1

INTRODUCTION

A book on navigation? Fine reading for a child of six![1]

1.1 NAVIGATION

During the European Age of Discovery, in the fifteenth to seventeenth centuries, the word *navigation* was synthesized from the Latin noun *navis* (ship) and the Latin verb stem *agare* (to do, drive, or lead) to designate the operation of a ship on a voyage from A to B—or the art thereof.

In this context, the word *art* is used in the sense of a *skill*, *craft*, *method*, or *practice*. The Greek word for it is *τεχνυ*, with which the Greek suffix -*λογια* (the study thereof) gives us the word *technology*.

1.1.1 Navigation-Related Technologies

In current engineering usage, the art of getting from A to B is commonly divided into three interrelated technologies:

---

[1] Truant Officer Agatha Morgan, played by Sara Haden in the 1936 film *Captain January*, starring Shirley Temple and produced by Daryl F. Zanuck for 20th Century Fox Studios.

Global Navigation Satellite Systems, Inertial Navigation, and Integration, Third Edition.
Mohinder S. Grewal, Angus P. Andrews, and Chris G. Bartone.

- **Navigation** refers to the art of determining the current location of an object—usually a vehicle of some sort, which could be in space, in the air, on land, on or under the surface of a body of water, or underground. It could also be a comet, a projectile, a drill bit, or anything else we would like to locate and track. In modern usage, A and B may refer to the object's current and intended dynamic *state*, which can also include its velocity, attitude, or attitude rate relative to other objects. The practical implementation of navigation generally requires observations, measurements, or sensors to measure relevant variables, and methods of estimating the state of the object from the measured values.
- **Guidance** refers to the art of determining a suitable trajectory for getting the object to a desired *state*, which may include position, velocity, attitude, or attitude rate. What would be considered a "suitable" trajectory may involve such factors as cost, consumables and/or time required, risks involved, or constraints imposed by existing transportation corridors and geopolitical boundaries.
- **Control** refers to the art of determining what actions (e.g., applied forces or torques) may be required for getting the object to follow the desired trajectory.

These distinctions can become blurred—especially in applications when they share hardware and software. This has happened in missile guidance [2], where the focus is on getting to B, which may be implemented without requiring the intermediate locations. The distinctions are clearer in what is called "Global Positioning System (GPS) navigation" for highway vehicles, where

- **Navigation** is implemented by the GPS receiver, which gives the user an estimate of the current location (A) of the vehicle.
- **Guidance** is implemented as *route planning*, which finds a route (trajectory) from A to the intended destination B, using the connecting road system and applying user-specified measures of route suitability (e.g., travel distance or total time).
- **Control** is implemented as a sequence of requested driver actions to follow the planned route.

1.1.2 Navigation Modes

From time immemorial, we have had to solve the problem of getting from A to B, and many solution methods have evolved. Solutions are commonly grouped into five basic navigation modes, listed here in their approximate chronological order of discovery:

- **Pilotage** essentially relies on recognizing your surroundings to know where you are (A) and how you are oriented relative to where you want to be (B). It is older than human kind.

- **Celestial navigation** uses relevant angles between local vertical and celestial objects (e.g., the sun, planets, moons, stars) with known directions to estimate orientation, and possibly location on the surface of the earth. Some birds have been using celestial navigation in some form for millions of years. Because the earth and these celestial objects are moving with respect to one another, accurate celestial navigation requires some method for estimating time. By the early eighteenth century, it was recognized that estimating longitude with comparable accuracy to that of latitude (around half a degree at that time) would require clocks accurate to a few minutes over long sea voyages. The requisite clock technology was not developed until the middle of the eighteenth century, by John Harrison (1693–1776). The development of atomic clocks in the twentieth century would also play a major role in the development of satellite-based navigation.
- **Dead reckoning** relies on knowing where you started from, plus some form of heading information and some estimate of speed and elapsed time to determine the distance traveled. Heading may be determined from celestial observations or by using a magnetic compass. Dead reckoning is generally implemented by plotting lines connecting successive locations on a chart, a practice at least as old as the works of Claudius Ptolemy (~85–168 AD).
- **Radio navigation** relies on radio-frequency sources with known locations, suitable receiver technologies, signal structure at the transmitter, and signal availability at the receiver. Radio navigation technology using land-fixed transmitters has been evolving for nearly a century. Radio navigation technologies using satellites began soon after the first artificial satellite was launched by the former Soviet Union in 1957, but the first global positioning system (GPS) was not declared operational until 1993. Early radio navigation systems relied on electronics technologies, and global navigational satellite system (GNSS) also relies on computer technology and highly accurate clocks. Due to the extremely high speed of electromagnetic propagation and the relative speeds of satellites in orbit, GNSS navigation also requires very precise and accurate timing. It could be considered to be a celestial navigation system using artificial satellites as the celestial objects, with observations using radio navigation aids and high-accuracy clocks.
- **Inertial navigation** is much like an automated form of dead reckoning. It relies on knowing your initial position, velocity, and attitude, and thereafter measuring and integrating your accelerations and attitude rates to maintain an estimate of velocity, position, and attitude. Because it is self-contained and does not rely on external sources, it has the potential for secure and stealthy navigation in military applications. However, the sensor accuracy requirements for these applications can be extremely demanding [26]. Adequate sensor technologies were not developed until

the middle of the twentieth century, and early systems tended to be rather expensive.

These modes of navigation can be used in combination, as well. The subject of this book is a combination of the last two modes of navigation: GNSS as a form of radio navigation, combined with inertial navigation. The key integration technology is Kalman filtering, which also played a major role in the development of both navigation modes.

The pace of technological innovation in navigation has been accelerating for decades. Over the last few decades, navigation accuracies improved dramatically and user costs have fallen by orders of magnitude. As a consequence, the number of marketable applications has been growing phenomenally. From the standpoint of navigation technology, we are living in interesting times.

1.2 GNSS OVERVIEW

There are currently four GNSSs operating or being developed. This section gives an overview; a more detailed discussion is given in Chapter 4.

1.2.1 GPS

The GPS is part of a satellite-based navigation system developed by the U.S. Department of Defense under its NAVSTAR satellite program [9, 11, 12, 14–18, 28–31].

1.2.1.1 GPS Orbits The fully operational GPS includes 31 or more active satellites approximately uniformly dispersed around six circular orbits with four or more satellites each. The orbits are inclined at an angle of 55° relative to the equator and are separated from each other by multiples of 60° right ascension. The orbits are nongeostationary and approximately circular, with radii of 26,560 km and orbital periods of one-half sidereal day (≈11.967 h). Theoretically, three or more GPS satellites will always be visible from most points on the earth's surface, and four or more GPS satellites can be used to determine an observer's position anywhere on the earth's surface 24 h/day.

1.2.1.2 GPS Signals Each GPS satellite carries a cesium and/or rubidium atomic clock to provide timing information for the signals transmitted by the satellites. Internal clock correction is provided for each satellite clock. Each GPS satellite transmits two spread spectrum, L-band carrier signals on two of the legacy L-band frequencies—an L1 signal with carrier frequency $f_1 = 1575.42\,\text{MHz}$ and an L2 signal with carrier frequency $f_2 = 1227.6\,\text{MHz}$. These two frequencies are integral multiples $f_1 = 1540 f_0$ and $f_2 = 1200 f_0$ of a base frequency $f_0 = 1.023\,\text{MHz}$. The L1 signal from each satellite uses *binary phase-shift keying* (BPSK), modulated by two *pseudorandom noise* (PRN) codes in phase quadrature, designated as the C/A-code and P-code. The L2 signal from each satellite is BPSK modulated by only the P(Y)-code. A brief description of the nature of these PRN codes follows, with greater detail given in Chapter 4.

Compensating for Ionosphere Propagation Delays This is one motivation for use of two different carrier signals, L1 and L2. Because delay through the ionosphere varies approximately as the inverse square of signal frequency f (delay $\propto f^{-2}$), the measurable differential delay between the two carrier frequencies can be used to compensate for the delay in each carrier (see Ref. 27 for details).

Code-Division Multiplexing Knowledge of the PRN codes allows users independent access to multiple GPS satellite signals on the same carrier frequency. The signal transmitted by a particular GPS signal can be selected by generating and matching, or correlating, the PRN code for that particular satellite. All PRN codes are known and are generated or stored in GPS satellite signal receivers. A first PRN code for each GPS satellite, sometimes referred to as a *precision code* or *P-code*, is a relatively long, fine-grained code having an associated clock or chip rate of $10f_0 = 10.23\,\text{MHz}$. A second PRN code for each GPS satellite, sometimes referred to as a *clear* or *coarse acquisition code* or *C/A-code*, is intended to facilitate rapid satellite signal acquisition and handover to the P-code. It is a relatively short, coarser-grained code having an associated clock or chip rate $f_0 = 1.023\,\text{MHz}$. The C/A-code for any GPS satellite has a length of 1023 chips or time increments before it repeats. The full P-code has a length of 259 days, during which each satellite transmits a unique portion of the full P-code. The portion of P-code used for a given GPS satellite has a length of precisely 1 week (7.000 days) before this code portion repeats. Accepted methods for generating the C/A-code and P-code were established by the satellite developer (Satellite Systems Division of Rockwell International Corporation) in 1991 [10, 19].

Navigation Signal The GPS satellite bit stream includes navigational information on the ephemeris of the transmitting GPS satellite and an almanac for all GPS satellites, with parameters providing approximate corrections for ionospheric signal propagation delays suitable for single-frequency receivers and for an offset time between satellite clock time and true GPS time. The navigational information is transmitted at a rate of 50 baud. Further discussion of the GPS and techniques for obtaining position information from satellite signals can be found in Chapter 4 of Ref. 24.

