A BRIEF SURVEY OF MODERN NAVIGATION

P. M. Janiczek
U. S. Naval Observatory, Ret.

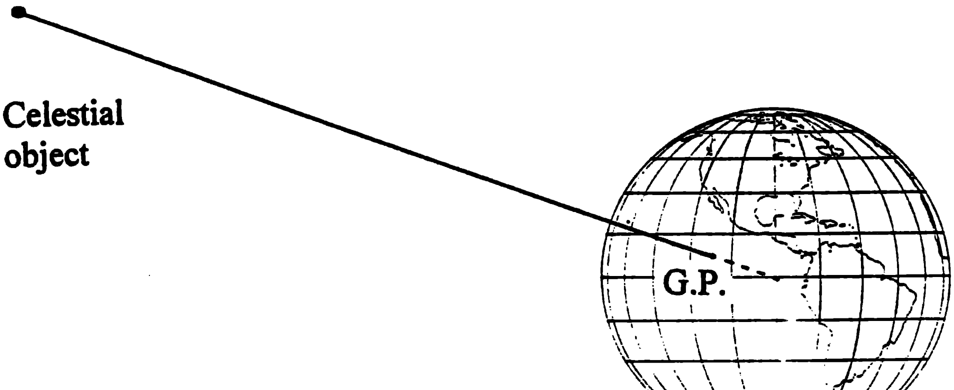
Introduction

There are many forms or systems of navigation that may be called modern. To qualify as a brief, a survey must be based on some limiting criterion. Here I propose first to consider some basic principles associated with all modern navigation and then to limit descriptions of individual navigation systems to a very few that fit into the first of two broad divisions. The first division is absolute navigation, wherein present position is known in relation to an overall coordinate system (latitude and longitude, for example). The second is relative navigation, wherein present position is known relative to some local, special coordinate or grid system. The difference between divisions may be thought of in terms of global versus local. As an example that qualifies as absolute, the Global Positioning System (GPS) is probably the best known. Using road maps or landmarks are everyday examples of relative navigation, as are Very high frequency Omnidirectional Range (VOR) systems used by aircraft as highways in the sky.

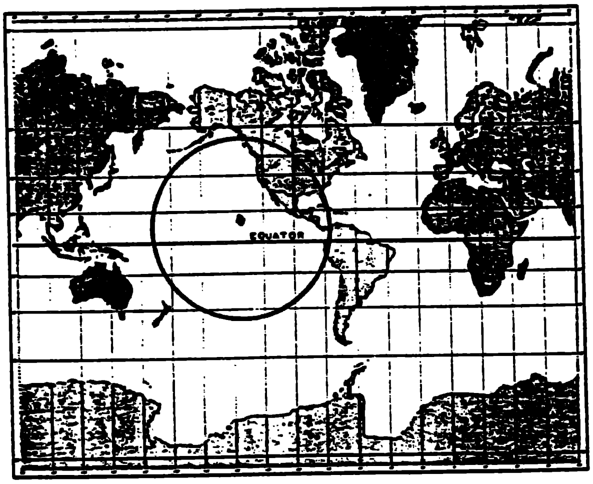
Before continuing, a few points should be noted.

- While navigation involves directing vehicle motion safely and efficiently from one location to another, key words in both absolute and relative navigation are "present position is known." Therefore, it is understandable that effort to devise methods, devices and systems has been concentrated on determining vehicle present position.
- For modern systems, the distinction between absolute and relative can become blurred. Soon, for example, we will likely find GPS officially approved and used for both relative and absolute navigation.

As representative of modern navigation by means of modern systems, I will describe briefly and with some arbitrariness, Loran-C, Omega, inertial, Transit and GPS. Omega and Transit are discontinued. They are nevertheless modern and I will describe them as still active.



a. Geographic position (G.P.) of celestial object.



b. Position line (circle).

c. Intersecting position circles.

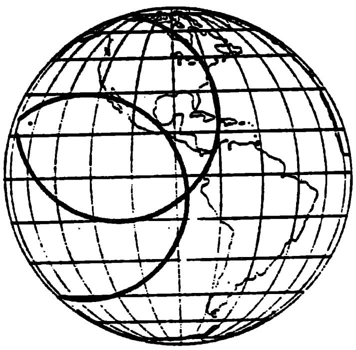


Fig. 1. Position finding by angle observations of celestial objects.

Basic Principles

A few basic principles are common to all modern navigation. It is worthwhile to examine those principles in a little detail.

The first principle is that of a position line. It was first introduced into open ocean navigation in 1837, by way of celestial navigation. To understand a position line requires only the ability to visualize some simple geometry on a sphere.

In Figure 1a, a line is drawn from the center of the Earth to a celestial object (Sun, Moon, star, planet). The line intersects the surface of the Earth at a point. At that point on the Earth, the object is directly overhead, and that point is called the geographic position (G.P.) of the celestial object. The geographic positions of celestial objects, especially those useful to navigation, can be calculated to very high accuracy for any specific time, well in advance of that time.

In celestial navigation practice, a device to measure angles is used to determine the angular distance of a celestial object from the horizon. At the instant the measurement is made, the object has a definite geographic position, as described above. The measured angle then defines a line on the Earth; the line having the property that at any point on it the celestial object will have the same angular distance from the horizon. We call the line a line of position. It has another interesting property. It closes on itself to form a circle. The importance of the position circle is the fact that the navigator's position is somewhere on that circle. Figure 1b shows the position line, or position circle. It also shows lines of latitude and longitude, and it can be seen that the position line intersects any number of latitude and longitude lines. As a result, a navigator needs additional information in order to determine known position. The additional information comes from making an angular measurement of another celestial body to produce a second position line, as shown in Figure 1c. The two position lines intersect in two places and the navigator's position is at one of the intersections. With celestial objects carefully chosen, the navigator can produce large position circles (lines) that also intersect at large angles, and can thereby decide which intersection actually represents his position.

