

R	
rate gyros	28
rebalance loop	25
rebalance loops	63
reference ellipsoid	70
restrained gyros	27
rigidity	12, 26
ring interferometer	35
S	
Sagnac effect	35
Schuler loop	56
Schuler period	55
Schuler tuning	55
selective availability	88
self alignment	71
semi-analytic	63
Shuler tuning	55
sidereal day	17, 85
signal generators	30
space loss	82
space stabilized	64
specific force	23, 44
spin stabilization	16
stable platform	29
stable table	29
standard positioning service	88
star-trackers	64
strapdown	63
T	
theorem of Coriolis	10
theoretical mechanics	3
torquers	30
U	
user clock bias	80
V	
vehicle rate	63
W	
wander angle	67
wander azimuth	64, 67
weighted positional number system	4
Z	
zenith	90
zero velocity update	74

gyro	26
gyrocompassing	71
gyroscope	26
H	
handover word	83
I	
inertial	6
inertial forces	11
inertial frames	7
inertial measurement unit	62
inertial reference unit	62
integrating gyros	28
interferometric GPS	96
invariant.	8
L	
laser	38
latitude	69
levelling	71
local vertical	64
locally level	64
lock-in	39
longitude	69
luminiferous ether	35
M	
mask angle	85
mean solar day	18
mechanize	42
meridians	69
meru	73
moving base alignment	76
N	
navigation message	83
northing	70
north-slaved	64
P	
parallels	69
pendulous	25
phase center	98
pickoffs	30
pole problem	66
precession	15, 26
precise positioning service	88
PRN codes	82
pseudoranges	80
pumped	38

Index

A	
aided	73
almanac	83
alongtrack	76
analytic	63
apparent forces	11
augmented	73
B	
base point	64
baseline	81
C	
codeless	98
crosstrack	76
cycle slip	97
D	
damping	73
differential carrier phase tracking	96
direction cosine matrix	52, 54, 63
dither	39
drift trimming	72
E	
earth fixed	64
earth stabilized	64
easting	70
ephemeris data	79
equivalent	8
F	
free azimuth	64
free gyros	27
G	
geodesy	22
geometric	62
geometric decorrelation	95
geometric dilution of precision	89
gimbal lock	30
gimballed	62
gimbals	26
GPS system time	86
gravitation	59
gravity	22, 48
grid	69
group delay	90

- [55] Newton, Isaac, "Mathematical Principles of Natural Philosophy", Mott (1979) and Cajori (trans), University of California Press, Berkeley, 1960
- [56] French, A. P., "Special Relativity", W. W. Norton & Co., 1968
- [57] Singer, C., Holmyard, E., et. al., "The History of Technology", Oxford University, Clarendon Press, 1958
- [58] Beer, F. P, and Johnson, E. R., "Vector Mechanics for Engineers", McGraw Hill, 1977
- [59] Ivey, D. G., "Physics", University of Toronto Press, 1982
- [60] Fowles G. R., "Analytical Mechanics", CBS College Publishing, 1986
- [61] Marion, J., "Classical Dynamics", Academic Press, 1965.
- [62] Kuritsky M. M. , and Goldstein M. S. , "Inertial Navigation", Proceedings of the IEEE, vol 71, no 10, 1156-1176
- [63] McGreevy, J. "Fundamentals of Strapdown Inertial Navigation", Litton Systems Inc, Moorpark Ca, Rev C, May 1986
- [64] Unknown, "An introduction to Inertial Navigation", Litton Systems Inc, Woodland Hills, Ca,
- [65] Broxmeyer, C., "Inertial Navigation Systems", McGraw Hill, 1964
- [66] Parvin, R., "Inertial Navigation", Van Nostrand, 1962
- [67] Farrell, J., "Integrated Aircraft Navigation", Academic Press, 1976.
- [68] Richardson, K., "The Gyroscope Applied", Philosophical Library Inc, New York, 1954
- [69] Scarborough, J., "The Gyroscope", Interscience Publishers, New York, 1958.
- [70] Deimel, R., "Mechanics of the Gyroscope", Dover, 1929.
- [71] Aronowitz, F. , 1978, in "Laser Applications", Academic, New York, pp 113-200
- [72] Martin, G. J. , "Gyroscopes May Cease Spinning", IEEE Spectrum, Feb 1986, pp 48-53
- [73] Chow, W. W. et al, "The Ring Laser Gyro", Reviews of Modern Physics, Vol 57, No. 1, January 1985
- [74] Post, E. J. , "Sagnac Effect", Reviews of Modern Physics, Vol 39, No. 2, April 1967
- [75] Shigeru, O. et al, "Optical Fiber Gyroscopes for Automobiles", Society of Automotive Engineers, Exposition, Detroit, Michigan, March 2, 1990