1.2.1.3 Selective Availability (SA) SA is a combination of methods available to the U.S. Department of Defense to deliberately derate the accuracy of GPS for "nonauthorized" (i.e., non-U.S. military) users during periods of perceived threat. Measures may include pseudorandom time dithering and truncation of the transmitted ephemerides. The initial satellite configuration used SA with pseudorandom dithering of the onboard time reference only [19], but this was discontinued on May 1, 2000.

Precise Positioning Service (PPS) Formal, proprietary service PPS is the full-accuracy, single-receiver GPS positioning service provided to the United States and its allied military organizations and other selected agencies. This service includes access to the encrypted P-code and the removal of any SA effects.

Standard Positioning Service (SPS) without SA SPS provides GPS single-receiver (stand-alone) positioning service to any user on a continuous, worldwide basis. SPS is intended to provide access only to the C/A-code and the L1 carrier.

1.2.1.4 Modernization of GPS GPS IIF, GPS IIR-M, and GPS III are being designed under various contracts (Raytheon, Lockheed Martin). These will have a new L2 civil signal and new L5 signal modulated by a new code structure. These frequencies will improve the ambiguity resolution, ionospheric calculation, and C/A-code positioning accuracy.

1.2.2 Global Orbiting Navigation Satellite System (GLONASS)

A second system for global positioning is the GLONASS, placed in orbit by the former Soviet Union, and now maintained by the Russian Republic [21, 22].

1.2.2.1 GLONASS Orbits GLONASS has 24 satellites, distributed approximately uniformly in three orbital planes (as opposed to six for GPS) of eight satellites each (four for GPS). Each orbital plane has a nominal inclination of 64.8° relative to the equator, and the three orbital planes are separated from each other by multiples of 120° right ascension. GLONASS orbits have smaller radii than GPS orbits, about 25,510 km, and a satellite period of revolution of approximately 8/17 of a sidereal day.

1.2.2.2 GLONASS Signals The GLONASS system uses frequency-division multiplexing of independent satellite signals. Its two carrier signals corresponding to L1 and L2 have frequencies $f_1 = (1.602 + 9k/16)$ GHz and $f_2 = (1.246 + 7k/16)$ GHz, where $k = -7, -6, \ldots 5, 6$ is the satellite number. These frequencies lie in two bands at 1.598–1.605 GHz (L1) and 1.242–1.248 GHz (L2). The L1 code is modulated by a C/A-code (chip rate = 0.511 MHz) and by a P-code (chip rate = 5.11 MHz). The L2 code is presently modulated only by the P-code. The GLONASS satellites also transmit navigational data at a rate of 50 baud. Because the satellite frequencies are distinguishable from each other, the P-code and the C/A-code are the same for each satellite. The methods for receiving and analyzing GLONASS signals are similar to the methods used for GPS signals. Further details can be found in the patent by Janky [19]. GLONASS does not use any form of SA.

1.2.2.3 Next Generation GLONASS The satellite for the next generation of GLONASS-K was launched on February 26, 2011 and continues to undergo flight tests. This satellite is transmitting a test CDMA signal at a frequency of 1202 MHz.

1.2.3 Galileo

The Galileo system is the third satellite-based navigation system currently under development. Its frequency structure and signal design is being developed by the European Commission's (EC's) Galileo Signal Task Force (STF), which was established by the EC in March 2001. The STF consists of experts nominated by the European Union (EU) member states, official representatives of the national frequency authorities, and experts from the European Space Agency (ESA).

1.2.3.1 Galileo Navigation Services The EU intends the Galileo system to provide the following four navigation services plus one search and rescue (SAR) service.

Open Service (OS) The OS provides signals for positioning and timing, free of direct user charge, and is accessible to any user equipped with a suitable receiver, with no authorization required. In this respect, it is similar to the current GPS L1 C/A-code signal. However, the OS is expected to be of higher quality, consisting of six different navigation signals on three carrier frequencies. OS performance is expected to be at least equal to that of the modernized Block IIR-M GPS satellites, which began launching in 2005, and the future GPS III system architecture currently being developed. OS applications will include the use of a combination of Galileo and GPS signals, thereby improving performance in severe environments such as urban canyons and heavy vegetation.

Safety of Life Service (SOL) The SOL service is intended to increase public safety by providing certified positioning performance, including the use of certified navigation receivers. Typical users of SOL will be airlines and transoceanic maritime companies. The European (also Geostationary) Navigation Overlay System (EGNOS) regional European enhancement of the GPS system will be optimally integrated with the Galileo SOL service to have independent and complementary integrity information (with no common mode of failure) on the GPS and GLONASS constellations. To benefit from the required level of protection, SOL operates in the L1 and E_5 frequency bands reserved for the Aeronautical Radionavigation Services.

Commercial Service (CS) The CS service is intended for applications requiring performance higher than that offered by the OS. Users of this service pay a fee for the added value. CS is implemented by adding two additional signals to the OS signal suite. The additional signals are protected by commercial encryption and access protection keys are used in the receiver to decrypt the signals. Typical value-added services include service guarantees, precise timing, ionospheric delay models, local differential correction signals for very high-accuracy positioning applications, and other specialized requirements. These services will be developed by service providers, which will buy the right to use the two commercial signals from the Galileo operator.

Public Regulated Service (PRS) The PRS is an access-controlled service for government-authorized applications. It is expected to be used by groups such as police, coast guards, and customs. The signals will be encrypted, and access by region or user group will follow the security policy rules applicable in Europe. The PRS will be operational at all times and in all circumstances, including periods of crisis. A major feature of PRS is the robustness of its signal, which protects it against jamming and spoofing.

SAR The SAR service is Europe's contribution to the international cooperative effort on humanitarian SAR. It will feature near real-time reception of distress messages from anywhere on Earth, precise location of alerts (within a few meters), multiple satellite detection to overcome terrain blockage, and augmentation by the four low earth orbit (LEO) satellites and the three geostationary satellites in the current Cosmitscheskaja Sistema Poiska Awarinitsch-Search and Rescue Satellite (COSPAS-SARSAT) system.

1.2.3.2 Galileo Signal Characteristics Galileo will provide 10 right-hand circularly polarized navigation signals in three frequency bands. The various signals fall into four categories: F/Nav, I/Nav, C/Nav, and G/Nav. The F/Nav and I/Nav signals are used by the OS, CS, and SOL services. The I/Nav signals contain integrity information, while the F/Nav signals do not. The C/Nav signals are used by the CS, and the G/Nav signals are used by the PRS.

E_{5a}–E_{5b} Band This band, which spans the frequency range from 1164 to 1214 MHz, contains two signals, denoted E_{5a} and E_{5b}, which are respectively centered at 1176.45 and 1207.140 MHz. Each signal has an in-phase component and a quadrature component. Both components use spreading codes with a chipping rate of 10.23 Mcps (million chips per second). However, the in-phase components are modulated by navigation data, while the quadrature components, called *pilot signals*, are data-free. The data-free pilot signals permit arbitrarily long coherent processing, thereby greatly improving detection and

tracking sensitivity. A major feature of the E_{5a} and E_{5b} signals is that they can be treated as either separate signals or a single wide-band signal. Low-cost receivers can use either signal, but the E_{5a} signal might be preferred, since it is centered at the same frequency as the modernized GPS L5 signal and would enable the simultaneous reception of E_{5a} and L5 signals by a relatively simple receiver without the need for reception on two separate frequencies. Receivers with sufficient bandwidth to receive the combined E_{5a} and E_{5b} signals would have the advantage of greater ranging accuracy and better multipath performance.

Even though the E_{5a} and E_{5b} signals can be received separately, they actually are two spectral components produced by a single modulation called alternate *binary offset carrier* (altBOC) modulation. This form of modulation retains the simplicity of standard BOC modulation (used in the modernized GPS M-code military signals) and has a constant envelope while permitting receivers to differentiate the two spectral lobes. The current modulation choice is altBOC(15,10), but this may be subject to change.

The in-phase component of the E_{5a} signal is modulated with 50 sps (symbols per second) navigation data without integrity information, and the in-phase component of the E_{5b} signal is modulated with 250 sps data with integrity information. Both the E_{5a} and E_{5b} signals are available to the OS, CS, and SOL services.

E_6 Band This band spans the frequency range from 1260 to 1300 MHz and contains a C/Nav signal and a G/Nav signal, each centered at 1278.75 MHz. The C/Nav signal is used by the CS service and has both an in-phase and a quadrature pilot component using a BPSK spreading code modulation of 5 × 1.023 Mcps. The in-phase component contains 1000-sps data modulation, and the pilot component is data-free. The G/Nav signal is used by the PRS service and has only an in-phase component modulated by a BOC(10,5) spreading code and data modulation with a symbol rate that is to be determined.

L1–E_1 Band The L1–E_1 band (sometimes denoted as L1 for convenience) spans the frequency range from 1559 to 1591 MHz and contains a G/Nav signal used by the PRS service and an I/Nav signal used by the OS, CS, and SOL services. The G/Nav signal has only an in-phase component with a BOC spreading code and data modulation; the characteristics of both are still being decided. The I/Nav signal has an in-phase and quadrature component. The in-phase component will contain 250-sps data modulation and will use a BOC(1,1) spreading code, but this has not been finalized. The quadrature component is data-free.

1.2.3.3 Updates The first Galileo In-Orbit Validation Element (GIOVE) satellites, designated as GIOVE-A, GIOVE-B, were launched in 1995 and 2008, respectively. The next two were launched in October 2011 and began broadcasting in December 2011. All Galileo signals were activated on December 17, 2011 simultaneously for the first time across the European GNSS system's three spectral bands known as E_1 (1559–1592 MHz), E_5 (1164–1215 MHz), and E_6 (1260–1300 MHz).

1.2.4 Compass (BeiDou-2)

The BeiDou Navigation Satellite System is being developed by the People's Republic of China (PRC), starting with regional services, and later expanding to global services. Phase I was established in 2000. Phase II is to provide for areas in China and its surrounding areas by 2012. Phase III will provide global service by 2020.