The second principle is the precisely known, constant velocity of electromagnetic radiation in all directions in a uniform medium. It is a physical constant expressed as miles per second, or as kilometers per second. The fact that radio waves, in particular, travel at constant speed, and do so in all directions is basic to modern radio navigation. In simplest

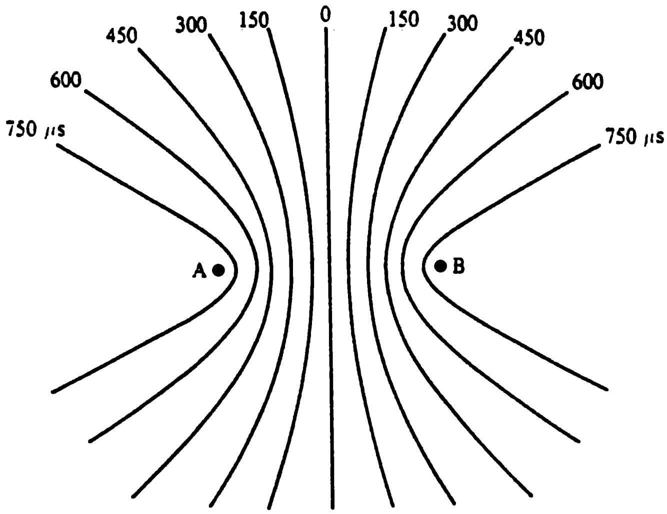


Fig. 2a. Hyperbolic position lines generated by measured time differences.

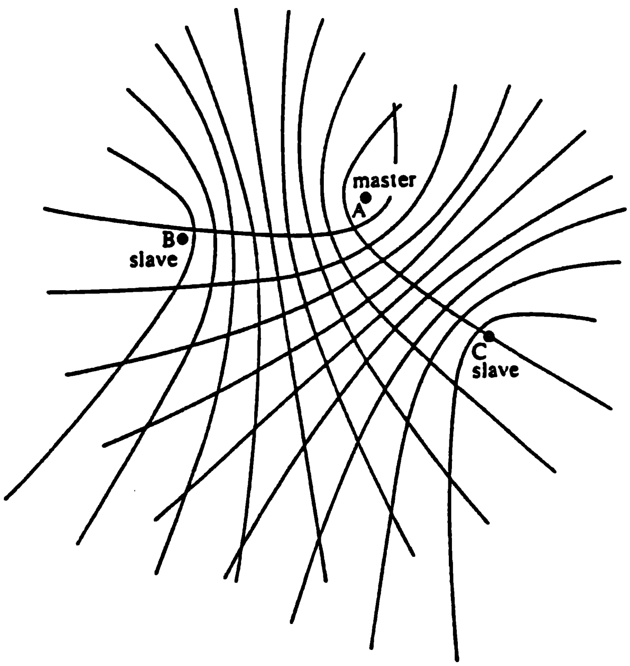


Fig. 2b. Two hyperbolic patterns obtained from three transmitters.

form, a radio station transmits a brief signal. A navigator has a receiver and an accurate clock that is synchronized with the clock at the radio station. By using the clock to know when the radio signal was sent, and by measuring the interval of time for the signal to arrive at the receiver, the distance to the radio station can be determined. However, the radio station sends the signal in all directions. Consequently, knowing the time interval, or distance, simply places the navigator on a position line similar to what was seen as basic to celestial navigation. The only difference is that the radio transmitter is physically located at the center of the position. Additional information is required and is supplied by a second transmitter, at some distance from the first and with its clock likewise synchronized. Again, the navigator's position is determined by the intersections of the position circles. In straightforward application, this method requires receivers coupled to atomic clocks. Atomic clocks are very expensive, need periodic calibration and, ultimately, replacement.

A simple technique avoids the need for an atomic clock in every receiver. Two radio stations transmit precisely synchronized signals. Then, instead of measuring absolute time of arrival, a navigator only needs to be able to measure the difference between the arrival times of both signals. Fortunately, by using relatively inexpensive clocks (or oscillators) in navigation receivers, this is possible. But, in this case, a constant time difference between two signals, rather than precise times of arrival, locates the navigator somewhere on a position line that is an hyperbola, as in Figure 2a (time differences are labeled in microseconds). A third transmitter, also synchronized, is required. The receiver can then measure a time difference between the third signal and either the first or second signal. This gives rise to a second hyperbola representing constant delay time difference. As in the case of intersecting position circles, the intersecting hyperbolas (Figure 2b) can determine a unique latitude and longitude. Used in this way, Loran-C and Omega qualify as hyperbolic systems. Clocks in receivers do not have to be synchronized with those at the transmitters, but only need to be stable for short intervals of time.

It has been said that signals from radio navigation transmitters travel outward in all directions. When considering the measurement of time delays of signals from far above Earth's surface, the geometry of position lines becomes what mathematicians call hyperboloids. We need not explore that geometric fact here, but the omnidirectional characteristic has made possible the use of hyperbolic radio navigation aboard aircraft as well as on Earth's surface.

When radio navigation transmitters are located in spacecraft, rather than on the ground, there is no great distinction to make in regard to geometry, only a reorientation. The signals again travel outward from a spacecraft in all directions. For navigation on the Earth's surface, the distance to a satellite is measured by determining the time taken for a signal to leave the satellite and arrive at the navigation receiver. In this case, a single distance measurement to one satellite locates the receiver on a line of position of the exactly same type as encountered in celestial navigation. The main distinction is one of angular measurement versus distance measurement. As in celestial, it takes a separate measurement to a second satellite to create intersecting position lines that determine position. The technique to overcome the problem of maintaining a very accurate clock within a navigation receiver is discussed below.

Another physical principle that has been basic to a navigation system is the Doppler effect. Simply stated, the frequency shift of received electromagnetic radiation depends upon the relative motion of the source, the receiver, or both. The geometry leading to a determination of position is not intuitive, and use of the principle as the basis for navigating ships or aircraft is practical only if the source of the waves, specifically radio waves, has sufficient velocity to cause an easily measured frequency shift. An artificial satellite answers the need. The U. S. Transit and Russian Cicada navigation systems were constructed to use the significant frequency shift produced in radio signals transmitted by orbiting satellites, but primarily for ship positioning.