- [27] A. J. Kelly, "Essential Kinematics for Autonomous Vehicles", CMU Robotics Institute Technical Report CMU-RI-TR-94-14.
- [28] A. J. Kelly, "A Partial Analysis of the High Speed Autonomous Navigation Problem", CMU Robotics Institute Technical Report CMU-RI-TR-94-16.
- [29] A. J. Kelly, "A Feedforward Control Approach to the Local Navigation Problem for Autonomous Vehicles", CMU Robotics Institute Technical Report CMU-RI-TR-94-17.
- [30] A. J. Kelly, "Adaptive Perception for Autonomous Vehicles", CMU Robotics Institute Technical Report CMU-RI-TR-94-18.
- [31] A. J. Kelly, "A 3D State Space Formulation of a Navigation Kalman Filter for Autonomous Vehicles", CMU Robotics Institute Technical Report CMU-RI-TR-94-19.
- [32] A. J. Kelly, "An Intelligent Predictive Controller for Autonomous Vehicles", CMU Robotics Institute Technical Report CMU-RI-TR-94-20.
- [33] A. J. Kelly, "Concept Design of A Scanning Laser Rangefinder for Autonomous Vehicles", CMU Robotics Institute Technical Report CMU-RI-TR-94-21.
- [34] Kleusberg, Alfred and Langley, Richard B., "The Limitations of GPS", *GPS World*, March 1990, pp 50-52
- [35] Kleusberg, Alfred ., "Comparing GPS and Glonass", *GPS World*, November 1990, pp 52-54
- [36] Klobuchar, John A., "Ionospheric Effects on GPS", *GPS World*, April 1991, pp 48-51
- [37] Krakiwsky, Edward J., "GPS and Vehicle Location and Navigation", *GPS World*, May 1991, pp 50-53
- [38] Langley, Richard P., "The GPS Receiver, An Introduction", *GPS World*, January 1991, pp 50-53
- [39] Langley, Richard P., "The Mathematics of GPS", *GPS World*, July 1991, pp 45-50
- [40] Langley, Richard P., "Time, Clocks, and GPS", *GPS World*, November 1991, pp 38-42
- [41] Langley, Richard P., "The Mathematics of GPS.", *GPS World*, July 1991, pp 45-50
- [42] Langley, Richard P., "The Orbits of GPS Satellites", *GPS World*, March 1991, pp 50-53
- [43] Langley, Richard P., "Why is the GPS Signal So Complex", *GPS World*, May 1990, pp 56-59
- [44] Leick, Alfred, *GPS Satellite Surveying*, John Wiley and Sons, 1990.
- [45] Milliken, R. J. and Zoller, C. J., "Principle of Operation of NAVSTAR and System Characteristics", in "GPS: Papers Published in Navigation, Institute of Navigation", pp 3-14
- [46] Nesbo, Inge., and Canter, Peter, "GPS Attitude Determination for Navigation", *GPS World*, September 1990, pp 36-41
- [47] Spilker, J. J. Jr., *Digital Communications by Satellite*, PrenticeHall, Englewood Cliffs, N. J. 1977, pp 495-60
- [48] Spilker, J. J. Jr., "GPS Signal Structure and Performance Characteristics", in *GPS: Papers Published in Navigation, Institute of Navigation*, pp 29-54
- [49] Stanzell, T. A. Jr., "Civil Marine Applications of the Global Positioning System", in *GPS: Papers Published in Navigation, Institute of Navigation*, pp 132-143
- [50] Unknown, "Civilian GPS Briefing", Magnavox Corporation
- [51] Unknown, "GPS: A Guide to the Next Utility", Trimble Navigation
- [52] F. van Graas, M. Braash, "Real-Time Attitude and Heading Determination Using GPS.", *GPS World*, March 1992, pp32-36
- [53] Zachmann, George W., "Differential GPS Transmissions By Geostationary L-Band Satellites", *Sea Technology*, May 1990, pp 57-63
- [54] Einstein, Albert, "Relativity - The Special and the General Theory", translated by Robert W. Lawson, Crown Publishers, New York, 1961

