coarse/Acquisition (C/A) code, which is unencrypted. The encryption of the P(Y) code prevents it from being spoofed. That is, no enemy can broadcast a false replica of the code in order to confuse a receiver because no enemy knows the code. The C/A code is subject to spoofing. The P(Y) code's nominal chipping rate is 10.23 MHz, which is 10 times faster than the C/A code chipping rate. This faster chipping rate offers two advantages. First, it reduce the error due to receiver thermal noise and the error due to multi-path. Second, it gives the signal an additional 10 dB of resistance to jamming. The military code is present on the two GPS frequencies, $f_{L1}$ and $f_{L2}$. This fact allows dual-frequency military receivers to measure the ionospheric delay and correct for it, which reduces the residual RMS effect of ionospheric errors from 4 m to 1 m\textsuperscript{14}. Specially-designed civilian receivers can use the encrypted military signal on $f_{L2}$ in order to make similar ionospheric corrections\textsuperscript{16}, but the cross-correlation process needed in order to work with the unknown encryption bits makes current civilian dual-frequency receivers expensive and prone to loss of lock.

There are two disadvantages to using the military P(Y) code signal. The first is that the necessary receiver hardware is more complex and more expensive. A typical handheld civilian receiver costs about $300 or less, whereas a similar military unit may cost $3000\textsuperscript{17}. Although part of the extra expense is probably a result of the military procurement process, some of the extra cost is caused by the signal's faster chipping rate. This faster rate necessitates more digital processing, which requires a faster chip and more power. The second disadvantage of the military code is its length. The P(Y) code repeats once per week, as opposed to once per msec for the C/A code. This fact requires that many more acquisition calculations be done in order to find the P(Y) code directly. One solution to this problem is to use the C/A code for acquisition and the P(Y) code only for signal tracking. This strategy is problematic if acquisition must occur under spoofing or jamming conditions.
C. Differential GPS and Augmentation Systems

Differential techniques have been developed as a means of reducing the various error sources that affect the measurements in eq. (3). Differential GPS exploits the fact that some of the principal errors are correlated in space and time. These include the ephemeris and clock errors for the GPS satellites, the ionospheric errors, and the tropospheric errors. The differential GPS concept is illustrated in Fig. 7. The basic idea is to measure the cumulative effect of the range measurement errors by using a reference receiver at a known location. These errors are then transmitted via an independent radio link to a mobile receiver. The mobile receiver uses these measured errors in order to correct its pseudorange measurements before it computes its navigation solution. The two main requirements for implementing differential GPS are to have a receiver at a known reference location and to have a data transmission link between that receiver and the mobile receiver.

There are a number of differential GPS systems. The Local Area Augmentation System (LAAS) is an FAA system that bases the reference receivers at airports and that broadcasts its corrections to mobile receivers using a VHF signal. Its 95% accuracy is 0.5 meters, and its range of operation is with 45 km of the airport in question. Its purpose is to improve the accuracy, availability, and integrity of GPS to the point where it can be used for certain phases of landing approach. The U.S. Coast Guard operates a differential GPS system that has similarities to LAAS, though its goal is to help shipping.

The Wide-Area Augmentation System (WAAS) is another FAA program that covers the entire continental U.S. It uses a network of reference receivers that are spread throughout North America in order to characterize the ionosphere, the satellite ephemeris errors, and the satellite clock errors. It transmits data from which pseudorange corrections can be computed that are tailored to the particular locale of the mobile receiver. The WAAS system broadcasts its data to the users from geostationary satellites using the GPS $f_{1}$ frequency and a PRN code that is available in GPS receivers. This broadcast scheme allows a receiver manufacturer to implement WAAS corrections without adding new hardware to the design. The absolute accuracy of a WAAS-equipped GPS receiver is about 1-2 m. The European Geostationary Navigation Overlay Service (EGNOS) provides differential GPS services over Europe that are similar to those provided by WAAS over the U.S. Japan is planning a similar system called the Quasi-Zenith Satellite System (QZSS), and India may develop its own system called GAGAN (the sky). All such systems are called Satellite-Based Augmentation Systems (SBAS) because they use satellites to transmit their correction signals to the user receivers.

Fig. 7. Illustration of a local area differential GPS system (drawing courtesy of B.W. Parkinson).
D. Carrier-Phase Measurements and Carrier-Phase Differential GPS

Carrier-Phase Differential GPS techniques offer an extremely high level of accuracy. The automatic landing system discussed in the introduction, the JPALS, makes use of CDGPS techniques. Many GPS receivers are equipped to measure the carrier phase of the received signal. Thinking in terms of eq. (2), the measured carrier phase is the argument of that equation's cosine function. Suppose that this measured value is \( \theta \). The receiver takes this measured value and uses it to compute the beat carrier phase, which is the difference between the nominal phase that would result if there were no Doppler shift and the actual measured phase:

\[
\phi(t) = f_{L1}(t+\delta t) + \gamma_0/(2\pi) - \theta(t)/(2\pi) \tag{4}
\]

where \( \phi(t) \) is the measured beat carrier phase at time \( t \) expressed in units of carrier cycles and \( \gamma_0 \) is the initial value of the receiver's replica of the nominal phase. The explicit inclusion of \( \delta t \) in the first term on the right-hand side of this equation amounts to a recognition that the nominal phase signal is corrupted by the receiver's clock error. This measurement recipe can be used to develop the following carrier-phase measurement model for the \( i \)th GPS satellite signal:

\[
\lambda_{L1}\phi^i = \sqrt{(r^i - r_{user})^2 (r^i - r_{user})} + c\delta t + \lambda_{L1}N^i + v_\phi^i \quad \text{for } i = 1, \ldots, n_{sats} \tag{5}
\]

where \( \lambda_{L1} = c/f_{L1} = 0.1903 \text{ m} \) is the nominal carrier wavelength at the \( f_{L1} \) frequency, \( N^i \) is a constant carrier phase bias, and \( v_\phi^i \) is the total carrier-phase measurement error. The bias \( N^i \) is the sum of \( (\gamma_0-\theta^i_0)/2\pi \) from eqs. (2) and (5) and an ambiguity in the integer number of cycles between the measured \( \theta(t) \) value in the receiver and the cosine argument in eq. (2).

The principal advantage of CDGPS is that some of the errors in \( N^i \) are very small while other errors are highly correlated between receivers that are located within about 10 km of each other. The uncorrelated errors include multipath and thermal noise, and these may be on the order of 0.005 m when sampled at 1000 Hz. The correlated errors come from the ionosphere, the troposphere, the GPS satellite ephemerides, and the GPS satellite clock corrections. If a reference receiver at a known location is used to difference these correlated errors out of the signal at a mobile receiver, then the resulting measurement can be used to determine the vector from the reference receiver to the mobile receiver to a precision on the order of 1 cm. Over a 10 km distance this represents 1 part in \( 10^6 \). Advanced dual-frequency techniques can be used to extend the region of applicability of this level of precision to 100 km and more.