All of the principles described so far, when applied to navigation, have in common an external source of radiation, whether the light of a celestial object or radio signals. A system that does not depend on external sources is desirable for several reasons. Such systems have been referred to as self-contained. The most familiar self-contained systems make use of the principle that I state simplistically as: the axis of spin of a spinning rigid body always points to a fixed point in inertial space, absent external forces. An obvious example of such a spinning body is a child's toy top. As adapted for navigation the spinning body is called a gyro. The ensemble of gyro, required sensors, mounting, etc., is called an inertial system. When the gyro is located in a vehicle that is in motion, forces act on it. The magnitude, direction of the forces, and length of time that they act are sensed and measured. The measurements can be used either to apply forces that counteract vehicle motion or restore the orientation of the gyro, or they can be converted to indicate changes in vehicle position and

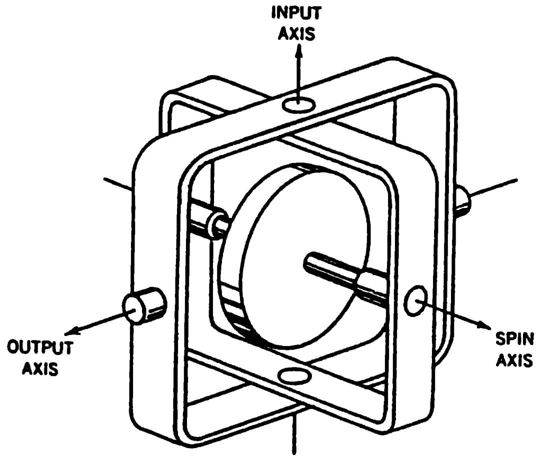


Figure 3. Two-degree-of-freedom gyro without sensors, mounting, etc.

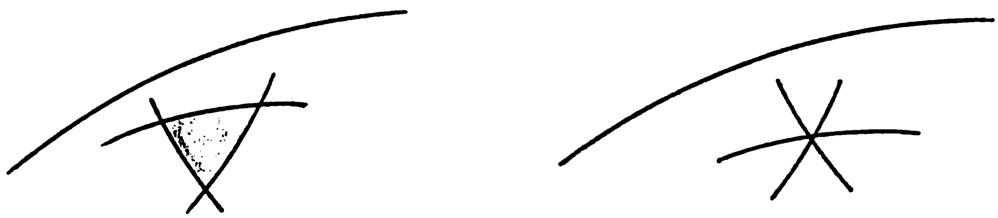


Figure 4. Intersections of three position lines before and after clock corrections. See descriptions of Omega and Global Positioning System.

displayed to the navigator. Figure 3 is a simplified diagram of a two-degree-of-freedom gyro, without sensors, to illustrate the basic device.

The great importance of inertial navigation arises from the fact that an inertial device responds to only to forces acting upon its host vehicle. It is completely self-contained and independent of external signals such as light or radio waves. Because of that importance, inertial systems are now to be found not only on ships and submarines, but also in aircraft, rockets, spacecraft and even in land vehicles. The Ship Inertial Navigator System is one example; the series of Carousel units found in aircraft is another.

During the last 30 years inertial devices have been built that do not use a mechanical spinning body. However, they do retain the "self-contained" characteristic and can operate without reference to stars or radio waves.

Modern Systems

Despite the simplicity of the underlying principles described above, any attempt to implement a system brings additional physical principles to bear. In the real world, neither light nor radio waves travel in straight lines or with unchanging speed when the mediums through which they pass differ. The actual shape of the Earth changes the elegant position circles and hyperbolas into more complex figures. Measuring instruments introduce errors, etc. Table 1 is provided as an indication of physical, geometric and other problems that must be accommodated by modern navigation systems. Problems inherent to celestial navigation are included for comparison. The table is an admittedly incomplete compilation. Space does not allow definition and discussion of every tabular entry; nevertheless some general points should be made.

- The entries are a mix of phenomena and problems that have been completely or partly overcome by the systems, or remain.
- The appearance of the same words in more than one column does not necessarily represent the same problem. For example, the ionosphere affects Omega in a different manner than it affects GPS or Transit. Also, refraction is considered and dealt with as it affects satellite signals differently than in case of celestial.

Loran-C

Loran-C radio stations broadcast precisely structured and timed signals. To create the geometry of useful hyperbolic position lines, three stations, separated by hundreds of miles are required. One station is designated Master, the remaining two or more are called Secondaries, or Slaves. A group of such stations is referred to as a Chain. The master station signal is a sequence of nine pulses. The signal from the master station is received by a navigator and by the secondary stations. After a short, defined time interval, one secondary transmits the same signal, minus the ninth pulse. Other secondaries continue the pattern until all stations in the chain have transmitted.

One of the primary qualities of a Loran-C signal is its pulse shape. The pulse shape, combined with a technique called phase reversal, allows a receiver to reject unwanted signals reflected from the ionosphere. Such skywave signals would otherwise confound a receiver's conversion of time differences to position lines. The pulse shape must be carefully structured at the transmitter and the receiver must be able to identify it and select the third cycle within a pulse.

A state of the art Loran-C navigation receiver will perform several functions. Among them are: automatically locate and track the selected master and secondary stations, automatically measure time differences, indicate when a signal is lost, attenuate interfering signals, convert time differences to lines of position, and display latitude and longitude.

At this time, all respectable receivers incorporate corrections for the primary and secondary phase factors. But the additional phase factor can still cause problems. It arises from the passage of signals over terrain composed of both land and water. Further, it is a seasonal effect. Consequently, corrections incorporated within receivers may not be accurate, especially when operating within 10 miles of a coastline.

Loran-C can provide a user with position accuracy of about 0.25 nautical mile. The system is useful to a distance of 1200 nautical miles (nmi), but waves reflected from the ionosphere can increase coverage to 2300 nmi with a reduced accuracy. The system has been so successful that chains have been built to cover most of the Northern Hemisphere.