1.2.4.1 Compass Satellites BeiDou will consist of 27 medium earth orbit (MEO) satellites, including 5 geostationary earth orbit (GEO) satellites and 3 inclined GEO satellites. The GEO satellites will be positioned at 58.75°E, 80°E, 110.5°E, 140°E, and 160°E.

1.2.4.2 Frequency The nominal carrier frequency of 1561.098 MHz with B1 signal is currently a quadrature phase-shift key in (QPSK) modulation. Compass now has 13 (as of 8/6/2012) BeiDou-2 satellites operating in its constellation [1]. Details of this section are given in Chapter 4.

1.3 INERTIAL NAVIGATION OVERVIEW

The purpose of this section is to explain how and why inertial navigation came about, how it works in general, and what it is used for—mostly from a historical standpoint. Although the development of inertial systems technology has involved perhaps hundreds of thousands of people, much of their work was so highly classified that little of their history is known today. Historical discussion here is very limited, mostly about those technologies still used today, and with limited references to the multitude of people it took to make it happen. More detail can be found in the historical accounts by Draper [5], Gibson [8], Mackenzie [26], Mueller [28], Wagner [34], and Wrigley [35]. For an account of the contemporary computer hardware and software developments through that period, see McMurran [27].

Chapter 3 has the essential technical details about hardware and software implementations, and Chapter 11 is about analytical methods for statistical characterization of navigation performance.

1.3.1 Theoretical Foundations

It has been called "Newtonian navigation" [33] because its theoretical foundations have been known since the time of Newton:

Given the position $x(t_0)$ and velocity $v(t_0)$ of a vehicle at time t_0, and its acceleration $a(s)$ for times $s > t_0$, then its velocity, $v(t)$, and position, $x(t)$, for all time $t > t_0$ can be defined as

$$v(t) = v(t_0) + \int_{t_0}^{t} a(s)\,ds$$

$$x(t) = x(t_0) + \int_{t_0}^{t} v(s)\,ds.$$

It follows that, given the initial position $x(t_0)$ and velocity $v(t_0)$ of a vehicle, its subsequent position depends only on its subsequent accelerations. If these accelerations could be measured and integrated, this would provide a navigation solution.

However, the technology of Newton's time was inadequate for practical implementation. What was missing included the following:

1. Sensors for measuring acceleration with accuracy sufficient for the intended mission. Two types of sensors would be required:
   (a) acceleration sensors for measuring each of the three components of acceleration
   (b) rotation sensor for keeping track of the directions of the acceleration components being measured from inside a moving vehicle.
2. Compatible methods for integrating the sensor outputs to obtain position and for putting the results into compatible formats for the application. This would include
   (a) integrating the outputs of rotation sensors to determine the orientation of the acceleration sensors
   (b) integrating the measured accelerations to obtain velocities and integrating the velocities to obtain position.
3. Hardware and software for implementing these methods and for putting the results into useable forms.
4. Applications that could justify the investments in technology required for developing the solutions to the capabilities listed above. It could not be justified for transportation at the pace of a sailing ship or a horse, and it would not happen until we had long-range missiles.

These issues are addressed in the remainder of this section.

1.3.2 Inertial Sensor Technology

The following accounts provide a rather small sample of the technologies developed for inertial navigation, covering mostly those that have remained in use today.

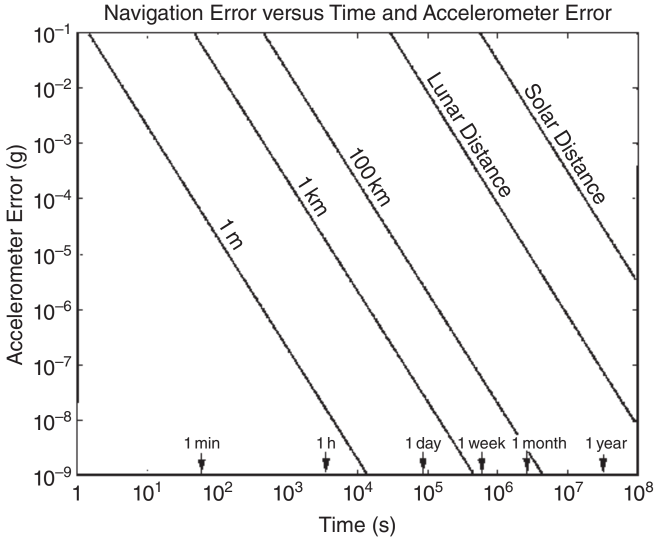

Fig. 1.1 Inertial navigation error as a function of sensor error and time.

1.3.2.1 Sensor Requirements Inertial navigation performance is primarily limited by inertial sensor performance because the sensor accuracies required for achieving even modest navigational performance can be difficult to attain. Figure 1.1 is a contour plot of the evolution over time of inertial navigation position error δ_{pos} resulting from accelerometer (acceleration sensor) error δ_{acc}, using Newton's model:

$$\delta_{\text{pos}} = \frac{1}{2}\delta_{\text{acc}}t^2,$$

where t is the time since navigation began. The plotted results would indicate, for example, that achieving 1 km of navigation accuracy after a week at sea would require acceleration sensor accuracies better than 10^{-9} g ($\approx 9.8 \times 10^{-9}$ m/s/s), and even something as modest as 1 km after an hour would require sensor accuracies in the order of 10^{-5} g. Achieving such accuracies aboard a moving vehicle was not going to be easy.

It turns out that, for terrestrial inertial navigation, a similar plot can be made in which accelerometer error (in units of 1 g) is replaced by attitude sensor error (in radians). An inertial navigation accuracy of 1 km after 1 h would then require attitude sensor accuracies in the order of 10^{-5} rad, or about 2 arc seconds. This is because an attitude error of 10^{-5} rad results in a miscalculation of gravitational acceleration[2] by about 10^{-5} g.

---

[2] It also turns out that, in the terrestrial environment, the shape of the gravitational field ameliorates the buildup of position errors with time. That subject is discussed in Chapter 11.

1.3.2.2 *Motivation* Inertial navigation is a product of the Cold War between the Soviet Bloc and NATO Allies. The United States and the Soviet Union had cooperated in defeating Nazi Germany in World War II. At war's end, both sides rushed to capture what they could find in the way of German military technology and those who developed it, and they did not wish to share. It was a harbinger of the coming Cold War.

The United States had developed and used nuclear weapons toward the end of World War II, and the Soviet Union was quick to develop its own nuclear capabilities. Both sides now felt they needed compatible long-range delivery systems, and both sides mounted well-funded programs to develop that capability. During that period, neither side could assume it had the option of doing only what it could afford to do. It was a time for doing whatever one cannot afford not to do. This was what motivated development of inertial navigation.

1.3.2.3 *Inertial Sensors Prior to Newton* Inertial sensors are actually older than Newton, although little was known about them at the time. Many species of flying insects have an extra pair of modified wings called *halters*, which function as rotation rate sensors during flight. Today, we would recognize the design as a *vibrating Coriolis gyroscope*.

Closer to home, you carry around in your head a pair of even more sophisticated inertial sensor systems. In the bony mass behind each ear is a *vestibular system.* It has been evolving since the time your ancestors were fish [32]. Each of your vestibular systems consists of a set of three rotation sensors (*semicircular canals*), with roughly orthogonal axes of rotational sensitivity. These are augmented by a set of acceleration sensors (*saccule* and *utricle*) indicating the direction and magnitude of accelerations due to forces applied to your head. Your vestibular systems play a major role in compensating your vision system during rotations of your head. They also help you maintain your balance and attitude awareness when you are without visual cues. Their performance is not what we would consider "inertial grade" (i.e., good enough for practical inertial navigation), but they have the essential elements of an inertial navigation system (INS):

1. a complete set of three quasi-orthogonal sensors for changes in the attitude of your head
2. sensors for three orthogonal components of acceleration of your head due to applied forces
3. a (neurological) processor for resolving these sensor outputs into applicable representations of the displacements and attitude changes you have experienced.

Man-made systems for inertial navigation have the same essential parts except that they are not biological (yet).

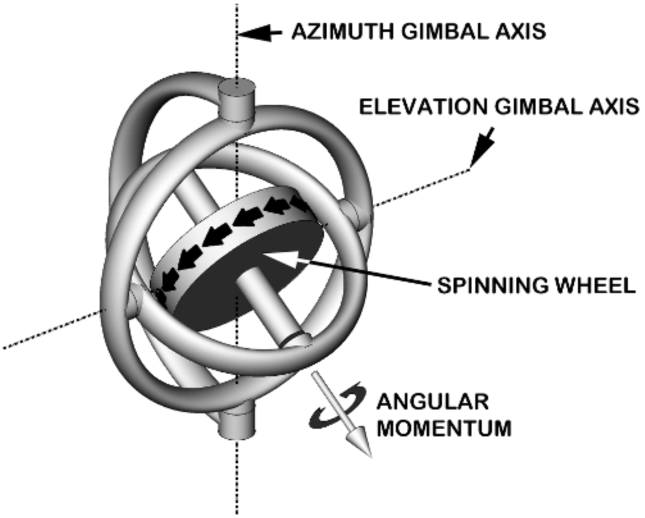

Fig. 1.2 Essential features of the Bohnenberger–Foucault gyroscope.

1.3.2.4 Early Momentum Wheel Gyroscopes (MWGs) The word *gyroscope* was coined by Jean Bernard Léon Foucault (1819–1868) in the mid-nineteenth century. The name was composed from Greek stem-words meaning essentially "rotation sensor," and that is what it still means. In 1852, Foucault used a design like that illustrated in Fig. 1.2 to measure the rotation of the earth [7]. It is called an *MWG* because it uses the conservation of angular momentum of the spinning wheel to maintain an inertially fixed (i.e., nonrotating, ideally) reference direction: the rotor spin axis. Foucault's gimbaled gyroscope was similar to an earlier design by Bohnenberger [3], but with a spinning wheel in place of Bohnenberger's spinning sphere.