A CDGPS solution algorithm needs to estimate the bias term \( N^i \). This necessity causes a CDGPS system to require more than 4 signals in order to make this determination. The best CDGPS systems use sophisticated estimation algorithms which exploit the fact that differences of the unknown \( N^i \) values between two different receivers and two different GPS satellite signals result in integer-valued double-differences of the biases. This happens because the differencing process removes the non-integer \( (\gamma_0-\theta^i_0)/2\pi \) terms from the biases. Exploitation of the integer nature of the double-differenced biases enables a CDGPS estimation algorithm to achieve significant improvements in accuracy or significant reductions in the observation time required in order to achieve a given level of accuracy.

The following case study illustrates the extremely high precision of GPS carrier phase measurements. A GPS receiver was flown on a sounding rocket that followed a ballistic trajectory for a significant amount of time after it exited the atmosphere. The receiver derived...
a velocity estimate by differentiating its measured carrier phases in order to measure the Doppler shifts. It used a differentiated version of the measurement model in eq. (5) in order to solve for the velocity. The velocity estimates were then time-differenced in order to estimate the sounding rocket's acceleration. Thus, the acceleration estimate was based, in effect, on the second time derivatives of the measured carrier phase signals. The resultant acceleration estimate was compared to that predicted by the Earth's gravity field. The GPS acceleration's high-frequency noise was on the order of $1/200^{th}$ of a g, and its bias was even smaller -- see Fig. 8. The GPS acceleration accuracy was good enough to clearly distinguish the effect on gravity of the Earth's oblateness because the GPS acceleration did not match the modeled gravitational acceleration until the oblateness term was included in that model. This level of acceleration accuracy even after double time differencing of the measured carrier phase signal is possible only because the carrier phase measurement has very little high-frequency random noise.

**Fig. 8.** Difference between GPS-derived and "truth" ECEF Z-axis acceleration time histories for the PHAZE II sounding rocket (Fig. 5 of Ref. 24).

### E. Integrity Monitoring

Integrity monitoring ascertains whether the GPS navigation solution is trustworthy. An integrity monitor warns the user if the receiver's position estimate is likely to be in error by more than some pre-specified performance limit. Integrity problems can arise due to problems with the transmitted signals, due to ionospheric disturbances, or due to jamming or other signal reception problems.

Receiver Autonomous Integrity Monitoring (RAIM) uses signal analysis at the receiver level in order to determine whether the signals are behaving in the expected ways. These techniques can be applied at the channel level in order to monitor signal strength and possible interference. They can also be applied when the receiver computes its navigation solution. RAIM analysis of the navigation equations considers the squared error in eq. (3) summed over all of the satellites. If this statistic fails to conform to the expected chi-squared distribution,
then the receiver sounds an integrity alarm, and further analysis may occur in an attempt to determine which signal or signals are having integrity problems.

Integrity monitoring is an important part of many differential GPS augmentation systems. The LAAS and WAAS systems use their reference receivers to check the integrity of each received signal \(^{19,20}\). The messages that get broadcast to the mobile receivers include warning messages if certain signals are not to be trusted. This warning system enables LAAS- and WAAS-equipped GPS receivers to achieve sufficient integrity to allow their use for various categories of aircraft approach.

III. Historical Notes about the GPS

A. Navigation before the use of Satellite Signals

The history of global navigation has passed through many phases, and each advance has saved lives, expedited commerce, or provided military advantages. During the early 1900s, global navigation was accomplished using a sextant to observe celestial objects and using a mechanical chronometer to determine universal time. Position fixes were computed by hand and were based on tables of the celestial positions of the stars, the Sun, and the Moon. The advent of world-wide radio timing signals lessened the need for accurate chronometers. The accuracy of celestial navigation was about 4 km (about 2 nm) \(^{25}\), and there was no means of achieving a position fix in overcast conditions. In fact, position fixes normally could be achieved only at sunrise and sunset, when the horizon and the stars were both visible. Complete position fixes during the day were possible only when the Moon was visible in addition to the Sun. Celestial navigation from a moving aircraft was more difficult because of the need to use a spirit level in order to define an artificial horizon \(^{25}\).

The 1940s brought the development of the first radio navigation system, the LORAN system. It works by sending precisely timed radio pulses from broadcast stations at known fixed locations on the Earth's surface. A LORAN receiver measures the time differences between signals from different stations. Each time difference corresponds to a particular hyperbola on the Earth's surface. Thus, the signals from 3 stations provide two time differences, which define two hyperbola whose intersection is the position of the receiver. LORAN is not a truly global system. Similar principles were used to develop the global system known as OMEGA \(^{25}\), which was operated from the early 1970s until the 1990s. These systems' accuracies are on the order of 2-7 km, and they only give 2-dimensional fixes; altitude must be known a priori.

The VOR navigation system also deserves mention as a radio-based system. Note, however, that it only gives bearing to a destination, not a position fix. Thus, it is not comparable to the LORAN and OMEGA systems.

B. Precursor Satellite Navigation Systems \(^{11}\)

During the 1960s and early 1970s, the U.S. military experimented with 3 different satellite systems that performed navigation or related functions. One was the Navy’s Transit satellite system. This system broadcast a continuous carrier wave. A navigation receiver on the Earth’s surface would measure the Doppler shift of this signal. The receiver would determine the inflection point of the Doppler shift time history in order to determine the time of the satellite’s closest approach along its orbit, and it would use the slope of the Doppler shift at the inflection point in order to determine the distance to the satellite. The ephemerides of the satellite were obtained over a separate radio link, and these were used in conjunction with the Doppler shift data in order to derive a position fix. The fix calculations required a
priori altitude and velocity information as inputs. Accuracies on the order of several hundred meters were achievable. A second precursor was the Timation Satellite system. This Navy program developed satellites that orbited very precise clocks, eventually using atomic clocks with frequency stabilities on the order of a few parts in $10^{12}$ per day. Timation satellites also broadcast a timing/ranging signal. Their stable clocks improved the ability to predict their orbits and reduced the update frequency required in order to maintain accurate orbital ephemerides and accurate clock calibration parameters.

The third system was the U.S. Air Force’s Project 621B. The most important feature of this project was its development and use of PRN codes to implement range measurements, as described for the $C_{PRN}(t)$ function of eq. (1), eq. (2), and Fig. 5. This development enabled accurate ranging and the use of a single broadcast frequency by multiple satellites that were simultaneously in view. The signals could be broadcast at low power levels, below the noise density floor, and still be detected and used. These signals provided resistance to jamming because the receiver worked by de-spreading the spectrum of the signal’s PRN code. This de-spreading of the PRN code had the inverse effect on a jamming signal: the jammer was spread out into something that looked like random noise. It could then be largely eliminated through the use of narrow-band filters. Another benefit of this signal scheme is that it provided a means of encoding data at a low bit transmission rate.

Each of these systems had deficiencies. The Transit system could not tolerate more than one satellite in view because the two satellites’ signals would jam each other. This system also suffered from a low position update rate, once per orbit of a spacecraft, and an inability to determine altitude. The Timation system ranging signals from different satellites also tended to interfere with each other. The 621B program envisioned a satellite system that required continuous updates from ground stations in order to keep its clocks synchronized, which made the system more expensive to operate and vulnerable to jamming of its up-link.