5 Bibliography

- [1] Broxmeyer, C., "Inertial Navigation Systems", McGraw Hill, 1964
- [2] Britting, K., "Inertial Navigation System Analysis", Wiley, 1971.
- [3] Draper, C. S., W. Wrigley, and J Hovorka: "Inertial Guidance", Pergammon Press, New York 1960
- [4] Farrell, J., "Integrated Aircraft Navigation", Academic Press, 1976.
- [5] Fernandez M., and Macomber G. R., "Inertial Guidance Engineering", Prentice Hall, Englewood Cliffs, NJ, 1962
- [6] Ivey, D. G., "Physics", University of Toronto Press, 1982
- [7] Kuritsky M. M. , and Goldstein M. S. , "Inertial Navigation", Proceedings of the IEEE, vol 71, no 10, 1156-1176
- [8] McGreevy, J. "Fundamentals of Strapdown Inertial Navigation", Litton Systems Inc, Moorpark Ca, Rev C, May 1986
- [9] Odonnell, C. F. (ed.), "Inertial Navigation Analysis & Design", McGraw Hill, New York, 1964
- [10] Parvin, R., "Inertial Navigation", Van Nostrand, 1962
- [11] Pittman, G. R. Jr (ed.), "Inertial Guidance", John Wiley Z& Sons Inc, New York 1962.
- [12] Sutherland, A. A. Jr, and Gelb, A., "Application of the Kalman Filter to Aided Inertial Systems", Analytical Sciences Corporation, Winchester, Mass.
- [13] Savage, P. G., "Strapdown Systems Algorithms", AGARD Lecture Series No. 133, Strapdown Associates Inc., Minnetonka, Minnesota.
- [14] Savant C. J. Jr. , R. C. Howard, C. B. Solloway, and C. A. Savant, "Principles of Inertial Navigation", McGraw Hill Book Company, Inc, New York, 1961
- [15] Unknown, "An Introduction to Inertial Navigation", Litton Systems Inc, Woodland Hills, Ca,
- [16] Inertial Guidance, Terrestrial and Interplanetary, Course 16.455 Notes. MIT Department of Aeronautics & Astronautics, 1960.
- [17] Ashkenazi, Vidal, and Ffoulkes-Jones, Geraint, "Millimeters Over Hundreds of Kilometers by GPS", GPS World, November 1990, pp 44-47
- [18] Arrondo-Perry, Jacki, "GPS World Receiver Survey", GPS World, January 1992, pp 46-58
- [19] Colombo, Oscar L., and Peters, Mary F., "Precision Long Range DGPS for Airborne Surveys", GPS World, April 1992, pp 44 - 50
- [20] Cox, D. B. Jr., "Integration of GPS with Inertial Navigation Systems", in GPS: Papers Published in Navigation, Institute of Navigation, pp 144-153
- [21] Georgiadou Yola, and Doucet, Kenneth D., "The Issue of Selective Availability", GPS World, September 1990, pp 53-56
- [22] Glazer, B. G., "GPS Reciever Operation", in "GPS: Papers Published in Navigation, Institute of Navigation", pp 81-86
- [23] Gold, R., "Optimal Binary Sequences for Spread Spectrum Multiplexing", IEEE Trans on Information Theory, Oct 1967, pp 619-621
- [24] Henderson, D. W., and Strada, J. A., "NAVSTAR Field Test results", in "GPS: Papers Published in Navigation, Institute of Navigation", pp 234-246
- [25] Hobbs, Richard R., "Marine Navigation. Piloting and Celestial and Electronic Navigation." Third Edition, Naval Institute Press, Annapolis, Maryland. 1990.
- [26] Kalafus, Rudolph M et al., "Differential Operation of NAVSTAR GPS", in "GPS: Papers Published in Navagtion", Institute of Navigation, pp 197-214

standard positioning service - the less accurate positioning service available to civilians.

space loss - the absorption of radio energy as it passes through the atmosphere.

user clock bias - the offset of a receiver's clock from GPS system time.

zenith - directly overhead.