Foucault's gyroscope had to be spun up manually and was only useful for a matter of minutes before bearing drag slowed it down. Methods for sustaining rotor spin (using compressed gas or electricity) would solve the run-down problem, and further improvements in bearing technology resulted in significant advances in the ability of MWGs to maintain a true inertial direction. Since Foucault's time, MWG bearings have included

1. thrust or sleeve bearings
2. pivot or jewel bearings
3. ball or roller bearings
4. gas bearings
5. magnetic bearings
6. electrostatic bearings.

Many of these improvements occurred early in the nineteenth century, by which time MWGs would replace the magnetic compass, which was not that reliable aboard iron ships. Soon after aircraft were introduced, MWGs were also being developed for flight instrumentation.

1.3.2.5 German Inertial Technology: 1930s–1945 The Treaty of Versailles restricted development of artillery in Germany. The response was to develop alternative weapons not covered by the Treaty [8]. The Nazi government put great effort into alternative means for delivering explosive projectiles over long distances. From this would come a cruise missile (V-1) and a ballistic missile (V-2). These had greater ranges than artillery, but both required means for autonomous guidance and control during flight. Their developers experimented with magnetic compasses, radio navigation, and inertial sensors.

World War II German technologies for inertial sensing and control were remarkably advanced for the time. Although the related technologies for onboard processing were severely limited, they were adequate for short-range missions of several minutes' duration. Most guidance computations were done on the ground and simplified to a form that could be implemented onboard a rocket or cruise missile using the electromechanical technology of the day. Onboard inertial guidance implementation had to be done "open loop," in the sense that control was used only to follow a preprogrammed trajectory without feedback related to trajectory errors. Control was achieved by using rotation sensors to detect deviations from the intended heading (yaw angle) and pitch angle, and feedback was applied to aerodynamic actuators (ailerons and elevons for cruise missiles, or equivalent vanes in rocket exhaust). Radio navigation was also used as an aid for maintaining the planned heading, and the cruise missile used a barometric altimeter to control altitude. The amount of fuel loaded into the V-1 cruise missile was used as a means for controlling its range. Range control for the V-2 rocket used a rather sophisticated integrating acceleration sensor to keep track of accumulated velocity for initiating thrust termination.

Gyroscopes had been used before this time for the control of torpedoes[3] and rockets.[4] German innovations in inertial technology would include

1. "Inertial platforms," capable of maintaining a nonrotating orientation with respect to the celestial sphere. This would remain an essential component of all INSs for decades. All spacecraft for the Apollo moon missions, for example, contained inertial platforms. Even today, the most accurate INSs use inertial platforms.
2. The first inertial grade integrating accelerometer, the output of which is proportional to accumulated acceleration (velocity change). This basic

---

[3] In the 1890s, by British torpedo pioneer Robert Whitehead.

[4] In 1932, by American rocket pioneer Robert H. Goddard.

design concept, with implementation refinements, has had different names at different times, but it is still being used today.

3. Electromechanical servomechanisms for feedback control, used in the implementations of the inertial platform, missile guidance, and fire-control systems.
4. Models and methods for predicting inertial system performance based on sensor error characteristics. These models were essential for developing unmanned missiles, with no operator available to observe performance and make corrections. These models enabled systems designers to predict relative performance of alternative designs before systems were flight-tested.

German scientists and engineers were successful in developing inertial guidance and control for missiles, which was all that was needed at the time. Inertial navigation would come later.

The Inertial Platform An inertial platform is a hardware solution to the problem of keeping track of the directions of the acceleration components being measured from inside a moving vehicle. It solves the problem by keeping the acceleration sensors in a known, inertially fixed (i.e., nonrotating) orientation, a concept developed and patented by Johann M. Boykow (1878–1935), who had also worked on aircraft autopilots.

Figure 1.3 is an illustration of the basic design features of an inertial platform, with a nested set of three gimbals used to allow the innermost member to have complete rotational freedom about three axes. Three gyros (represented by blocks with "G" on all faces) on the innermost member are used to sense rotations about their respective input axes (i.e., axes of rotational sensitivity, represented by arrows), and any sensed rotation is nulled using feedback loops with torque motors in the gimbal bearings. Three accelerometers (represented by blocks with "A" on all faces) are used to measure accelerations along their input axes (also represented by arrows).

The inertial sensor design for the V-2 rocket was similar to the configuration shown in the figure, with the top open for easier access during assembly and testing. The sensors looked quite different, but the operational principles were the same.[5]

Integrating Accelerometers The first acceleration sensor that might be considered "inertial grade" was invented in Germany in the late 1930s by Fritz Mueller (1907–2001) [28] and was named (in English) the *Mueller mechanical integrating accelerometer* (MMIA). Improved models have been called *pendulous integrating gyroscopic accelerometers* (PIGAs), or *specific*

---

[5] MWGs are capable of sensing rotation about two axes, which complicates the implementation just a bit.

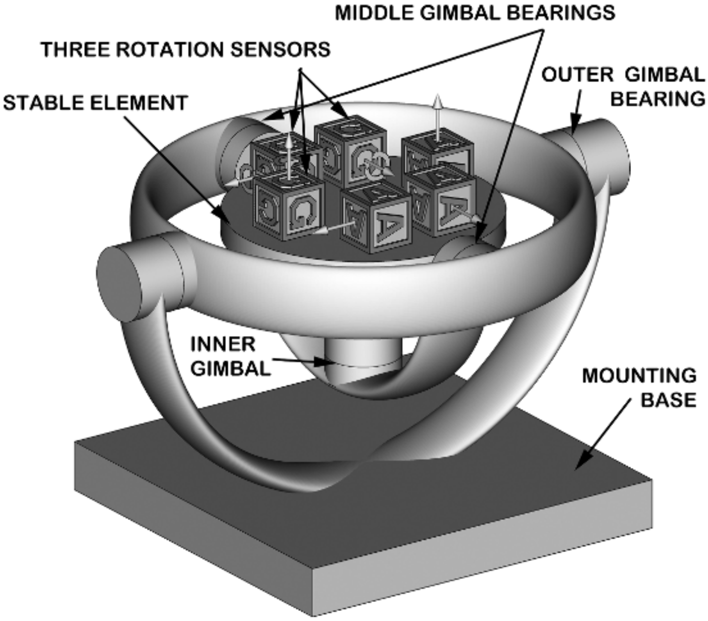

Fig. 1.3 Basic design features of a gimbaled inertial platform.

force integrating receivers (SFIRs, pronounced “siffers”), but the underlying principles have remained the same [13].

The basic features of the MMIA are illustrated in Fig. 1.4. As shown, it is similar in design to the Bohnenberger–Foucault gyroscope illustrated in Fig. 1.2 except that the inner gimbal and rotor assembly has been deliberately unbalanced so that its center of mass is offset from its center of support (along the inner gimbal rotation axis) in the direction along the rotor spin axis. The implementation of the MMIA in hardware requires a beefier gimbal structure than that shown in the figure, and the better implementations use a fluid-filled cylinder as the outer gimbal bearing.

Figure 1.5 illustrates how the MMIA functions. Any force F, applied to the rotor and inner gimbal assembly (including the added mass), is applied through the inner gimbal bearings. Because the inner gimbal rotation axis is offset from the center of mass of the rotor assembly by some distance D, the reactive force ma creates a coupling torque $T = maD$, causing the angular momentum of the rotor assembly to precess about the outer gimbal axis. The angular rate of precession will be proportional to the applied acceleration a, and the total precession angle will be proportional to the integral of acceleration. The MMIA performs this integration rather accurately—more accurately than any of the other methods for implementing integration available in the 1940s. The MMIA used in German V-2 rockets had scale factor errors in the order of

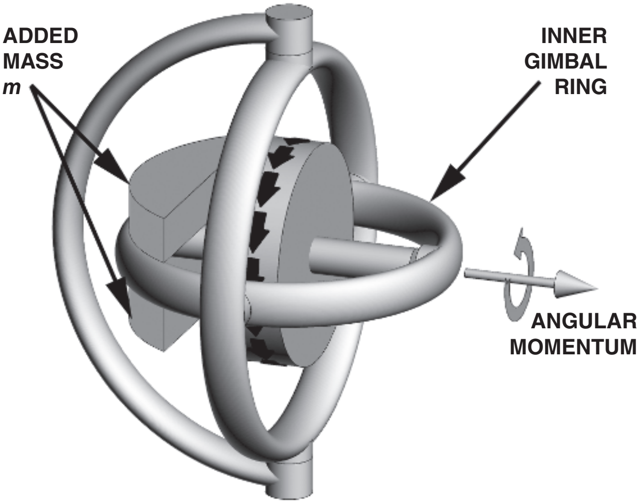

Fig. 1.4 Unbalanced inner gimbal ring of MMIA.

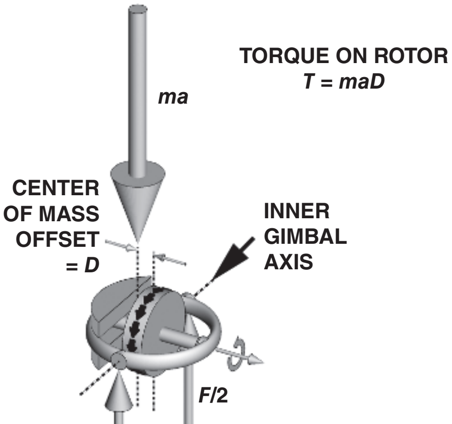

Fig. 1.5 Force–balance mechanics of the MMIA.

0.1%, or a relative error of about 0.001. Today's models have scale factor errors several orders of magnitude smaller.