Omega

Omega (nominally 10 kilohertz) signals can reach virtually any location using only eight transmitters. Worldwide coverage is obtained by taking advantage of the fact that very low frequency radio waves tend to follow Earth's curvature and can be received at enormous distances. Omega position determinations are not as accurate as with Loran-C, but adequate for enroute navigation on or over open ocean. Each Omega transmitter operates independently of the others, but the transmissions are

Table 1. Physics and Geometry Affecting Performance and Accuracy

Loran-C	*Omega*	*Inertial*
Phase:	Signal propagation mode:	Alignment
Primary	attenuation rate	Bias
Secondary	excitation factor	Coriolis "acceleration"
Additional	phase velocity	24 hour oscillation
Ground vs. sky waves	Ionosphere:	Schuler oscillation: 84 m.
Station (clock) sync	day / night	Gravity anomalies
Envelope / cycle match	sudden disturbance	Vehicle:
Crossing angle	polar cap disturbance	roll
Interference	latitude	pitch
Receiver:	Ground conductivity:	yaw
oscillator	normal / ice caps	acceleration
3rd cycle ID	Arctic	Analog / digital converter
cycle match	aurora	Reset / update
other circuitry	Antipode phase confusion	Other electronics
Fix ambiguity	Geomagnetism (East-West)	
	Lane slip	
	Lane ambiguity	
	Receiver quality	

synchronized by atomic clocks at the stations. Every station transmits on four common frequencies, and each also transmits on its own unique frequency. No two stations transmit on the same frequency at the same time, so there is no overlap. But because of the multiple frequencies and stations, information flowing to an Omega receiver is almost continuous. At any receiver location and time, most of 40 possible signals are useable. Basically, a receiver measures a phase difference between signals from three or more stations to produce hyperbolic position lines and a position.

However, it is also possible to use Omega in direct ranging mode, also called range-range mode. For this mode, a receiver generates a reference signal that replicates the actual Omega signal. As the real signal is received, the replica is shifted in phase until it coincides with the real signal in the receiver circuitry. The phase shift is equivalent to a time interval, which equates quite simply to distance as

$$\text{Distance} = \text{velocity of light (radio)} \times \text{time interval.}$$

The distance measurement is interpreted as defining a circle of position, as described above. Repeating the process using a signal received

Table 1. *Continued*

Transit	*GPS*	*Celestial*
Refraction:	Refraction:	Personal equation
ionosphere	ionosphere	Sextant: (7 error sources)
troposphere	troposphere	Height of eye
Height:	Height:	Refraction:
geoid at vehicle positions	geoid at vehicle position	air temperature
vehicle antenna	vehicle antenna	atmos. pressure
Chart congruence	Chart congruence	inversions
Velocity:	Multipath reflection	Cloud cover
satellite	System clocks	Horizon:
vehicle	Ephemerides quality	night
Earth	Message quality	false
Position estimates	Satellite geometry	Object:
Satellite crossing	Receiver type:	semidiameter
Ephemeris quality	access capability	phase
Message quality	correlation ability	parallax
Satellite geometry		Time
Receiver & computer		Estimated position
		Geoid
		Calculation accuracy
		Position line geometry

from a second station defines a second position circle. Ideally, the receiver must be located at one of the two points where the position circles intersect.

Two measured distances and position circle intersections are suspect. That is because the velocity of light is so great that a very small clock error can produce a large distance error that carries over into latitude and longitude. A simple technique borrowed from long-standing celestial navigation practice takes care of the situation. A range to a third station is measured and a third position circle is generated. The three position circles do not generally intersect at a point (Figure 4). The receiver clock is then adjusted by a constant amount, which causes changes to the phase values, distances and position circles. This adjustment process is repeated until all position circles do intersect at a common point, hence an accurate position.

State of the art technology has benefited Omega receivers, as it has for Loran-C. All of the circuitry and computations implied by the above descriptions, from signal reception to direct display of latitude and longitude, can now be carried out automatically. It is no longer necessary, for example, to plot lines on specially printed charts when using either system.

Inertial

For many years the spinning gyro, as a navigation device, had the limited capability of indicating heading of a ship. Called a gyrocompass, it could be considered a navigation aid. An inertial navigator requires some mechanism to sense the forces that produce vehicle motion and to integrate those to estimate changes in speed and position. The sensing element is usually called an accelerometer. Integration is accomplished by electronics. To endow an inertial system with the ability to indicate position and speed of a vehicle accurately and reliably requires very considerable effort and ingenuity. Part of that effort must be directed to methods of stabilizing the gyro itself. With few exceptions, modern inertial systems are usually a part of an integrated system in which a radio navigation system provides periodic updates. The state of inertial development is such that no implementation can or should yet stand unaided for long periods.

Transit

The Transit system was mentioned earlier as an example of applying the Doppler effect in navigation. In simple terms, a navigator with a radio receiver acquires the signal from a satellite and measures the frequency shift caused by the satellite's high velocity. This must be a repetitive process so that changing frequency can be related to slant range changes to the satellite and the resulting data accumulated. By itself, the collection of range differences tells little. It is necessary for the navigator, his computer more precisely, to have the satellite's orbital positions when the Doppler measurements are made. It is also necessary to have approximately known position and motion for the receiver as well. The message transmitted by a satellite contains values for the parameters that define its orbit, so that its position relative to the receiver can be computed as the satellite makes its pass.

In contrast to other navigation systems, a position in latitude and longitude is determined for the navigator by use of one satellite and its signal. The drawback is that it takes between 10 and 16 minutes to determine a present position. During this interval, and to determine the navigator's position, his computer combines calculated satellite positions, range difference measurements (counting Doppler cycles) and information regarding vessel motion.