Schuler loop - the loop which employs gravity feedback in order to level a table.

Schuler period - the period of oscillation of the Schuler pendulum, or 84 minutes.

Schuler tuning - the process of designing a system to exhibit the Schuler natural frequency.

self alignment - the process of aligning an INS without the aid of external measurements.

semi-analytic - a designation for systems which instrument only the navigation frame.

signal generator - see pickoff.

space-stabilized - designation for systems which maintain constant inertial platform orientation.

specific force - the net inertial acceleration minus gravitational attractions.

stable platform - a platform whose orientation is stable in a particular reference frame.

strapdown - see analytic.

star-tracker - a device which tracks and reports the bearing of a star.

torquer - device for applying torque to a gyro or gimbal.

vehicle rate - the angular velocity of the vehicle with respect to the center of the earth.

wander angle - the rotation of the coordinate axes with respect to north.

wander azimuth - a designation for a system which torques the azimuth to follow the earth rate only.

almanac - imprecise satellite ephemeris data broadcast to allow receivers to predict when satellites will rise above the horizon.

baseline - the distance between two antennae in differential GPS.

codeless - a designation for receivers which compute position from only the signal carriers.

cycle slip - a temporary loss of lock in the carrier tracking loop of a receiver.

differential carrier phase tracking - a technique which improves accuracy by measuring the carrier phase shift.

ephemeris data - the data specifying the orbital position of a satellite.

geometric decorrelation - the degree to which system error is a nonuniform function of position.

geometric dilution of precision - the dilution of precision which occurs when pseudoranges are transformed into cartesian position.

GPS system time - the time standard to which all satellites are actively synchronized by MCS.

group delay - the propagation delay when radio passes through the atmosphere.

handover word - a component of the data message which facilitates the transfer of tracking from C/A code to P code.

interferometric GPS - see differential carrier phase tracking.

mask angle - the elevation angle to the horizon above which a satellite is practically visible.

navigation message - a digital data message modulated onto both carriers which provides ephemeris data, satellite clock offsets, and many other useful pieces of information.

phase center - the point on the antenna which is actually being positioned by the receiver.

precise positioning service - the most precise dynamic positioning service available with GPS.

pseudorange - satellite to receiver range errors which include the user clock bias.

PRN codes - digital codes which have characteristics similar to noise.

selective availability - the deliberate degradation of ephemeris data accuracy.

sidereal day - the time required for the earth to rotate once with respect to the stars. It is shorter than the mean solar day by four minutes.

4 Glossary

aided inertial systems - inertial systems which make use of redundant external measurements of navigation quantities.

alongtrack - the direction parallel to the direction of motion.

analytic - a designation for systems which compute rather than instrument the navigation frame quantities.

base point - designation for system which maintain the platform orientation at the initial orientation with respect to the earth.

crosstrack - the direction perpendicular to the direction of motion, measured horizontally.

damping - the process of filtering redundant external measurements with the INS output in order to reduce the magnitude of system error.

direction cosine matrix - the matrix which describes the transformation of the body axes into the navigation frame.

earth fixed - see base point.

earth-stabilized - designation for systems which maintain a navigation frame which is defined with respect to the earth in some way.

easting - eastern coordinate in UTM coordinates.

free azimuth - a designation for systems providing no platform rate about the azimuth.

geometric - designation for systems which perform geometric stabilization utilizing a stable platform.

gimballed - designation for systems incorporating a stable platform.

gimbal lock - singularity of the platform gimbal mechanism.

gravity - vector sum of gravitation and centrifugal force.

grid - a regular coordinate grid placed on a map. Unlike latitude and longitude, grid lines are parallel and equally spaced.

gyrocompassing - the process of using the earth's spin to find the direction of north.

inertial measurement unit - the sensor suite consisting of the gyros and accelerometers.

inertial reference unit - see inertial measurement unit.