The MMIA was used to control the range to impact of the V-2 by metering the accumulated acceleration due to thrust and signaling engine cutoff when a preset threshold was reached.

After the war, Mueller came to the United States and continued development of inertial guidance and control systems, including the one that helped put the first American satellite in orbit in 1958.

1.3.2.6 Charles Stark Draper (1901–1987), "The Father of Inertial Navigation" Although, as one of its early pioneers put it, the INS "was apparently evolved rather than invented,"[6] Draper had a lot to do with reducing it to practice. Born before the Wright brothers made their first flight in 1903, he developed an early interest in aviation, obtained a civil aviation license, and flew his own airplane. This experience made it obvious to him that better flight instrumentation was required. His early success in research and development of flight instrumentation led to his founding the Instrumentation Laboratory at MIT in the early 1930s.

During World War II, Draper and his associates at MIT developed antiaircraft fire-control systems used aboard Navy ships and military aircraft. Much of this technology depended on gyroscopic instruments to measure angular rates, analog computation of firing solutions, and servomechanisms to control firing directions.

Beginning in 1946, Draper led a defense-funded research and development project at the Instrumentation Laboratory to develop an inertial navigator for manned missions.[7] Its intended applications included long-range bombers, surface ships, and submarines.

In February of 1953, Draper's team demonstrated an INS aboard a World War II-vintage Boeing B-29 bomber on a flight from Bedford, Massachusetts to Los Angeles, California—a distance of about 2250 nmi (~4167 km). It was the first successful demonstration of acceptable inertial navigation performance over a representative mission distance. Draper's INS was called Space Inertial Reference Equipment (SPIRE). It was the size of a small automobile and weighed around 2700 lb, but it worked.

The Instrumentation Laboratory at MIT would go on to develop successive generations of INSs, including those for the NASA Apollo Command Module and Lunar Excursion Module (LEM), the Air Force Atlas, Thor, Titan, and MX intercontinental ballistic missiles (ICBMs), and the Navy Polaris, Poseidon, and Trident submarine-launched ballistic missiles (SLBM). The Advanced Inertial Reference Sphere (AIRS) designed at MIT for the MX missile is perhaps the most accurate system ever developed for ICBMs. It uses a floated sphere in place of gimbals, called a floated inertial measurement ball (FLIMBAL).

---

[6] G. R. Pitman, editor, *Inertial Guidance*, Wiley, 1962.

[7] At the time, Russian-born theoretical physicist George Gamow was a member of one of the government science advisory boards and an early critic of Draper's inertial navigation project—on the grounds that inertial navigation was theoretically unstable in vertical navigation. Gamow was correct, but Draper was able to prevail anyway. Navigators aboard surface ships generally knew their altitudes; barometric altimeters would solve the vertical navigation problem for aircraft, and depth sensors would do the same for submarines.

TABLE 1.1. U.S. Ground-Launched Missile Projects Begun in 1946

Project Number	Contractor	Missile Type	Missile Name(s)
MX-770	North American Aviation	Cruise	Navaho
MX-771	Glen L. Martin Co.	Cruise	Matador
MX-772	Curtiss-Wright Corp.	Cruise	[a]
MX-773	Republic Aviation	Cruise	[a]
MX-774	Consolidated Vultee	Ballistic	Hiroc, Atlas
MX-775	Northrup Corp.	Cruise	Snark, Boojum

[a] Cancelled 1947.

The MIT Instrumentation Laboratory was renamed the Charles Stark Draper Laboratory (CSDL) in 1970 and spun off as an independent, MIT-owned, not-for-profit corporation in 1973. It would later play a major role in the development of micron-scale inertial sensors. The designs of all systems developed at MIT and CSDL were turned over to the U. S. Department of Defense, which then put production contracts out for bids by commercial manufacturers.

1.3.2.7 Aerospace Inertial Technology All successful inertial instruments and systems designed at MIT were eventually manufactured by commercial aerospace companies, which helped develop the industry. In addition, however, many aerospace companies were discovering inertia navigation on their own.

Table 1.1 lists some of the major military-funded projects started in 1946 for the purpose of developing practical ground-launched delivery systems for nuclear weapons. All were to be unmanned, requiring automated guidance and control. Most of these were cruise missiles, which were thought to have greater range potential because they could use air-breathing propulsion to eliminate the oxidizer weight needed for rockets. However, as the missile technologies matured, so did nuclear weaponry. By the late 1950s, nuclear payloads had shrunk and rocketry had improved to point that, of these, only the Atlas ballistic missile survived. However, inertial navigation technologies developed for many of the other projects had been so successful that they found applications on other projects.

Project MX-770, for example, was cancelled in 1957, after more than a decade of development. By then, its inertial guidance technology was so advanced it could be successfully applied to inertial navigation of Navy ships, including submarines. A year after Project MX-770 had been terminated, a modified version of the INS developed for Navaho was used for navigating the nuclear submarine *USS Nautilus* under the ice at the North Pole.

By the 1960s, many of the major aerospace and commercial companies were involved in developing inertial sensors and systems. These would come to include such names as Litton, Sperry, Teledyne, Honeywell, and Delco.[8]

---

[8] Originally named AC Spark Plug, AC being the initials of its founder, Albert Champion (1828–1927).

TABLE 1.2. A Sampling of Inertial Sensor Types

What It Measures	Sensor Type	Physical Phenomenon	Implementation Method
Rotation (gyroscope)	Momentum wheel gyro	Angular momentum	Displacement Torque rebalance
	Coriolis gyro	Coriolis effect	Rotation Vibration
	Optical gyro	Sagnac effect	Fiber-optic gyroscope
		Laser	Ring laser gyro
Acceleration (accelerometer)	Gyroscopic	Precession due to mass unbalance	Displacement Torque rebalance
	Electromagnetic	Induction	Drag cup
		Electromagnetic force	Force rebalance
	Mass–spring	Strain	Piezoelectric Piezoresistance
	Electrostatic	Electrostatic force	Force rebalance

Innovations in inertial technology during this period included a number of new sensor designs based on new physical principles, some of which are listed in Table 1.2.

Improvements in MWGs The major sources of sensor error in MWGs include those due to rotor mass unbalance (the same precession mechanism exploited by the MMIA) and those due to bearing torques.

The mass-unbalance problem was solved by a combination of improved manufacturing tolerances and the ability to calibrate it and compensate for it during operation. It is an acceleration-sensitive effect, and it can be compensated using the accelerations measured by the accelerometers. The most accurate inertial sensors developed during this period would all rely on a combination of high-precision manufacturing, sensor calibration, and run-time compensation.

The effects of bearing torques were reduced significantly by two new bearing technologies. One of these uses gas as a lubricant; another uses electrostatic forces to support the rotor in a vacuum.

Early gas-bearing gyroscopes used compressed gas (much like an air hockey table). Joseph C. Boltinghouse (1909–2009) and John M. Slater (1908–1987) developed an alternative bearing design using precise spherical bearing surfaces in very close proximity, such that no gas pumping was required. The spinning of the rotor was sufficient to maintain gas lubrication of the bearing surfaces. This design was used in the INS for Minuteman missiles.

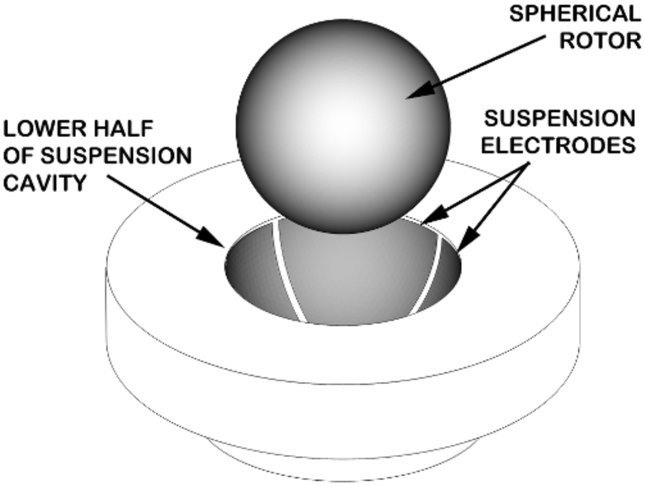

Fig. 1.6 The Boltinghouse micro-ESG.

A further reduction in bearing torques was achieved by using electrostatic pressure[9] to suspend a spherical rotor in a spherical electrode cavity. The electrostatic gyroscope (ESG) was first developed by Arnold T. Nordsieck (1911–1971), then a professor of physics at the University of Illinois, and was first applied to inertial navigation in the 1960s by the Honeywell Corporation. The Nordsieck design uses a hollow beryllium ball about the size of a golf ball, which is somewhat difficult to manufacture to the close tolerances required on its inside and outside surfaces. Autonetics instrument designer Joseph C. Boltinghouse developed an alternative design, illustrated in Fig. 1.6. It has a smaller (1-cm diameter) solid beryllium rotor, which is much easier to manufacture. The U.S. Navy Electrostatically Suspended Gyro Navigator (ESGN), using the Boltinghouse ESG, would be the primary INS aboard Trident-class submarines for decades. Its performance is classified, but it is probably the most accurate INS ever built for submarine navigation.

The most accurate gyroscopes made to date have been ESGs. However, these were not used for inertial navigation, but for a theoretical physics experiment named "Gravity Probe B." This was a NASA-funded program started in 1976 and continued until 2011. Its mission was to resolve two fine points of Einstein's theory of gravitation, the most demanding of which was to measure an effect called "frame dragging," predicted to cause an effective inertial coordinate rotation rate of around 37 milliarcseconds/year ($\sim 10^{-9}$ deg/h). A team at Stanford University, led by Dr. Francis Everitt, designed the scientific payload,

---

[9] It is easier to create negative (attractive) pressure with electrostatics, so the rotor is suspended by overhead electrostatic attraction.

a satellite containing four superconducting ESGs. These were able to measure the frame-dragging rate with a 95% confidence level of about $\pm 2.3 \times 10^{-10}$ deg/h [6].