C. Synthesis of the GPS from its Precursors

In 1973, the U.S. Department of Defense (DoD) decreed the creation of a Joint Program Office (JPO) that would develop a satellite navigation system for all 3 military branches. The reasons for creating a joint project were to increase efficiency and to eliminate inter-service rivalry over resources. Dr. (Col.) Bradford W. Parkinson of the U.S. Air Force was the first program director.

The JPO’s first proposal for a navigation satellite system did not win approval. The rejected design was similar to the Air Force 621B program. The DoD was supportive of the concept of a global navigation satellite system, but the initial proposal was rejected because it was not viewed as being truly reflective of all three services’ requirements. This seeming failure turned out to be a blessing in disguise because it forced the eventual GPS design to shed the faults of the 621B program by incorporating useful technology from the Transit and Timation programs.

The reaction of the JPO to the initial rejection in August 1973 was to call a meeting in order to develop a new design that truly reflected the inputs of all 3 services. The meeting took place at the Pentagon during the Labor day holiday weekend in early Sept. 1973, and the resulting design won approval in Dec. of that year. The GPS as it exists today is essentially the same system that was designed during that holiday weekend.

D. Shepherding the Project Along

The world takes the GPS for granted today, but its development and deployment were not foregone conclusions during the middle and latter half of the 1970s. Dr. (Col.) Parkinson
had to make many trips to Washington in order to hold at bay the many people who could have cancelled the project. Several important events during the project's history probably played significant roles in its eventual success at getting funded and built by the DoD. One such event was a chance meeting between Dr. (Col.) Parkinson and Dr. Malcolm Currie, who was then the head of Defense Development, Research, and Engineering (DDR&E), which made him the No. 3 man in the U.S. DoD. Dr. Currie found that he had extra time on his hands during one of his visits to the Los Angeles Air Force Base. It was suggested that he use this time to learn about the satellite navigation project from the director of its JPO. The result was a 3 hour meeting with Dr. (Col.) Parkinson that convinced Dr. Currie to become a firm supporter of the project.

Even the naming of the system has its roots in the struggle to keep it funded. The name “Global Positioning System” was suggested by a general who was then the Director for Space for the U.S. Air Force Deputy Chief of Staff Research and Development. This name was adopted because the general was a sponsor of the JPO. The name “NAVSTAR” was added as a prefix because it was suggested by an associate director of DDR&E who liked its sound -- NAVSTAR is not an acronym. This person’s budgetary decision making powers made his support critical. His level of enthusiasm for the program had been somewhat lacking, and the JPO reasoned that his support might solidify if he had a hand in naming the system. Thus, the system’s entire name became "NAVSTAR the Global Positioning System." The program’s fast development timeline contributed to its success by shortening its opponents’ window of opportunity for having it cancelled. February 1978 saw the launch of the first prototype satellite, just 44 months after the contract was awarded to build it. The fourth satellite was launched by the end of 1978, at which point the basic 3-dimensional navigation concept could be demonstrated.

Although the initial contract covered the 4 satellites needed to demonstrate the GPS concept, the JPO wanted to get funds for additional satellites in case of launch failures or other mishaps. They succeeded in obtaining the necessary funds from the Trident missile program by showing that signals from the prototype GPS satellites could be used to support the tracking of submarine-launched Trident missiles during booster test firings.

The JPO developed a motto in order to maintain the program’s focus: “The mission of this Program is to: 1. Drop 5 bombs in the same hole, and 2. Build a cheap set that navigates (< $10,000), and don’t you forget it!” The success of the first program goal has been demonstrated by the data shown in Fig. 2. The second goal has been wildly exceeded. The $10,000 receiver price tag in 1973 dollars is equivalent to a $45,000 cost in 2006 dollars after adjusting for inflation based on the consumer price index. Nowadays, a typical civilian handheld receiver sells for about $300, and it is possible to buy a simple one for $100. Although military receivers are more expensive, military handholds only cost on the order of $3,000. The JPO did not foresee two important factors that led to this positive result: the development of cheap, low-power, high-performance DSP hardware and the massive civilian use of the GPS that has led to rapid improvements in receiver technology.

E. Comparisons with Previous Navigation Technologies

Three comparisons can be made between the GPS and its navigation technology predecessors. First, the critical role of time determination is not new to the GPS. Time determination was a major challenge to celestial navigation because time was needed in order to compute longitude. The solution was worked out by an Englishman named John Harrison. The main difference between Harrison's time determination problem and the GPS problem lies in the required level of accuracy, which is dictated by the speed involved. Harrison needed to achieve an accuracy on the order of seconds because points on the Earth's equator

13
move 1 nm with respect to Earth-centered inertial coordinates every 4 seconds. The GPS needs to achieve an accuracy on the order of 30 nsec because this is the amount of time required for light to travel 10 m.

The second comparison is with the LORAN and OMEGA radio navigation systems. The GPS implicitly exploits their strategy of using time difference measurements to place the receiver on a hyperbolic curve. Although not obvious from the measurement model in eq. (3), a GPS receiver's navigation solution algorithm effectively differenciates the pseudorange measurement equations for different GPS signals in order to eliminate the \( c/dt \) terms from the equations. The resulting differenced equations place the user receiver on intersecting hyperbolic surfaces, and the position fix algorithm solves for the intersection of these surfaces.

The last comparison is with the Transit navigation satellite system. The Transit system relied on the curved shape of the received Doppler shift time history in order to deduce navigation information. The basic GPS system does not do this, but state-of-the-art CDGPS techniques rely on this same phenomenon. The curved Doppler shift time histories of the various received signals enable CDGPS solution algorithms to accurately estimate the residual integer values of double-differences of the \( N^j \) bias terms that remain in double-differenced versions of eq. (5).

F. Significant Developments after Initial Proof of Concept

A large number of GPS satellites have been designed, built and launched since the success of the initial prototype testing in the late 1970s and early 1980s. The system reached Initial Operational Capability (IOC) on 8 December 1993. At that point its constellation had 26 Block I and Block II GPS satellites, 24 of which were in their assigned orbital locations and transmitting the Standard Positioning Service (SPS) signal, that is, the civilian signal. The Block I satellites were development versions, and the Block II satellites are operational versions. The system reached Full Operational Capability (FOC) on 27 April 1995, at which point it included 24 Block II or Block IIA satellites.

The U.S. government has committed itself to make the civilian SPS available to the world. This process started during the Reagan administration in response to the downing of Korean Airlines flight 007 in 1983 after it strayed into Soviet air space. The U.S. government has promised to keep the SPS available free of charge for the foreseeable future and to give at least six years notice prior to termination of the service.