Global Positioning System

The Global Positioning System combined some of the best known, basic principles and techniques of navigation with innovations to become a modern system that should meet most military and civil requirements for accurate position determination for a long time. For a navigator on Earth's surface, the circular line of position reappears in this system because distance measurements are made between receiver and orbiting satellites. Distance measurements to two satellites produce two circles that intersect, one intersection being at the navigator's position. As in the Transit system, each GPS satellite provides the navigation receiver with a message. In GPS, the message contains the data necessary to calculate the satellite's position, but also includes clock correction parameters and a parameter that permits an approximation to be made for atmospheric delay of the signal. It also contains a reduced accuracy 'almanac' containing similar information for all other satellites in the system.

For a receiver to operate successfully, it has to have an internal replica of the satellite's timing signal. It must also be able to compute the satellite's position using the data in the satellite message. As the satellite signal is received, the replica signal is shifted in phase to agree with the satellite signal. The phase shift is equivalent to a time interval, which equates simply to distance as described in connection with the Omega ranging process.

Each of the satellites contains an atomic clock carefully maintained in synchronization with a GPS system time, which is, in turn kept in step with the master clock at the U. S. Naval Observatory. As mentioned above, it is simply too expensive and logistically impossible to have such a synchronized clock in every receiver. Distances found by the above process, and the intersection of two position circles, would be inaccurate, at least to the extent that even a stable receiver clock is not quite accurate enough for the complete task. A third satellite distance is computed from measured phase shift. Again, the distance translates to a third line of position that does not generally intersect the first two lines at the same point. By repetitively adjusting the receiver clock by a constant amount (assuming a constant clock error) followed by recomputing each distance until the three lines of position all do intersect at one point, the correct, pinpoint position is determined.

The case of an aircraft in flight introduces a third dimension to the problem, its altitude above the Earth surface. Extension of the above operational description to this case is straightforward. Instead of measuring the distance to two satellites, the receiver must capture and measure signals from three, and compute their positions as well. For this case, the measured distances define radii of spheres, each sphere centered at a satellite. Theoretically, the common point of intersection for the three spheres determines altitude as well as latitude and longitude. But again, it is not to be expected that all of the spheres will intersect at a point, so a fourth satellite is used. As above, adjustments are made to the receiver clock until a pinpoint intersection is found. The number of active satellites in the GPS system is 24. For an aircraft and ships at least, there are always a sufficient number of available satellite signals to carry out the positioning process.

Since GPS broadcasts on two frequencies, the difficulty of propagation delay downward through the ionosphere is virtually eliminated by applying a formula that relates delay to the mathematical squares of the two frequencies. This technique was also used successfully by the Transit system. A tracking network and frequent uploads of orbit ephemerides also maintain available accuracy for the navigator. Use of a special coding in the navigation message (called pseudo random noise) combined with high frequency transmission enable the satellites to conserve power, and the user to access and process the signal reliably in the presence of considerable noise.

To some extent the high accuracy of the system can be a problem. Full accuracy has made necessary the compilation of a geodetic reference system that is commensurate, so that coordinates derived from GPS are correctly related to chart positions based on the same geodetic system. Not being aware of the situation, some navigators have placed exclusive, blind faith in a combination of GPS and charts based on a regional or an outdated geodetic datum, ultimately to find their vessel in peril or grounded.

Since the requisite messages for both GPS frequencies are not available to all, civil use accuracy does not match the full accuracy. Nevertheless, attainable accuracy using only the civil availability frequency is more than enough for the majority of users. On the other hand, with a special receiver that can reproduce and track the carrier frequency of the GPS signal, it is possible to obtain a precision of about 2 millimeters. While precision is not the same as accuracy, this tracking technique does have applications in surveying and geophysics.

An Eclectic History

Table 2 lists some dates that were particularly significant relative to transforming radio signals into simple lines of position and instantaneous indication of latitude and longitude. Also included, set apart, are four approximately dated periods during which significant research was being carried on that had direct impact on the ability to field modern systems. All modern systems that we have at present have been made possible by a combination of research, development and engineering in physics, radio circuitry and wave propagation, digital computer science, space science, materials science and microfabrication. Not revealed by Table 2 is the course of political events, particularly during the last 60 years, that had as much or more influence on the specific array of navigation now at our disposal. In what follows I briefly mention some particular events in the history of the few navigation systems I discussed above. That is not to say a very complete account would be misplaced or boring; rather, such a discussion is best left to the history of science.

Loran-C shares a name with Loran-A, but it is a different system. Work to improve Loran began in 1943 but, for a long time, the thrust of almost all research in radio was at higher and higher frequencies. There were problems with interference, bandwidth and propagation that had to be addressed. Utility of the low frequency, long wave part of the spectrum, once heavily used for long distance maritime communications, seemed to hold little interest. The final selection of the 90 – 110 kilohertz band for Loran-C resulted from the study and experimentation of relatively few people. Meanwhile, many experimental systems, with names largely forgotten, were tested and discarded.

As a direct benefit of three patents issued between 1974 and 1980, most Loran transmitters are now solid state, with the benefits of reliability and economy. Reliability is further enhanced by redundancy, and it is possible for stations to operate virtually unmanned.

Military requirements for Loran-C ended in 1994. However, at that time Loran-C could boast the highest number of users of any precise radio navigation system, and the number of users was continually growing. The system should remain viable for several years, even though many users will shift to GPS.