Optical Gyroscopes The first optical gyroscopes were designed in the 1960s, not long after the first functioning helium neon laser was demonstrated.[10] These were ring laser gyroscopes (RLGs), which use a closed-loop lasing path with mirrors at the corners and laser beams propagating in both directions. The phase coherence of the two counter-rotating beams can be measured by optical interferometry of the two beams, deliberately allowed to leak through one of the corner mirrors. Rotation of the device about an axis orthogonal to the plane of the laser path will (due to the finite velocity of light) cause one beam to advance in phase relative to the other. Early designs were plagued by a phenomenon called "lock-in," in which the two counter-rotating beams would remain phase locked at low rotation rates. RLG designers at Honeywell discovered that lock-in was not instantaneous and were successful in avoiding lock-in by adding zero-mean pseudorandom dither in rotation of the optical element relative to its mounting base. In the 1980s, instrument designers at Litton Guidance & Control Systems (now the Navigation Systems Division of Northrop Grumman) solved the fundamental problem with a design using a combination of dual-frequency lasing and a nonplanar lasing path, as illustrated in Fig. 1.7. Appropriately enough, the design is called the Zero Lock Gyro™, or ZLG™ (both registered trademarks).

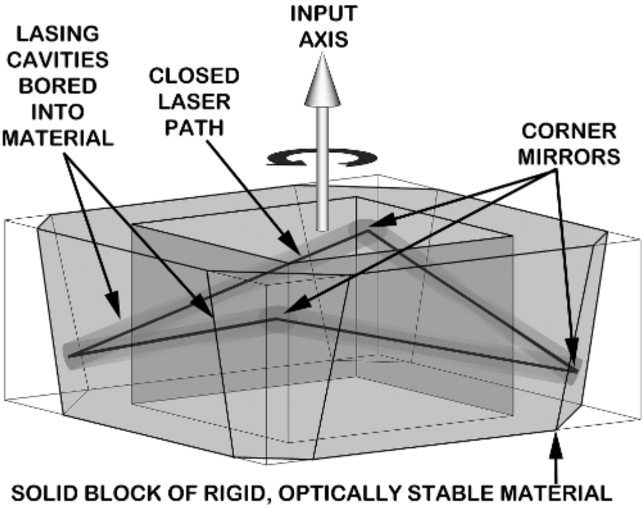

Fig. 1.7 Basic design features of a ring laser gyroscope (RLG).

---

[10] By A. Javan, W. Bennett, and D. Herriott at Bell Labs in 1960.

RLGs function as *rate-integrating gyroscopes* in that the output phase-shift rate is proportional to the incremental rotation angle. As a consequence, each bit of output represents a fixed rotation angle.

The second fundamental type of optical gyroscope is the fiber-optic gyroscope (FOG). Its development was able to piggyback on commercial optical fiber developments in the 1970s. As illustrated in Fig. 1.8, the FOG uses a common source to transmit laser light both ways through a coil of very long optical fiber. Its function depends on the differential delay due to rotation of the coil and the finite speed of light, a physical phenomenon called the *Sagnac effect*. Optical interferometry of the laser light exiting the fiber at opposite ends will show a phase shift proportional to the rotation rate. As a consequence, the FOG is a rate gyroscope. Its output is proportional to rotation rate.

All optical gyroscope designs require geometric stability to subwavelength levels, and the optical fibers in FOGs are particularly sensitive to stress.

Vibrating Coriolis Gyroscopes The Coriolis effect (derived in Appendix B) is one of the corrections required for modeling Newtonian mechanics in a rotating coordinate frame. It is modeled in the form of an apparent acceleration (*Coriolis acceleration*) experienced by a moving mass whose coordinates are represented in a rotating coordinate system, and has the form

$$a_{\text{Coriolis}} = -2\omega \otimes v_{\text{rotating}},$$

where a_{Coriolis} is the apparent acceleration, ω is the rotation rate (a vector with three components representing the coordinate rotation rates about the three

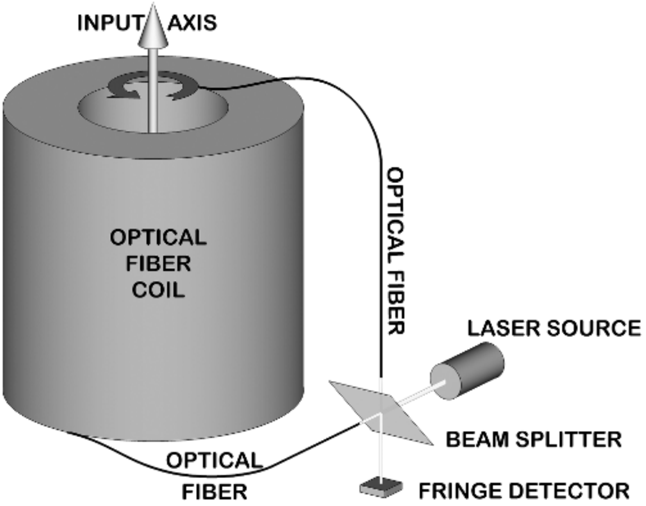

Fig. 1.8 Essential elements of the fiber-optic gyroscope.

rotating coordinate axes), ⊗ represents the vector cross product, and v_{rotating} is the velocity of the mass represented in rotating coordinates. The Coriolis effect couples velocity—even vibrational velocity—into acceleration, and the resulting accelerations are orthogonal to the velocities.

The tuning fork gyroscope in Fig. 1.9 illustrates how this effect works to make a rotation rate sensor. The tuning fork is well designed such that the in-and-out motion of its tines is balanced, and no stress is transmitted to the handle. However, when the tuning fork is rotated about its handle, the resulting Coriolis effect couples the balanced in-plane vibration mode into an unbalanced, twisting out-of-plane vibration mode, which produces vibrational torque on the handle. This output vibration can be sensed by using strain sensors between the handle and its holder, but a better solution is to attach another tuning fork handle to handle and end to end such that the twisting vibration mode of the attached tuning fork has the same resonant frequency as that of the in-plane vibration mode of the sensing tuning fork. Perhaps the best models use quartz as the tuning fork material. Quartz is piezoelectric, which means that the vibrational mode of the sensing fork can be controlled electronically. It is also stiff and light, which ups the resonant frequencies (a good thing), and practically lossless, which gives the resonator a high Q-factor (another good thing).

Balancing the vibration modes of a Coriolis gyroscope is very important, because any vibration transmitted through its support translates into energy loss (a bad thing) and potential signal coupling with other vibration sensors. Better balance overall can be achieved by using the three-dimensional rotational equivalent of a tuning fork: the **wine glass**. The vibrational modes of wine glasses have been known for some time. In 1890, George H. Bryan (1864–1928) discovered that, when a vibrating wine glass is rotated about its

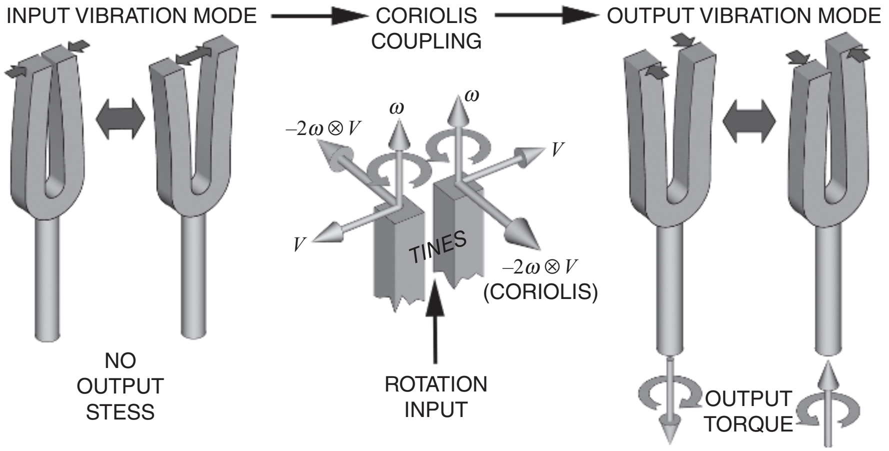

Fig. 1.9 Vibration modes of the tuning fork gyroscope.

stem, the vibrational nodes on its rim rotate at a rate different from that of the input rotation. Bryan called it the "wave inertia effect." It is now called the **Bryan effect**. Early gyroscopes based on the Bryan effect were called "wine glass gyros" (not a particularly prestigious name). They are now called "**hemispherical resonator gyroscopes**" (HRGs). The Cassini spacecraft sent to Saturn in 1997 uses hemispherical resonator gyroscopes for controlling its orientation.

Microelectromechanical systems (MEMS) are tiny electromechanical devices fabricated using wafer-scale processes based on semiconductor manufacturing technology. MEMS technologies were first developed in the 1970s, and an early application was for the acceleration sensors used in detecting automobile collisions for initiating air-bag deployment. Vibration frequencies of structures tend to increase significantly as the structural dimensions decrease, which meant that MEMS devices could be made to vibrate at frequencies up to ~10^5 Hz. Vibrational velocity scales up with frequency, which meant that the Coriolis effect was strong in MEMS devices.