The history of GPS receiver sizes, weights, and costs has been similar to that of the personal computer: the trends have been steeply downward since the late 1970s. Initial development receivers were large and expensive. The military's first "portable" GPS receiver was called the Manpack and was the size of a backpack, weighed 8 kg, cost $45,000, and had 1 channel. One-channel operation meant that the time to obtain a fix was at least 2 minutes and probably much longer. By contrast, today's handheld military receivers weigh less than 0.5 kg, have 12 channels, use both GPS frequencies, and cost about $3,000. A typical civilian handheld receiver weighs 230 gr, costs about $300, has 12 channels on the \( f_{L1} \) frequency, and can often get its first fix in less than 15 seconds.

The military realization of the importance of GPS occurred during operation Desert Storm, the Gulf War with Iraq that occurred in 1991. The GPS enabled the U.S.-led coalition to efficiently navigate and direct fire in difficult desert conditions 24 hours per day and 7 days per week. The system's success out-ran the U.S. military's ability to provide receivers, which caused it to buy over 10,000 civilian receivers. The military also temporarily suspended the use of Selective Availability (SA), which is an intentional degradation of the civilian signal.

A dominant aspect of recent GPS developments is the U.S. government’s reaction to the impending development of Galileo. The European Galileo system has been conceived...
primarily as a commercial system. Its design goals put it on track to provide superior accuracy than was originally provided by the civilian signal of the GPS.

This "competition" from Galileo has put pressure on the GPS to improve its service. Otherwise, the GPS might be abandoned by future civilian users, which would make the U.S. Congress less likely to fund GPS maintenance and improvements. Another concern is that domination of the civilian market by Galileo might decrease the trickle-down effect whereby marked-driven advances in civilian receivers eventually find their way into military units. Perhaps the most significant concern is that U.S. civilian receiver manufacturers might be put at a competitive disadvantage if Galileo becomes the dominant system because the Europeans might limit access to Galileo technical information.

The GPS has begun its response to Galileo. It implemented a major improvement in system accuracy by the “flick of a switch” that discontinued SA. SA deliberately degraded the accuracies of the reported broadcast times $t_b$ and of the ephemerides that are used to calculate the satellite positions $r_i$. These degradations increased the civilian system's RMS positioning errors from about 10 m to about 50 m. Their intent was to deny precise GPS navigation accuracy to any potential U.S. military adversaries. SA was turned off permanently on 1 May 2000 by order of President Clinton. This change instantly made millions of civilian GPS receivers about 5 to 10 times more accurate without making a single modification to any of them, as evidenced by the navigation error time history plotted in Fig. 9. The plot includes the time at with SA was turned off, 04 hours UTC. The sudden accuracy improvement is striking: peak positioning errors were almost 100 m with SA turned on, but they immediately decreased to about 10 m when SA was turned off.

The GPS has plans to improve the accuracy of the broadcast values of the satellite ephemerides and the transmitter clock calibration parameters. There is evidence that this accuracy has been gradually improving since the end of SA, but specific numbers were not available for publication here.

An interesting historical perspective on SA is that it provided one of the main motivations for the original work on differential GPS techniques. Differential techniques offered an effective means of countering SA because the time correlations of its deliberate errors were long enough to allow a reference receiver to transmit the necessary corrections to a mobile receiver. These corrections could enable a mobile receiver to achieve an absolute position accuracy on the order of 10 m or better, which was a tremendous improvement over the 100 m peak errors of a stand-alone civilian receiver operating under SA conditions. Once developments began, differential techniques took on a life of their own and produced all sorts of useful technology as embodied in the WAAS and LAAS systems and especially as embodied in CDGPS systems. The usefulness of these systems goes far beyond their ability to
counteract SA, especially in the case of CDGPS systems with cm-level accuracy. Thus, these important developments represent positive legacies of the U.S. military's attempt to deny accuracy to civilian users. These developments, in turn, made SA largely obsolete so that it became easier for the U.S. military to accept the decision to turn off SA in response to competition from Galileo.

IV. Current Trends in Civilian GPS Technology
A. Economically Important Applications with Significant Technical Challenges

The civilian GPS market is large and is growing rapidly. Two of the primary drivers for this growth are GPS-equipped cell phones and GPS-equipped automobiles. The potential size of such markets is huge. Researchers and receiver manufacturers are “following the money” by targeting developments for these two areas.

The interest in GPS for cell phones is driven by several factors. One is the desire for location-based services. Cell phone users would like their handsets to tell them about available resources near their current location, resources such as public bathrooms or good restaurants, and they would like to get directions to these resources. Another significant driver is the FCC’s requirement for Phase II E911 service for cell phone users. That requirement stipulates that the wireless carrier must provide the position of a cell phone that generates an E911 call. This position must be accurate to within 50 m or 100 m on 67% of the calls and to within 150 m or 300 m on 95% of the calls.

It is difficult to provide accurate position fixes for cell phones in all environments. Good systems employ a combination of GPS signals and other radio signals, often including the cell phone’s own communications signal. GPS tends to work better in rural settings, and other signals work better in urban settings where GPS signals are attenuated too much in urban canyons or indoors in large buildings. Assisted GPS (AGPS) uses various information and computational aids that are available from the cell phone network in order to improve the sensitivity of the GPS receiver to weak signals. AGPS improves sensitivity by reducing the search space for signals, both in time and in Doppler shift, and by allowing longer signal-processing integration intervals based on knowledge of the data bit time history $D(t)$ of eq. (1). These techniques can enable the use of GPS for E911 service in buildings such as 2-story malls, but they lack sufficient sensitivity to operate inside the lower floors of very tall buildings.

Further developments to look for in GPS-based cell-phone positioning include an increasing number of GPS-equipped handsets, the multiplication of commercial location-based services, and increased receiver sensitivities that allow improved performance in challenging environments. The E911 requirement has forced the production and deployment of a sufficient number of position-capable handsets to make location-based services profitable. The number of such services should grow, which should create more interest and more demand. This positive feedback might result in a sort of “snowball” growth effect over a period of several years, until the market saturates. New signals that are discussed in Section VI should improve the sensitivity of GPS and AGPS by providing stronger signals, by simplifying the process of using longer signal integration times during detection and tracking, and by reducing inter-channel interference. MEMS inertial measurement devices and improved miniature oscillators will also enable longer integration times. The result will be an increased ability to use GPS signals in larger buildings and in urban canyons. Note that the increased multipath in such environments will require the development of new mitigation techniques in order to achieve the best possible position accuracy.
The demands for GPS automobile navigation stem from the obvious desire not to get lost when driving to new locations, but location-based services are also an important consideration in the economics of this market. Current automobile systems work well in many rural and suburban settings, but they have problems in urban canyons. One of the technology drives in the GPS industry is to improve automobile navigation performance in the latter environment.

Automobile systems have several advantages over cell-phone systems, and these advantages can be exploited in order to improve performance in urban settings. Automobile receivers need not be extremely power-efficient or small. This allows them to use better antennas, better antenna locations, and better electronics, all of which improve sensitivity. Automobile navigation systems can incorporate larger inertial measurement devices than can be used in a cell phone, and dead reckoning can be performed based on the vehicle’s odometer and steer angle. A further possible enhancement is to employ street map information directly in the estimation calculations. The availability of an increased number of satellites from the Galileo system or locally in Japan from the QZSS should improve the performance of automobile receivers in an urban canyon by increasing the likely number of available signals, as illustrated in Fig. 10.