Table 2. Brief Chronology of Navigation Related Events

1837 – line of position by celestial navigation discovered
1887 – electromagnetic (radio) waves produced
1895 – transmission and reception of radio waves demonstrated
1896 – theory of the ionosphere proposed
1897 – electron discovered
1904 – first broadcast of radio time signals
1907 – triode vacuum tube announced
1922 – idea of hyperbolic radio navigation patented
1925 – height of ionosphere measured by pulse ranging
1937 – first hyperbolic navigation system (Gee) proposed
1942 – Gee system operational
1942 – Gyro system used to stabilize rockets
1943 – Standard Loran (later Loran-A) operational
1943-58 – research to improve Loran
1946 – first electronic digital computer
1947 – Radux system proposed
1947-66 – research on radio wave propagation, systems tested
1948 – invention of transistors
1951 – Ship Inertial Nav. System development initiated
1953 – highly stable crystal oscillator for radio frequencies
1955 – hybrid Radux-Omega system studied
1955-58 – stability of cesium beam frequency standards demonstrated
1956 – inertial navigation system installed aboard ship
1957-pres. – space research
1957 – Doppler shift of Sputnik radio signals analyzed
1958 – Ship Inertial Navigation System deployed
1958 – Loran-C operational
1958 – Transit (Doppler) system approved
1962 – Radux abandoned, Omega system development pursued
1964 – Transit system operational
1970-pres. – microelectronic development
1971 – Omega system operational
1972 – pseudorandom noise code ranging signal demonstrated
1973 – GPS development initiated
1993 – GPS fully deployed (24 satellites)
1996 – Transit updates discontinued
1997 – Omega transmitters turned off

The major problem with current inertial navigation is the tendency for existing systems to accumulate errors fairly rapidly and pass those on as erroneous indications of position and speed. Consequently, periodic position input is required to reset or update the inertial system. Updated position input must be obtained from other systems, and the frequency with which that is required equates to more or less autonomy for the navigator and his vehicle.

Inertial navigation is only one of several uses for inertial systems, and research to create better inertial systems has some interesting aspects. To a large extent, focus has been away from spinning wheels and their mountings and toward use of other physical principles and dependence on computers. New devices, some having no gyros or gimbals are still referred to as gyros. They are also known by esoteric names such as electrostatically supported, hemispherical resonator, ring laser and interferometric fiber optic gyros. The last named two are optical gyros, based on a general relativistic effect. Impetus is another aspect. Navigation being but one application of inertial devices, many improvements and innovations appear to be driven at present more by demand for smaller size, less weight and cost to own, and less by a quest for greater accuracy and reliability. It may be a long time until an inertial, or other self-contained system, fulfills all navigation criteria or competes with radio based systems.

The U. S. Navy originally developed Transit to update the inertial navigation system aboard Polaris submarines. Transit began operation in January 1964 and, because of the accuracy it afforded in determining position, it has been credited with giving birth to the science of satellite geodesy.

Transit satellites were exceptionally reliable and, when the last satellites were placed in operation, the system reached peak capability and had about 100,000 users. Operation of the system has been discontinued in favor of the Global Positioning System, but the Russian Cicada system, also operating on the Doppler principle, continues.

Except for a few requirements, the Global Positioning System offers all the performance and accuracy that the military needs. To everyone else, it appears to be the best of all systems. In fact, it is now evolved to the position of creating its own requirements in the commercial arena.

GPS was almost cancelled at one point and was degraded at another. A joint program for development of GPS (also called NAVSTAR then and at times since) came before the Defense System Acquisition Review

Council (DSARC) in 1973. The program, as presented, was essentially an Air Force system having some undesirable aspects. The review failed but, fortunately, strong support was expressed for a broadened system concept that would represent the views and requirements of all services. Work then began on synthesis of the best of all extant concepts for a system.

In 1979 another DSARC review gave approval for continued development of the system with its 24 satellites as planned. However, at a higher level review, at which money was the sole unit of measurement, it was decided to reduce the GPS constellation to 18 satellites. This triggered studies for the purpose of redesigning the orbit plane configurations for the administratively revised system. What seemed the best alternatives would, it appeared to me, demand perfect reliability of orbit insertion and operation for all 18 satellites. Further, there would be some locations on the Earth for which the geometry of available satellites would at times be unfavorable, particularly for aircraft. Fortunately, the 24 satellite constellation was reinstated.

The Cardinal Principle

To this point, I have described a few navigation systems particularly as they illustrate the use of some basic principles. There is one principle, neither geometric nor physical, as old as navigation beyond familiar landmarks and still applicable to modern navigation. The principle itself is simple but cardinal: A navigator should use every available means to determine his position.

A navigator who must steadily rely on an inertial system adheres to the principle whenever he updates, or resets, the inertial system by using an external source. Of course modern navigation is heavily reliant on electronics, and electronics is readily adaptable to combining two or more navigation system signals in various ways to provide an optimum result. In fact, this has been done many times. A simple, obviously obsolete example is the combination of Omega and Transit. Omega as a global system could provide a position almost continuously (every 10 seconds), with successive positions enabling a calculation of estimated speed. Transit provided a more accurate position, but on an irregular basis. Transit also required an estimate of position and speed as input to its data processing. By combining the two systems electronically, at the receiver, Omega could be considered the primary system, with accurate Transit positions used to minimize the errors in Omega. In this context the system becomes a global Differential Omega system. Complementary systems,