latitude - angle of a point on the earth measured from the equatorial plane.

levelling - the process of levelling a stable platform during alignment.

locally level - designation for systems which maintain the platform level.

local vertical - see locally level.

longitude - angle of a point on the earth measured from the prime meridian.

mechanization - the process by which the navigation equations are implemented in hardware and software.

meridian - a line of constant longitude.

northing - northern coordinate in UTM coordinates.

north-slaved - designation for systems which physically track the direction of north by rotating the platform.

parallel - a line of constant latitude.

pickoff - the transducer which measures a gimbal angle.

pole problem - the degradation of performance experienced by inertial systems as they approach the earth's poles.

rebalance loops - a control loop which actively prohibits deflection of a compliant transducer.

reference ellipsoid - an idealized mathematical model of the earth.

3.6.2 Configuration

Some units are available for stand-alone use; others are configured in a printed circuit board set. There are also many complete navigation systems available incorporating integrated GPS and inertial navigation systems. Volume, weight, power consumption, port baud rates, and environmental hardness can vary significantly.

3.6.3 Input/Output

Some sets report position, velocity, and time and some report only position. Special purpose multi antenna sets can measure orientation. It may be necessary to accept differential range corrections according to a standard protocol. Some differential sets are intended for postprocessed surveying applications only and postprocessing software may or may not be supplied.

Of particular importance is the coordinate systems in which position can be reported. Options include cartesian earth-centered coordinates; latitude, longitude, and altitude on a reference ellipsoid; easting, northing and altitude; and range, bearing, speed, and crosstrack error with respect to a stored sequence of waypoints.

3.6.4 Antenna

Several types of antenna are available including monopole or dipole configurations, quadrifilar helices, spiral helices, and microstrips. Provision for an external antenna is not always available. For precision applications, an important parameter is the stability of the antenna **phase center**, the electrical center of the antenna to which the receiver output actually refers.

3.6.5 Channels

Most receivers will operate solely on the C/A PRN code, and the number of channels of radio input can vary significantly. Some sets integrate GPS and Glonass tracking. Precision **codeless** sets use only the carrier signal. A set with several channels can track several satellites simultaneously, rather than sequentially, and since it will exhibit better signal to noise ratios, it can track them closer to the horizon.

3.6.6 Update Rate

Update rates and the acceptable vehicle velocity and acceleration increase with the number of channels. Some sets require an initial estimate of position to start, and others can cold start. The time to the first fix can vary from seconds to 30 minutes depending on the initial information supplied and the receiver mode of operation.

of the line of sight requires that the system be reinitialized. Receivers may sometimes miss a carrier cycle in their computations, and special techniques are employed to eliminate this **cycle slip** problem. Accuracies increase by an order of magnitude at the cost of limited excursion, higher cost equipment, and sometimes reduced update rate. Relative accuracies of a few parts per million of excursion are the rule. For systems measuring rotation on a rigid vehicle, 1 milliradian of accuracy is achievable over a baseline of 1 meter.

The advantages and disadvantages of each mode of operation are summarized in Figure 39

Mode	Accuracy	Range	Why
Absolute	120 m	global	
Relative	100 m	local	remove position and time independent bias
Repetitive	100 m	global	removes time independent bias
Coded Differential	3-5 m	20 Km	remove position indep and time dependent bias
Codeless Differential	2 ppm	10 Km	bypasses codes and uses interferometry

Notes:
a) All coded modes using C/A code, update rate ~ 1 Hz
b) Bias is error varying slower than 1 Hz

Figure 39 GPS Modes of Operation

3.6 GPS Receivers

The generic GPS receiver consists of an antenna and associated preamplifier, an RF section, tracking loops, a microprocessor, and optionally includes its own power supply, a command and display unit, and an external data and control port. Receivers exist which are designed to suit a number of diverse applications including aviation, land vehicles, marine, military, surveying, timing, and even handheld varieties. Some of the important features of commercial sets are summarized below.

3.6.1 Application

For robotics applications, a unit which is designed as a component of a larger navigation system is appropriate. This often implies facilities for external power, communication with an external computer, an external antenna, and significant internal storage may or may not be required.