Expected signal improvements that are discussed in Section VI will allow increased receiver sensitivity and will reduce inter-channel interference. These improvements should help in urban canyons and in situations where the automobile drives near dense foliage in a rural setting. Current systems can experience signal drop-out in this latter situation.

B. Reprogrammable Receivers

A looming need is for GNSS receivers that have flexible architectures that can handle new signal structures. As will be discussed in Section VI, a number of new and useful GNSS signals are starting to become available, and this trend is expected to accelerate. The new signals have structures that differ from the standard C/A PRN codes that are currently being broadcast on the $f_{L1}$ frequency. Almost all existing civilian receivers de-spread the PRN codes in a dedicated, custom-designed DSP chip that works only for a specific set of C/A codes.

The use of custom-designed DSP chips for the new signals is problematic. A receiver will require several different chip (or sub-chip) designs in order to process several different PRN codes if it wants to use several different types of signals. Otherwise, it will be restricted to one signal type, which will limit its performance. This trade-off will be particularly troublesome during the next 5 years. The availability of new signals will increase gradually as new satellites get launched, either to replace retired GPS satellites or to assemble the Galileo constellation. A receiver that cannot process the new signals will become obsolete when enough new satellites get launched, but a receiver that can process the signals will see much of its processing capacity lying dormant early in its life cycle, before many of the new signals are available.
The solution to this problem is to develop a flexible receiver that is adaptable to new signals without the need for new hardware. Field-Programmable Logic Gate Arrays (FPGAs) provide one route to a flexible receiver design. An FPGA is a custom-designed DSP whose design can be modified by downloading a new program. An FPGA offers flexibility and the potential to implement powerful parallel computations, but the re-programming of an FPGA can be a complicated task.

Another way to achieve flexibility is to use real-time software radio technology. A real-time software radio performs its PRN code dispreading functions using a general-purpose DSP chip programmed in a conventional programming language such as C. This approach allows the receiver to deal with new PRN code merely through moderate software changes.

Several real-time software GPS receivers have been developed and tested\textsuperscript{34-37}. They can have equivalent performance to receivers that implement code dispreading in a dedicated hardware chip\textsuperscript{38}. Two of these works have demonstrated the adaptability of this type of receiver to the new signals that are beginning to appear\textsuperscript{34,37}. Evidence of the likely future importance of such systems is provided by the continued health and growth of a start-up company that is based entirely on this technology\textsuperscript{39}.

C. New Applications in Space

GPS has been used successfully onboard individual LEO spacecraft for a number of years. New developments in GPS space applications center on its use for formation flight and for high-altitude applications. CDGPS techniques can be used to measure the relative positions of formations of spacecraft to an accuracy on the order of 1 cm\textsuperscript{40,41}, similar to what can be done in terrestrial applications. The resulting estimates can be generated in real-time and used to control the shape of the formation. They can also be used in processing measurements made by radar or other instruments mounted on the elements of the formation. One possible application for such technology is to develop a large aperture phased-array radar for Earth observation that consists of a formation of small free-flying spacecraft. The resolution of such a system could be increased by increasing the formation spacing without the need to increase the size of any individual spacecraft. This approach could enable the development of a high-performance system at a fraction of the cost of a system that was based on one large spacecraft.

The use of GPS in high-altitude orbits is difficult because the user receiver lies above the GPS constellation, as illustrated in Fig. 11. The figure depicts three possible locations of a formation of 3 user spacecraft that fly in geosynchronous Earth orbits (GEOs). The GEO altitude is about 16,000 km above the GPS constellation. The only useable signals come from GPS spacecraft that orbit on the other side of the Earth; the nearest GPS spacecraft have their transmitting antennas pointed away from the user spacecraft. The high-altitude spacecraft need to be able to acquire and track weak signals that lie in the side lobes of the GPS transmitting antennas because there is only a small window of availability for...
each main-lobes signal, as illustrated by the figure. If the receiver is restricted to use only main-lobes signals, then signal availability is severely limited: There are never more than 2 usable main-lobes signals, and often there are none.

Therefore, it is necessary to use ultra sensitive receivers in order to make effective use of GPS at high altitudes. This ability is especially important if one wants to implement CDGPS relative navigation calculations for a constellation of high-altitude spacecraft, as in Ref. 41. To this end, NASA has begun to develop receivers that have the necessary sensitivity 42, and one of its ultra-sensitive receivers is slated to fly on a future geostationary mission. This need for increased receiver sensitivity is similar to the need of the terrestrial E911 GPS application. Therefore, similar techniques are sometimes used in the two applications.

V. Issues of Military Interest

A. Jamming

The valuable contributions of GPS in the two wars with Iraq have increased the U.S. military’s reliance on this system for functions that range from accurate weapons delivery to rapid location of ground forces and precise coordination of their actions. This increased reliance has brought an increased military vulnerability to jamming of the system. The P(Y) military code cannot be spoofed, and it is 10 dB more jam resistant than the C/A code, but it can still be overwhelmed by a sufficiently powerful jammer. Therefore, the subject of anti-jamming techniques continues to be a field in which the U.S. military has a research and development interest. Some of the more promising techniques resemble some of the techniques used for assisted GPS in cell phones. These approaches include the use of inertial measurement units and known time histories of the D(t) data bit stream in order to allow longer integration times in the receiver’s signal processing algorithms. These longer integration times have the effect of filtering out more of the jammer signal so that the received signal-to-noise ratio is sufficiently high to enable successful navigation.

B. The Temptation to Use Civilian GPS Signals

Another concern of the military is the temptation to use civilian receivers in military operations. At the top end of the military command structure, officers in charge of procurement budgets have been heard to complain about the high cost of military-grade units in comparison to seemingly equivalent civilian units 17. At the bottom end of the command structure, a private is not above bringing his own personal civilian handheld unit into the field when he finds out that the army will not issue him his own personal military receiver.

This use of civilian receivers in the military environment poses three hazards. First, the civilian receivers are not built as ruggedly as their military counterparts and, therefore, they are more likely to fail under field conditions. Second, an enemy could spoof the civilian systems or jam them with a low-level jamming signal that was harder to detect and destroy. Third, the U.S. military might need to jam the civilian signal in order to deny its use to an enemy. If some functions of the U.S. military depended on civilian receivers, then these functions would be lost.

C. The Potential use of Civilian GPS by Adversaries

Another significant concern of the U.S. military is the possibility that adversaries might use the civilian GPS service to improve their military capabilities. For example, they might use it to develop an accurate missile guidance system. The U.S. government restricts the export of civilian GPS receivers that can operate above an altitude of 18 km (60,000 ft) or above a speed of 515 m/sec (1,000 knots) 17. Unfortunately, it is possible to buy commercial
GPS chips and source code from foreign vendors that can be used to design receivers that violate these limits.