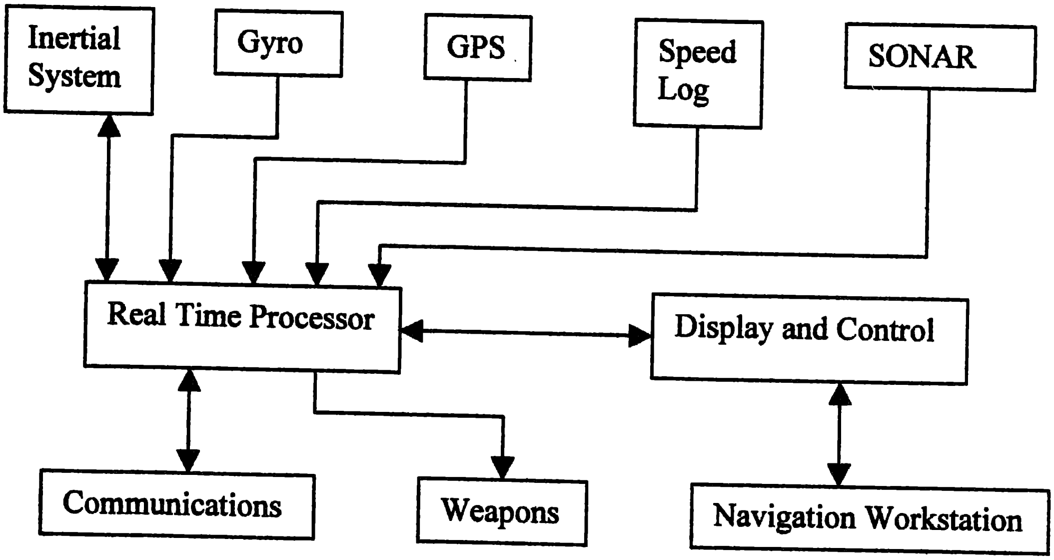


Figure 5. Elementary integrated navigation system

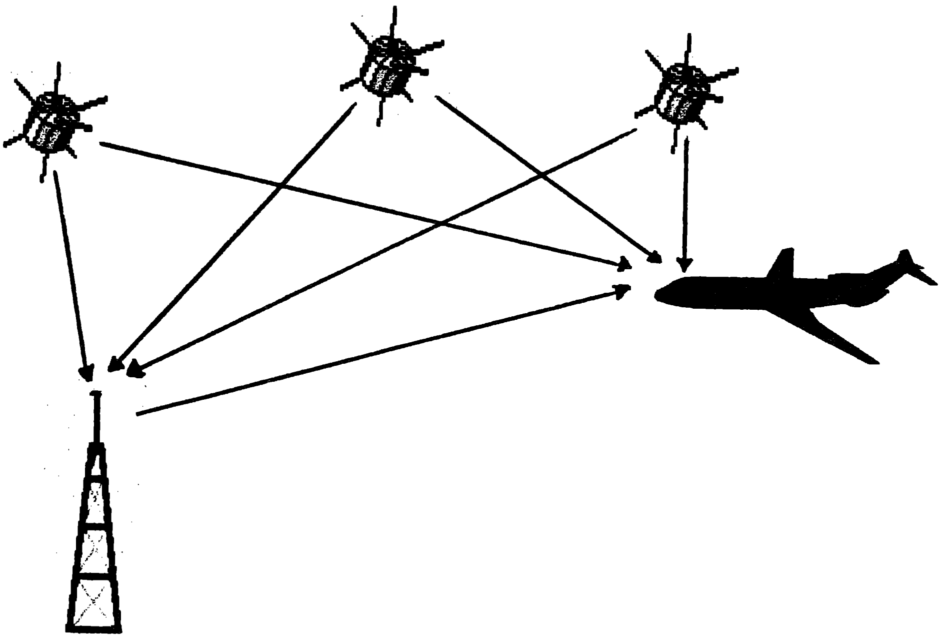


Figure 6. Differential (GPS) navigation.

such as described and whether proposed or built, have been variously called composite, hybrid, and integrated.

The GPS is continuously available worldwide, so that it is natural to ask whether it is any longer worthwhile to consider integration with another system. A positive answer is derived from considering issues of vulnerability and of operations in locations where GPS signals are blocked by natural or artificial objects. Taking the issues into account, the U. S. Navy, for one, has been developing massively integrated systems for use aboard ships. These systems are complex and evolving. However, Figure 5 is a block diagram of an elementary system that illustrates how integration could proceed for a combat vessel. Note that all navigation signals are directed to the real time processor, which quickly combines the inputs according to their relative accuracy and reliability. Also note that the processor provides updates to the inertial system. The "Gyro" block refers to the gyrocompass, which displays direction only. The navigator can adjust or override the processor using the display and control unit.

Aside from the benefits of integration, there is a situation that calls for a different solution. From the beginning of GPS development, it was intended that its highest accuracy capability would be withheld from all but authorized users. For GPS civil use, especially on or over vast stretches of terrain, the restriction is of little consequence. There are however, areas of operation by aircraft and ships in which highest accuracy is needed. It can be obtained by an investment in a differential system. The idea is conceptually simple and illustrated by Figure 6. A GPS receiver is monitored at an accurately known position. Any positions determined at the receiver that show differences from the known position are considered the result of errors in the system; in particular, signals that don't provide full system accuracy. It is also considered that those deviations from the known position are exactly the same anywhere in the vicinity of the known position. A transmitter at the accurately known location proceeds to broadcast a message quantitatively informing all GPS users in the vicinity what the deviations are. GPS users, on receiving that message, then have the opportunity to apply the deviations as corrections to positions determined directly from the GPS satellites. This Differential GPS concept has been extensively tested, automated, and found very successful.

A Last Word

The preceding paragraphs illustrate that absolute navigation is concerned primarily with determining present position and is based on straightforward geometric and physical principles. Also, from the history outlined by Table 2, it is seen that modern systems had to await developments in science and technology in order to be built and to overcome, or compensate for, additional physical and geometric effects. As a result, the equipment that constitutes a modern system is not only highly complex, but any such system requires an extensive, supporting infrastructure also. Complexity and extensive infrastructure render modern systems vulnerable, and something must be said about alternatives.

There are numerous possibilities for complex, high-tech systems to fail or become unavailable. Any practicing navigator with experience knows the value of having alternative methods of navigation at hand. At least one alternative must be independent of primary systems; it would be best if it has no point in common with a primary system such that failure at a common point eliminates both alternatives. The most obvious example of a common point failure is an electronic suite or integrated system in which all component devices depend on a single electric power source. The example is easily understood by considering Figure 5. Additional examples could be cited.

Until Loran-C became widely available, celestial was the standard of excellence for determining position in deep water. Celestial navigation is both an alternative and modern in the sense that it is still available in classic form and procedures when other systems cannot be used. Precisely because of its classic procedures, it lately stands as a weak backup to electronic systems. With the exception of the rigorous and flexible computer program STELLA (System To Estimate Latitude and Longitude Astronomically), there has been no attempt to automate the process of celestial aboard ship. Further, the extension of celestial capability to 24-hour capability has not gone beyond successful demonstrations of feasibility. In contrast, there are many applications of modern technology to at least some elements of celestial navigation in the form of automatic star trackers for missiles, spacecraft and long-range aircraft. I believe that an attempt to automate the total process of celestial for surface vessels is possible and reasonable, certainly to proceed along lines that maintain independence from a primary or other alternative system. For instance, power consumption by an automated celestial system ought to be minimal so that battery operation is a consideration. Other symposium participants address this subject more fully.