In the early 1990s the CSDL developed a MEMS gyroscope resembling a slice through a tuning fork, as illustrated in Fig. 1.10. It has a pair of thin rectangular masses like cross sections of the tines, and they are driven to vibrate 180° out of phase, just like the tines of the tuning fork. The difference is that the input rotation axis is in the plane of the structure, and the output vibration mode is normal to the plane of the substrate, so that one mass moves upward while the other moves downward. This mode is detected by sensing capacitive changes between the vibrating masses and the underlying surface. In-plane vibration is controlled using opposing pairs of "comb drives," electrostatic force transducers developed at the University of California at Berkeley. The supporting electronics fit into a single application-specific integrated circuit (ASIC) chip. The Draper tuning fork gyro was licensed and further developed

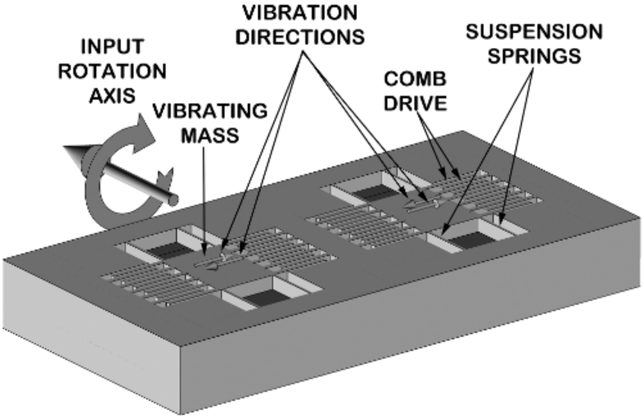

Fig. 1.10 Essential features of the CSDL MEMS gyro.

by Honeywell, which also produces vibrating Coriolis gyroscopes using the plate thickness vibrating mode at ~10^5 Hz.

Another gyroscope design resembles a slice through a HRG, just as the Draper gyro resembles a slice through a tuning fork. The Coriolis effect causes the nodes of the vibrating modes of the resulting ring structure to precess when the device is rotated about the axis normal to the plane of the ring.

Proof Mass Accelerometers A proof mass accelerometer measures the force F required to keep an otherwise isolated "proof mass" m from moving relative to its enclosure. If its enclosure is being accelerated at a rate a due to forces applied to it, then, knowing the force F it is applying to its proof mass m, a can be calculated as

$$a = \frac{F}{m},$$

the force per unit mass, also called *specific force*.

Proof mass accelerometers were developed as alternatives to the MMIA integrating gyroscopic accelerometer mentioned above. The gyroscope in the MMIA makes it sensitive to rotation, which is why it had to be mounted on an inertially stabilized base. It is also rather expensive.

The electromagnetic accelerometer (EMA) is a popular proof mass force–rebalance accelerometer using what is essentially a permanent-magnet speaker drive, as illustrated in Fig. 1.11. The cylindrical speaker coil is mounted on what

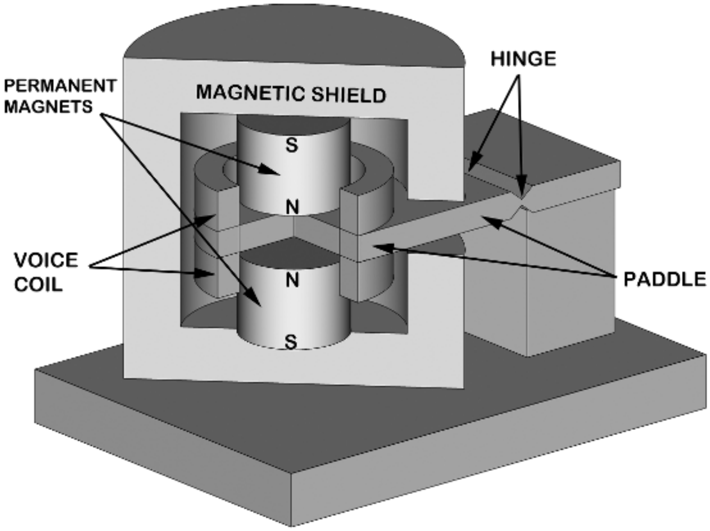

Fig. 1.11 Essential design elements of electromagnetic accelerometer.

is called a "paddle" attached to a compliant hinge, and the current through the coil is servoed to prevent the paddle from moving relative to its enclosure. The measured current is then proportional to the specific force being applied to the proof mass.

MEMS proof mass accelerometers are mostly mass–spring accelerometers, measuring the stress in the structure supporting a proof mass as a measure of the applied acceleration. Many use piezoresistive or piezoelectric films for measuring surface strain. Capacitance variation can also be used to measure displacement, although capacitors also create electrostatic forces corrupting the force measurement.

INS Signal Processing Hardware There was really no computer industry until the early 1950s and no flightworthy computers for inertial navigation until the 1960s. Inertial system implementations in the 1950s used a variety of interim technologies, including digital differential analyzers (DDAs) and magnetic drum memories. DDAs are digitized circuits specifically designed for integration. Enormous effort had to be put into making do with the processor technology of the time. One hybrid missile computer of that era used 128 DDAs together with a small general-purpose computer with only 3 kB of active memory [29].

Silicon transistors and integrated circuits began to revolutionize computer technology in the 1960s. The Apollo moon missions (1969–1972) used onboard computers with magnetic core memories. Magnetic core memory would dominate the market until semiconductor memories appeared in the 1970s. By that time, the cost of magnetic core memory had gotten down to pennies per bit. Memory prices would fall by several orders of magnitude over the next few decades. The introduction of the microprocessor in 1971 marked the beginning of a major downslide in the cost of computing.

Strapdown Systems Strapdown[11] systems use software to replace gimbals by processing the gyro outputs to maintain the coordinate transformation between accelerometer-fixed coordinates and inertial coordinates. The accelerometer outputs can then be transformed to inertial coordinates and processed just as they had been with a gimbaled system—without requiring an inertial platform.

Faster, cheaper computers enabled the development of strapdown inertial technology. Some vehicles (e.g., torpedoes) had been using strapdown gyroscopes for steering control since the late nineteenth century, but now they could be integrated with accelerometers to make a strapdown INS. This eliminated the expense of gimbals, but it also required considerable progress in attitude estimation algorithms [4]. Computers also enabled "modern"

---

[11] The terminology refers to the idea that the inertial sensors can be "strapped down" to the vehicle frame, although some form of vibration isolation is generally required.

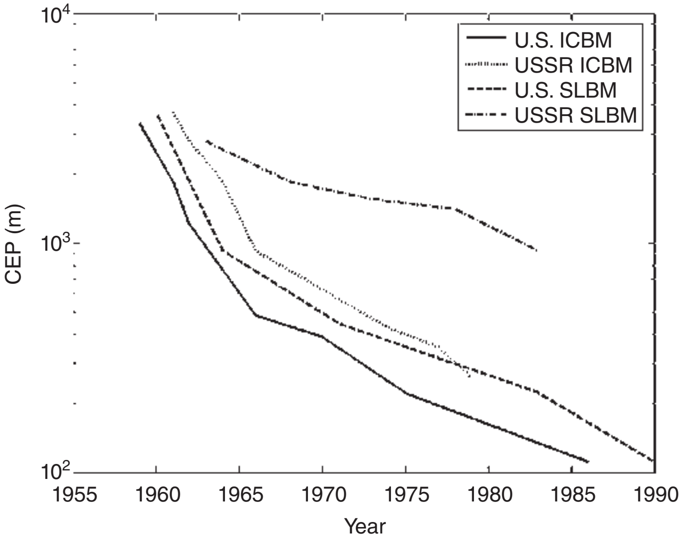

Fig. 1.12 Cold War missile accuracy improvements.

estimation and control, based on state space models. This would have a profound effect on sensor integration capabilities for INS.

A gimbaled INS was carried on each of nine Apollo command modules from the earth to the moon and back between December 1968 and December 1972, but a strapdown INS was carried on each of the six[12] Lunar Excursion Modules (LEMs) that shuttled two astronauts from lunar orbit to the lunar surface and back.

By the mid-1970s, strapdown systems were able to demonstrate navigational accuracies in the order of 1 nmi/h (CEP[13] rate). This was considered adequate for commercial aircraft at that time.

The Race for Accuracy Figure 1.12 is a plot of strategic missile capabilities developed during the Cold War, in terms of inertial guidance accuracies achieved over that period. Expected miss distances are given in meters, CEP. The different plots are labeled according to whether they are for U.S. or USSR missiles, and whether they are for surfaced-launched ICBMs or SLBMs. These data are from Tables A.1 and A.2 of Ref. 26, which should be consulted for

---

[12] Two additional LEMs were carried to the moon but did not land there. The Apollo 13 LEM did not make its intended lunar landing but played a far more vital role in crew survival.

[13] CEP is an acronym for "circle of equal probability." It is the radius of the 50% confidence circle.

additional clarifying information. According to Mackenzie [26], contributions of the inertial systems to these miss distances are actually rather minor.

1.3.2.8 Developments Since the Cold War The Cold War ended around the time GPS was becoming operational. By that time, INS technology had matured considerably. Not only had achievable accuracies improved by orders of magnitude but so had cost, weight, and power requirements. As a consequence, markets had expanded beyond strategic military applications to include tactical and commercial applications.

The availability of GNSS allowed for integrated GNSS/INS navigation systems accurate enough for automated grading, plowing, and mining. It has also lowered costs to the point that GNSS/INS systems can be embedded in consumer products.

1.4 GNSS/INS INTEGRATION OVERVIEW

1.4.1 The Role of Kalman Filtering

It has been called "navigation's integration workhorse" [23] for the essential role it has played in navigation and especially for integrating different navigation modes. Ever since its introduction in 1960 [20], the Kalman filter has played a major role in the design and implementation of most new navigation systems as a statistically optimal method for estimating position using noisy measurements. Because the filter also produces an estimate of its own accuracy, it has also become an essential part of a methodology for the optimal design of navigation systems. The Kalman filter has been essential for the design and implementation of every GNSS. It is unlikely that the first GNSS (GPS) could have been built without it.

Using the Kalman filter, navigation systems designers have been able to exploit a powerful synergism between GNSSs and INSs, which is possible because they have very complementary error characteristics:

- Short-term position errors from the INS are relatively small, but they degrade significantly over time.
- GNSS position accuracies, on the other hand, are not as good over the short term, but they do not degrade with time.