The author was involved in just such a project to design a GPS receiver for a sounding rocket system. It took only about 1-2 man years to develop the system based on publicly available information and parts and source code that were available from a foreign vendor. The system successfully flew on the SIERRA sounding rocket mission in January of 2002. The GPS receiver maintained lock and navigated through a peak acceleration of 16 g’s, a peak velocity of 3,500 m/sec, and a peak altitude of 735 km. All the while, it maintained a position accuracy of about 10 m or better. This was not extremely difficult to do.

A key contributor to the successful development of this system was the availability of a GPS signal simulator such as the one shown in Fig. 12. This type of simulator allows a receiver developer to generate RF GPS signals that appear to the receiver as though they come from an antenna mounted on a moving vehicle. Available motion scenarios are virtually limitless, and it is relatively straightforward to simulate and entire missile flight using such a device. The SIERRA receiver was tested on hundreds of simulated sounding rocket flights using an equivalent simulator. This simulation test program enabled the development team to work out all of the problems associated with flight at high altitude, high velocity, and high acceleration.

![Multi-output GPS Simulation System](image)

**Fig. 12.** A GPS RF simulator that can be used to test a receiver using realistic dynamic scenarios that include missile launch and spaceflight (courtesy of Spirent Communications at http://www.spirentcom.com).

A related concern is the recent successful flight of a radio-controlled model airplane over the Atlantic Ocean using civilian GPS to guide it. The UAV had a 2 m wingspan and carried a 5 kg payload. Although its mission was entirely peaceful, the concern is that an adversary could implement a sort of cruise missile using this technology. Its GPS guidance system could enable it to deliver its weapons payload to a specific room of a specific building, such as the oval office in the White House. Such a system could use a simple $100 handheld receiver that was incapable of violating the 18 km altitude limit or the 515 m/sec velocity limit.

As has been mentioned already, the original intent of Selective Availability was to degrade the civilian signal so that it would not be very useful to an enemy guidance system.
Pressure from the success of differential GPS and from the planned development of Galileo has forced SA to be turned off. This leaves the frightening potential for an educated enemy to use the civilian signal to do significant harm to the U.S. or to her allies. Therefore, the U.S. military has had to develop a new strategy for denying this capability to an adversary in time of war. This strategy employs intentional jamming of the civilian signal by the military in limited combat regions. Additional discussion of this subject will be presented in Section VI.

D. Possible Elements of a Small Country's Policy on the Military Use of GNSS Signals

A small country's defense forces should be concerned about GNSS signals for two reasons. First, they should want to use them in order to gain battlefield advantages over their adversaries. Second, they should want to deny the use of these signals to their adversaries, especially in the area of missile guidance. There are several possible strategies for achieving these twin objectives, but they have differing price tags and differing levels of likely effectiveness.

The second concern can be treated first. The way to deny the advantages of the civilian GPS signal to an adversary is to jam the signal or to spoof it. It is a relatively weak signal. Therefore, it is easy to jam, and there are Russian jamming devices available on the international market. One might want to develop one's own jamming signal, however, depending on how one wants to solve the problem of retaining use of GNSS signals.

The one problem with jamming or spoofing is that one cannot turn it on whenever one feels threatened. There are international agreements about reserved portions of the electromagnetic spectrum, and jamming of the GPS civilian signal would violate these agreements. Therefore, jamming can be employed only during a time of confirmed dire military necessity.

The most desirable means of assuring access to high-quality GNSS signals during a conflict would be to use U.S. military receivers that can process the encrypted military signals. It may be possible for a small country that has friendly relations with the U.S. to get access to such receivers. The U.S. military is broadening the access to these receivers because it has upgraded its signal encryption keying approach in a way that reduces the risk that an adversary will decipher the encryption by gaining access to a military receiver. The availability of such receivers would allow a small country to deny the use of the civilian signal to an enemy via jamming or spoofing while maintaining its own access to a precise navigation capability.

If this option is not feasible, then another option is to develop a Local Positioning System (LPS). This might involve a set of 5 or 6 geosynchronous satellites. They would form a formation that orbited around a stationary point above the equator in a way that maintained a visibility of 4 or 5 satellites in the small country and the surrounding region but that also maintained sufficient spacing between the satellites to produce a reasonably low GDOP. The satellites might not need to use atomic clocks because it might be possible to maintain almost continuous ground contact with them so that they could be re-calibrated from the ground at a high update rate. Of course, the development, deployment, and maintenance of an LPS would be an expensive undertaking.

A less expensive option would be to jam the civilian signal with a known pseudo-random jamming signal. The PRN code of the jamming signal would be kept secret. The small country could develop GPS receivers that had the capability to track this jamming signal and subtract it out of the received signal. The receiver could then process the resulting signal using normal civilian GPS techniques. The only costs of the system would be for the PRN jammers and for the new receivers. The receivers would be somewhat more expensive than standard receivers because their RF front ends would have to use additional bits of ADC conversion in order to maintain a usable signal level after the cancellation of the strong
jammer signal. The one disadvantage of this approach is that the new receivers would still be vulnerable to jamming and spoofing by an adversary. Therefore, a strategy would have to be developed to monitor the GPS signals’ integrity using reference receivers, and a separate communications link would need to transmit the integrity message to the GPS receivers. The integrity message could be transmitted on the small country's jammer signal. This strategy would also require the ability to rapidly locate and destroy enemy jammers and spoofers in order to maintain a reasonable level of availability.

VI. The Future of GNSS

The dominant changes expected in the future of GNSS will be caused by the introduction of a number of new signals. The GPS has begun implementing new signals, and has plans for more. The Galileo system will add yet more GNSS signals. The availability of these new signals will provide many enhancements to GNSS services, but they also raise some concerns.

A. New GPS Signals

The new GPS signals are shown in Fig. 13 along with the existing signals. Each graph on the figure is a plot of power spectral density vs. frequency for the given signal. The top line shows the two existing signals, the civilian C/A signal on the \( f_{L1} \) frequency and the encrypted military P(Y) signal on both \( f_{L1} \) and \( f_{L2} \). The middle line shows these two signals plus two new signals that started to become available in Oct. 2005 with the launch of the first Block IIR-M GPS spacecraft. These signals are the L2 civilian signal and the military M code. The L2 civilian signal has a PRN code chipping rate of 1.023 MHz, like the C/A code, which is why the power spectra of the two signals look similar. The encrypted M code uses a new signal structure that is known as a BOC(10,5) structure. Referring to the model form in eq. (1), a BOC(10,5) signal uses a \( C(t) \) time history of +1/-1 values that is the product of a PRN code and a square-wave sub-carrier. The PRN code has a chipping frequency of \( 5 \times 1.023 = 5.115 \) MHz, and the square-wave sub-carrier has an oscillation frequency of \( 10 \times 1.023 = 10.23 \) MHz. The sub-carrier is called a Binary Offset Carrier (BOC). It causes the two main power lobes of the signal to be offset +/-10.23 MHz from the nominal carrier frequency. The bottom line of Fig. 13 shows a third civilian signal that will be added at the nominal carrier frequency \( f_{L5} = 1176.45 \) MHz when the first Block IIF satellite gets launched, perhaps some time in 2007. The L5 signal uses a PRN code with a chipping rate of 10.23 MHz. If the current satellite replacement schedule is maintained, then the M code signal and the civilian L2 signal are likely to reach FOC around 2014, and the civilian L5 signal should reach FOC around 2016.