Acknowledgements

I am indebted to Dr. George H. Kaplan for providing some critical details and discussions, and for reading an early version of the manuscript. Since I was, for a time, editor of the journal *Navigation*, and was engaged in the production of navigational almanacs for a longer time, some of the preceding material is from personal knowledge of the subject. For many quantitative and historical details, I have also relied on a few references that should be mentioned.

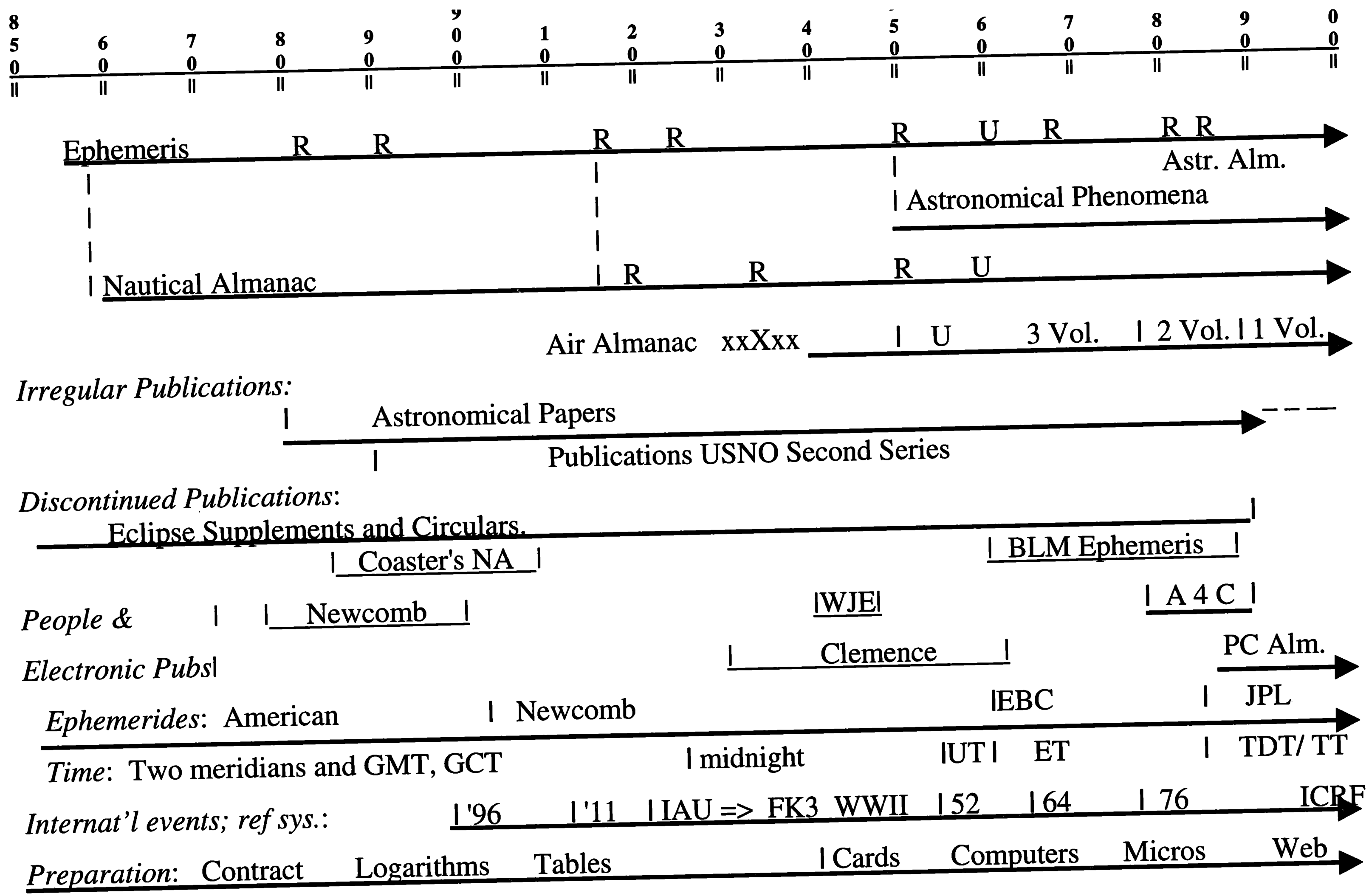
1. Except for too much emphasis on radio direction finding, S. F. Appleyard, *Marine Electronic Navigation* (Routledge & Kegan Paul, London, Boston and Henley, 1980) contains very readable discussions of basic principles and the operations of some modern navigation systems. Figure 2 in this paper was adapted from that book.

2. Unsurpassed for first-hand accounts of the development of modern systems (to 1995) are two anniversary issues of *Navigation*: Vol. 18, No.1, 1971 and Vol. 42, No.1, 1995 (Alexandria Va.: The Institute of Navigation).

3. The standard reference for an introduction to the engineering aspects of GPS is *Global Positioning System*, Vol. I (The Institute of Navigation, 1980). Advanced concepts and applications are found in succeeding volumes in the series.

4. The many aspects of applied marine navigation (to 1977) are fully explored in two volumes of the *American Practical Navigator* (Washington: The Defense Mapping Agency, Pub. No. 9, 1977).

5. An extensively revised edition of *The American Practical Navigator,* containing much new material, was published in 1995. The National Imagery and Mapping Agency is currently responsible for this standard reference.



Abbreviations:
R = Revision
U = Unification
X = Experimental

WJE = Wallace J. Eckert
EBC = Eckert-Brouwer-Clemence Integration and Brown's Lunar Theory
A 4 C = Almanac For Computers
PC Alm. = Floppy Almanacs, MICA, STELLA, etc.

Figure 1. Timelines for Publications, People, and Major Influences