The Kalman filter takes advantage of these characteristics to provide a common, integrated navigation implementation with performance superior to that of either subsystem (GNSS or INS). By using statistical information about the errors in both systems, it is able to combine a system with tens of meters position uncertainty (GNSS) with another system whose position uncertainty degrades at kilometers per hour (INS) and achieve bounded position uncertainties in the order of centimeters (with differential GNSS) to meters.

1.4.2 Implementation

The Kalman filter solves for the solution with the least mean-squared error by using data weighting proportional to statistical information content (the inverse of uncertainty) in the measured data. It combines GNSS and INS information to

1. track drifting parameters of the sensors in the INS, so that INS performance does not degrade with time when GNSS is available
2. improve overall performance even when there are insufficient satellite signals for obtaining a complete GNSS solution
3. allow the INS to navigate with improved initial error whenever GNSS signals become unavailable
4. improve GNSS signal reacquisition when GNSS signals become available again by providing better navigation solutions (based on INS data)
5. use acceleration and attitude rate information from the INS for reducing the signal phase-tracking filter lags in the GNSS receiver, which can significantly improve GNSS reliability during periods of high maneuvering, jamming, or reduced signal availability.

The more intimate levels of GNSS/INS integration necessarily penetrate deeply into each of the subsystems in that it makes use of partial results that are not ordinarily accessible to users. To take full advantage of the offered integration potential, we must delve into technical details of the designs of both types of systems.

1.4.3 Applications

1.4.3.1 Military Applications The rationale for developing the Navistar GPS system was based, in part, on economic considerations—in terms of how many inertial systems it could replace. However, the ability to integrate GPS with INS also enabled military applications that were not possible before. It would lead to a new generation of high-precision military weaponry, improving military effectiveness while reducing collateral damage. Most missiles were already using inertial sensors for guidance and control, so the transition to integrated GNSS/INS navigation was natural.

Most military applications of inertial navigation were already using other navigation aids for limiting the growth of inertial navigation errors with time. The U.S. Navy had begun using satellites for aiding shipboard inertial navigation decades before GNSS became available, and most military INSs were being adapted to use GPS before it was operational. It has resulted in superior navigation performance at low marginal cost.

1.4.3.2 Civilian and Commercial Applications The availability of GNSS also allowed for integrated GNSS/INS navigation systems accurate enough for automated grading, plowing, and surface mining. The resulting relaxation of inertial sensor stability requirements and advances in fabrication technologies have also combined to lower costs to the point where low-performance GNSS/INS systems can be embedded in high-end consumer products. This market is likely to grow even more as costs fall due to increasing production volumes.

Details of Section 1.4 are given in Chapters 10 through 12.

PROBLEMS

1.1 How many satellites and orbit planes exist for GPS, GLONASS, and Galileo? What are the respective orbit plane inclinations?

1.2 List the differences in signal characteristics between GPS, GLONASS, and Galileo.

1.3 What are the reference points for GNSS and INS navigators? That is, when one of these produces a position estimate, what part of the respective system is that the position of?

REFERENCES

[1] "BeiDou Navigation Satellite System Signal in Space, Interface control Document," China Satellite Navigation Office, December 2011.

[2] D. J. Biezad, *Integrated Navigation and Guidance Systems*. American Institute of Aeronautics and Astronautics, New York, 1999.

[3] J. G. F. Bohnenberger, "Beschreibung einer Maschine zur Erläuterung der Geseze der Undrehung der Erde um ihre Axe, und der Verländerung der Lage der Letzteren," *Tübinger Blätter für Naturwissenschaften und Arzneikunde* Tübingen, Germany, **3**, 72–83 (1817)

[4] J. E. Bortz, "A New Mathematical Formulation for Strapdown Inertial Navigation," *IEEE Transactions on Aerospace and Electronic Systems* **AES-6**, 61–66 (1971).

[5] C. S. Draper, "Origins of Inertial Navigation," *AIAA Journal of Guidance and Control* **4**(5), 449–456 (1981).

[6] C. W. F. Everitt, D. B. DeBra, B. W. Parkinson, J. Turneaure, J. W. Conklin, M. I. Heifetz, G. M. Keiser, A. S. Silbergleit, T. Holmes, J. Kolodziejczak, M. Al-Meshari, J. C. Mester, B. Muhlfelder, V. G. Solomonik, K. Stahl, P. W. Worden, Jr., W. Bencze, S. Buchman, B. Clarke, A. Al-Jadaan, H. Al-Jibreen, J. Li, J. A. Lipa, J. M. Lockhart, B. Al-Suwaidan, M. Taber, and S. Wang, "Gravity Probe B: Final Results of a Space Experiment to Test General Relativity," *Physical Review Letters* **106**, pp. 221101-1–5 (2011).

[7] L. Foucault, "Sur les phénomènes d'orientation des corps tournants entraînés par un axe fixe à la surfaces de la terre," *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences* **35**, 424–427 (1852).

[8] J. N. Gibson, *The Navaho Missile Project: The Story of the "Know-How" Missile of American Rocketry*. Schiffer Military/Aviation History, Atglen, PA, 1996.

[9] *Global Positioning System, Selected Papers on Satellite Based Augmentation Systems (SBASs) ("Redbook")*, Vol. VI. ION, Alexandria, VA, 1999.

[10] "GPS Interface Control Document ICD-GPS-200," Rockwell International Corporation, Satellite Systems Division, Revision B, July 3, 1991.

[11] T. A. Herring, "The Global Positioning System," *Scientific American*, February 1996, pp. 44–50.

[12] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *GPS: Theory and Practice*. Springer-Verlag, Vienna, 1997.

[13] R. E. Hopkins, F. K. Mueller, and W. Haeussermann, "The Pendulous Integrating Gyroscope Accelerometer (PIGA) from the V-2 to Trident D5, the Strategic Instrument of Choice," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Montreal, Canada, August 6–9, 2001.

[14] Institute of Navigation, *Monographs of the Global Positioning System: Papers Published in Navigation ("Redbook")*, Vol. I. ION, Alexandria, VA, 1980.

[15] Institute of Navigation, *Monographs of the Global Positioning System: Papers Published in Navigation ("Redbook")*, Vol. II. ION, Alexandria, VA, 1984.

[16] Institute of Navigation, *Monographs of the Global Positioning System: Papers Published in Navigation ("Redbook")*, with Overview by R. Kalafus, Vol. III. ION, Alexandria, VA, 1986.

[17] Institute of Navigation, *Monographs of the Global Positioning System: Papers Published in Navigation ("Redbook")*, with Overview by R. Hatch, Vol. IV. ION, Alexandria, VA, 1993.

[18] Institute of Navigation, *Monographs of the Global Positioning System: Papers Published in Navigation ("Redbook")*, Vol. V. ION, Alexandria, VA, 1998.

[19] J. M. Janky, *Clandestine Location Reporting by a Missing Vehicle*, U.S. Patent 5629693, May 13, 1997.

[20] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *ASME Transactions, Series D: Journal of Basic Engineering* **82**, 35–45 (1960).

[21] M. Kayton and W. L. Fried, *Avionics Navigation Systems*, 2nd ed. Wiley, New York, 1997.

[22] A. Leick, "*GPS: Satellite Surveying*, 2nd ed. Wiley, New York, 1995, pp. 534–537.

[23] J. J. Levy, "The Kalman Filter: Navigation's Integration Workhorse," *GPS World*, September 1997, pp. 65–71.

[24] T. Logsdon, *The NAVSTAR Global Positioning System*. Van Nostrand Reinhold, New York, 1992.

[25] P. F. MacDoran (inventor), *Method and Apparatus for Calibrating the Ionosphere and Application to Surveillance of Geophysical Events*, U.S. Patent 4463357, July 31, 1984.

[26] D. Mackenzie, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*. MIT Press, Cambridge, MA, 2001.

[27] M. W. McMurran, *Achieving Accuracy: A Legacy of Computers and Missiles*. Xlibris, Bloomington, IN, 2008.

[28] F. K. Mueller, "A History of Inertial Navigation," *Journal of the British Interplanetary Society* **38**, 180–192 (1985).

[29] B. W. Parkinson and J. J. Spilker, Jr. (Eds.), *Global Positioning System: Theory and Applications*, Vol. 1, Progress in Astronautics and Aeronautics (series). American Institute of Aeronautics and Astronautics, Washington, DC, 1996.

[30] B. W. Parkinson and J. J. Spilker, Jr. (Eds.), *Global Positioning System: Theory and Applications*, Vol. 2, Progress in Astronautics and Aeronautics (series). American Institute of Aeronautics and Astronautics, Washington, DC, 1996.

[31] B. W. Parkinson, M. L. O'Connor, and K. T. Fitzgibbon, "Aircraft Automatic Approach and Landing Using GPS," in B. W. Parkinson and J. J. Spilker, Jr. (Eds.), Chapter 14 in *Global Positioning System: Theory & Applications*, Vol. II, Progress in Astronautics and Aeronautics (series), Vol. 164, Paul Zarchan editor-in-chief. American Institute of Aeronautics and Astronautics, Washington, DC, 1995, pp. 397–425.

[32] N. Shubin, *Your Inner Fish: A Journey into the 3.5-Billion-Year History of the Human Body*. Random House, NY, 2009.

[33] J. M. Slater, *Newtonian Navigation*, 2nd ed. Autonetics Division of Rockwell International, Anaheim, CA, 1967.

[34] J. F. Wagner, "From Bohnenberger's Machine to Integrated Navigation Systems, 200 Years of Inertial Navigation," in Dieter Fritsch (Ed.), *Photogrammetric Week 05*. Wichmann Verlag, Heidelberg, 2005.

[35] W. Wrigley, "History of Inertial Navigation," *Navigation: Journal of the Institute of Navigation* **24**, 1–6 (1977). Baltimore.