The new GPS signals will offer a number of improvements. The new civilian L2 signal will enable dual-frequency civilian receivers to directly measure and cancel ionospheric effects. Such receivers will no longer need to use the expensive and unreliable process of P(Y) code cross correlation between the L1 and L2 signals. The civilian L2 signal will also enable increased receiver sensitivity. It uses longer PRN codes that reduce both inter-channel interference and the possibility that a strong signal will interfere with a weak signal. Each channel includes a pilot signal that does not carry a data message. This will enable easier detection and tracking of weak signals because un-assisted GPS receivers will be able to use long signal-processing integration intervals in order to increase their sensitivities.

The new military M code offers several advantages. It is encrypted like the P(Y) code so that it cannot be spoofed. The M code transmitters have the option to transmit spot patterns that have much higher power in regions where an adversary may be jamming the signal, thereby reducing or eliminating the impact of the jammer on receiver performance. The M
code's BOC structure allows the U.S. military to jam the civilian signals at the center of the spectrum without jamming the M code. Conversely, the M code spot pattern can be used with increased power without jamming the civilian signals.

The new civilian signal at the frequency $f_{L5}$ will offer several benefits. Its higher power level and higher chipping rate increase its immunity to jamming, whether intentional or unintentional. Its higher chipping rate also decreases its multipath errors. Its short PRN codes (they repeat every 1 msec) make acquisition relatively easy. It has two channels that operate in phase quadrature. One of these carries data, and the other does not. The data-less channel allows increased signal acquisition and tracking sensitivity. Increased receiver sensitivity, when combined with increased signal power, will make these signals more accessible in urban canyons, in tall buildings, and in high-altitude Earth orbits.

An important property of the L5 frequency band is that it lies in a spectrum that is aviation-protected. The L1 signals also lie in a protected band, but the L2 signals do not. Therefore, the introduction of L5 signals will enable aviation certifiable receivers to use dual-frequency techniques for mitigating the impact of ionospheric errors.

A big uncertainty about the GPS regards GPS III. GPS III will build new satellites that incorporate improvements over the Block-II versions, and it will improve the ground control segment. As far as the user is concerned, its goals are to improve signal accuracy, availability, resistance to jamming. One of the improvements may be a new civilian L1 signal that will use a BOC structure. The legacy C/A signal will also be broadcast on the L1 frequency in order to maintain the system's backwards compatibility. At present, there is much uncertainty about the form of GPS III and about the U.S. commitment to develop it, and that uncertainty has caused concern about the long-term future of the GPS.

Fig. 13 Frequency structure and power levels of existing and planned GPS signals (plots courtesy of B.W. Parkinson).

The new civilian signal at the frequency $f_{L5}$ will offer several benefits. Its higher power level and higher chipping rate increase its immunity to jamming, whether intentional or unintentional. Its higher chipping rate also decreases its multipath errors. Its short PRN codes (they repeat every 1 msec) make acquisition relatively easy. It has two channels that operate in phase quadrature. One of these carries data, and the other does not. The data-less channel allows increased signal acquisition and tracking sensitivity. Increased receiver sensitivity, when combined with increased signal power, will make these signals more accessible in urban canyons, in tall buildings, and in high-altitude Earth orbits.

An important property of the L5 frequency band is that it lies in a spectrum that is aviation-protected. The L1 signals also lie in a protected band, but the L2 signals do not. Therefore, the introduction of L5 signals will enable aviation certifiable receivers to use dual-frequency techniques for mitigating the impact of ionospheric errors.

A big uncertainty about the GPS regards GPS III. GPS III will build new satellites that incorporate improvements over the Block-II versions, and it will improve the ground control segment. As far as the user is concerned, its goals are to improve signal accuracy, availability, resistance to jamming. One of the improvements may be a new civilian L1 signal that will use a BOC structure. The legacy C/A signal will also be broadcast on the L1 frequency in order to maintain the system's backwards compatibility. At present, there is much uncertainty about the form of GPS III and about the U.S. commitment to develop it, and that uncertainty has caused concern about the long-term future of the GPS.
B. The Galileo System

The first test satellite of the Galileo system was launched in Dec. 2006. Its name is GIOVE-A. Test navigation signals were successfully transmitted and received by mid Jan. 2006. The age of Galileo has begun.

The Galileo system will resemble the GPS in many respects. It will have 30 satellites orbiting in 3 inclined planes with orbital properties that are similar to the GPS. The Galileo satellites will transmit similar PRN-encoded signals. Some of the Galileo signals will be transmitted on the two GPS frequencies \( f_{L1} \) and \( f_{L5} \), as shown in Fig. 14. The intent of these partially overlapping signal structures is to allow the development of receivers that are capable of processing signals from both systems in a single radio-frequency front-end, as in Ref. 34. The navigation accuracies that will be provided by the Galileo signals will be similar to those provided by the GPS signals once all of the GPS upgrades discussed in Section VI.A have been implemented.

The various Galileo signals have been designated to provide components in a range of services that will be offered by the system. The L1B, L1C, E5a, and E5b signals are Open-Source (OS) signals or include OS components. Safety-of-Life (SoL) signals will also be present on the E5b and L1B or L1C signals. The structures of the OS and SoL signals are slated to be published so that any receiver manufacturer can build equipment to use these signals. Users will be able to buy and operate such receivers without paying a fee to the Galileo system. In this respect, the OS and SoL signals will be like the L1C/A-code signal, the L2 civilian signal, and the L5 signal of the GPS.

The Galileo system will provide 2 other signal types that will have restricted access. Commercial Service (CS) signals will be available on E5b and E6B, and Public Regulated Service (PRS) signals will be available on L1A and E6A. CS signals will be encrypted and usable only by subscribers. PRS signals will be encrypted and usable by governments that are part of the Galileo consortium. Thus, these two signal types are more like the encrypted P(Y) and M codes of the GPS, although the CS is presumably not intended for military use. The European Union hopes to finance the multi-billion dollar cost of Galileo through the subscriptions that will be paid for the CS.

---

Fig. 14. A comparison of the frequency structures of the Galileo signals (top line) and the GPS signals (bottom line) (Fig. 1 of Ref. 48)
C. The Effects of Galileo on GNSS

The development of the Galileo system offers several advantages and challenges for users of GNSS services. One large advantage is the intentionally designed interoperability of some of the Galileo signals with GPS signals. The BOC L1B and L1C signals use the nominal carrier frequency $f_{L1}$, and the E5a signals use the carrier frequency $f_{E5}$. This will allow a manufacturer to design a dual-frequency L1/L5 receiver that can use both civilian GPS signals and OS Galileo signals. Such a receiver will be able to perform robust dual-frequency ionospheric corrections, and it will have access to signals from 54 or more satellites. Absolute position accuracy could approach 1 m in normal situations because of the ionospheric corrections and because of the additional data from the larger number of visible satellites. Availability will be increased in urban canyons and in other areas that have restricted views of the sky, as depicted in Fig. 10, because the larger number of satellites will increase the probability of 4 satellites being visible. An additional benefit will be the ability to increase the integrity of autonomous receivers because an increased number of signals makes RAIM techniques more powerful.

There are several possible pitfalls on the road to a GPS/Galileo world. One of them is the economic implication of the plethora of possible receiver designs. It will be expensive to design a single receiver that can receive all available signals. Therefore, mass-market manufacturers will be forced to choose between different possible signal combinations. This will create a market in which there are more varieties of receivers, but fewer units of each receiver type will be sold. This fact may force manufacturers to raise prices in order to amortize development costs over fewer units.

An analogy for this conundrum of the LINUX/Windows debate in the world of PC operating systems. Windows has dominated the market and created a standard that lets applications software designers focus their attention on one operating system. LINUX allows more freedom, but there is a loss in availability of application software because it is more difficult for a software company to find a sufficiently large market in the LINUX world to justify the development of a particular product. The advent of Galileo may change the GNSS world forever from a Windows-type world to a LINUX-type world.

An alternate possibility is that one basic preferred signal combination may win out and become the new standard. If this happens and if the preferred signal structure does not include any Galileo CS signals, then the European Union may fail to recover the costs of Galileo through subscriptions to its CS.

Another concern that applies to the Galileo signals and to the new GPS signals (except for the L2 civilian signals) is their increased bandwidth, as shown in Fig. 14. Increased bandwidth gives improved performance, but at a cost: The receiver's digital electronics must function at a higher frequency, which raises power consumption and, possibly, weight and cost.

One of the big questions hanging over the advent of Galileo concerns its economic relationship to GPS. It is possible that both systems will cooperate, peacefully co-exist, and prosper. It is also possible that there will be stiff competition between the two systems in order to attract users and that one system will drive the other system out of business.

The have been rumblings of competition between the GPS and the Galileo system ever since Galileo was seriously proposed. The U.S. is concerned that Galileo could make the GPS obsolete. This would make the GPS harder to fund through the U.S. federal budget, and the U.S. military might lose all or part of the system that it has come to rely on. Alternatively, the GPS might undercut Galileo and cause Galileo's CS market to dry up. This would be a financial disaster for Galileo, and might cause the system to be abandoned.
Another U.S. concern is whether Galileo will negatively impact the U.S. civilian GPS receiver industry. Presumably, U.S. manufacturers will not be able to build receivers for Galileo's CS and PRS signals because of the Europeans' concern to protect their encryptions. If the CS signals capture a large fraction of the market, then the U.S. GNSS industry will suffer. Another concern is that significant technical details about the OS and SoL Galileo signals may not be made public so that U.S. manufacturers would be effectively cut off from that market too.

U.S. concerns over this issue have heightened with the launch of the first Galileo test satellite GIOVE-A and the successful demonstration of navigation signal transmission and reception. The OS signal structures were supposed to have been published before this event, but they have yet to be made public. A request by this author for information about the GIOVE-A L1B and L1C OS PRN codes was met with a polite apology and the explanation that only a select few European entities are allowed access to these codes at the present time. This withholding of information has raised concerns among U.S. GNSS researchers and manufacturers.

The U.S. military is concerned that Galileo will provide high-quality guidance and navigation capabilities to potential adversaries. The biggest concern is about China, which has joined the Galileo consortium. The U.S. will have to prepare to jam Galileo in the event of a conflict with a Galileo-equipped foe. The development of capabilities to jam Galileo signals will cost money and military resources. In addition, it may increase political tensions between the U.S. and the European Union.

Galileo is slated to reach FOC as early as 2008. That schedule seems optimistic and is probably out of date, but even with a slowed development and launch schedule, the Galileo system could reach FOC years ahead of the new GPS signals' FOC. This timing issue may be the critical factor in determining the relationship between GPS and Galileo and in determining whether one system dominates the other. The U.S. is feeling pressure in this regard and is being advised by its GPS experts to meet the Galileo challenge by expediting the upgrades to the GPS in order to achieve a competitive FOC date.

**D. Use of Terrestrial Signals**

Other avenues are being pursued to provide navigation signals in difficult environments, especially indoors and in urban canyons. One possibility is to use the synchronization signals from digital TV broadcasts. These signals can have 40 dB more power than GPS signals, and their timing preambles can be used to derive range information relative to TV broadcast towers. Another signal type that is being considered for similar use is a wide-band communications link that is being implemented in South Korea to provide wireless internet access over a broad area. These new methods have the potential to vastly improve cell-phone-based GPS and other navigation services in dense urban areas, but their usefulness will be restricted to localities where there are sufficiently many suitable terrestrial signals.

**E. Concerns about Ultra-Wideband Communications Systems**

There is increasing concern about the use of Ultra-Wideband (UWB) devices. These devices work by sending out sharp pulses in the time domain rather than modulated carrier signals. They are useful for imaging buried objects and objects behind walls, for communications, and for ranging. These signals operate over a broad spectrum, including the GPS spectrum. The concern of the GNSS community is that the FCC will license UWB signal levels that cause serious degradation of GNSS receiver performance because the UWB manufacturers will convince the FCC that the likely GNSS performance degradations are acceptable. The GNSS community is responding by testing the effects of UWB signals and by
trying to convince the FCC to strictly regulate the power levels of these new signals in the
GNSS frequency bands so that GNSS receiver performance will not be adversely affected 53.

VII. Summary and Conclusions

The Global Positioning System has come a long way since its conception on Labor Day weekend of 1973. It has been a major technical and economic success and has provided significant military capabilities to its owner: the U.S. military. It is used in many military and civilian applications and is considered an essential utility of the 21st century.

The system works by sending one-way ranging signals from a constellation of Earth-orbiting satellites. User receivers passively receive the GPS signals and use their spread-spectrum pseudo-random number codes to measure the time of flight of the signal, but with a bias due to receiver clock errors. This information is combined with information about the satellite positions and clock corrections, with the latter information having been decoded from the received signals. Measurement model equations are derived and solved in order to determine the 3-dimensional user position to within about 10 meters and absolute time to within about 30 nsec anywhere on or near the surface of the Earth. Advanced versions of the system can determine the relative positions of 2 or more receivers to cm-level accuracies over baselines as long as 10 km or more.

The success of the GPS has motivated the European Union to develop an equivalent system called Galileo. Within the next 3 to 5 years, the completion of the Galileo system and planned upgrades to the GPS will give rise to further improvements in the accuracy, availability, and integrity of Global Navigation Satellite Systems. An important open question concerns how the GPS and Galileo will coexist. Each system has reasons to try to dominate the GNSS field. It is possible that one of the two systems will win a competition while the other fades out of existence. Another possibility as that both systems will flourish side by side. The latter result is the one for which GNSS users should hope because it will yield the best level of service.

Acknowledgements

The author wishes to thank four colleagues who have voluntarily provided information and materials for use in this paper. They are Bradford W. Parkinson, Paul M. Kintner, Jr., Steven P. Powell, and Todd E. Humphreys.

References