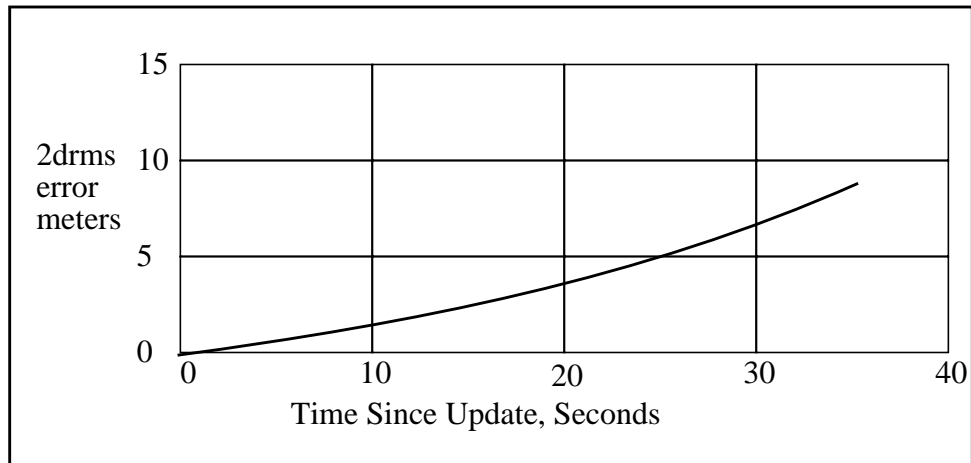


Figure 37 Temporal Variation of System Error

3.5.5 Codeless Differential GPS

This technique relies on bypassing the PRN codes entirely and using only the high resolution carriers. **Differential carrier phase tracking** consists of measuring the phase shift between the same signal received at two different locations. The phase shift is a direct measure of the difference in path length from the satellite to each antenna. Also known as **interferometric GPS**, the principle is indicated in Figure 38.

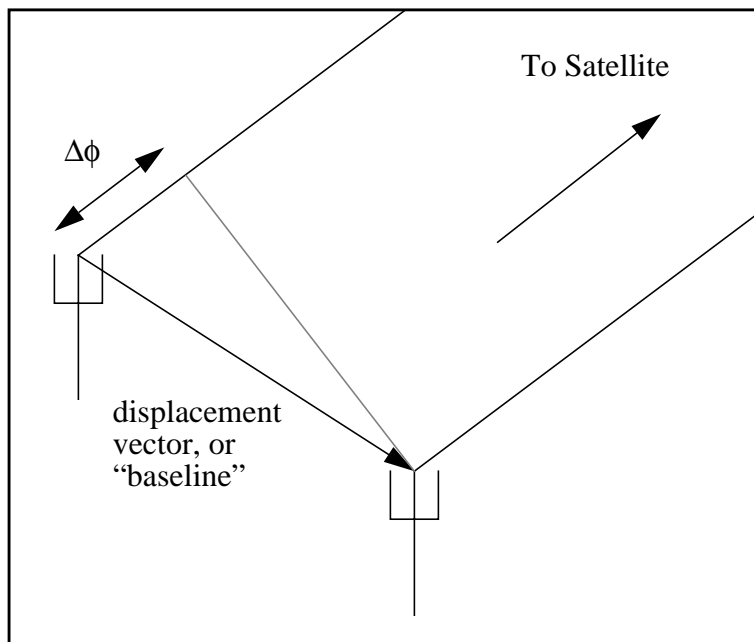


Figure 38 GPS Interferometry

Such measurements are subject to the phase ambiguity problem and special techniques are used to resolve it. Also, since a continuous lock must be kept on the carriers at all times, any interruption

Differential techniques can remove only that component of error which is common to both the reference and the moving set. Pseudorange error varies with both position at a fixed time and with time at a fixed position. Position variation is primarily due to the dependence of atmospheric delay and ephemeris errors on elevation angle. Luckily, this **geometric decorrelation** is relatively insignificant because of the small change in elevation angle associated with moderate excursion from the reference. This is a direct result of the substantial altitude of the spacecraft as indicated in Figure 36. The remaining uncanceled error is roughly linear in distance. Useful differential pseudorange corrections can be obtained over distances as large as 1,000 kilometers.

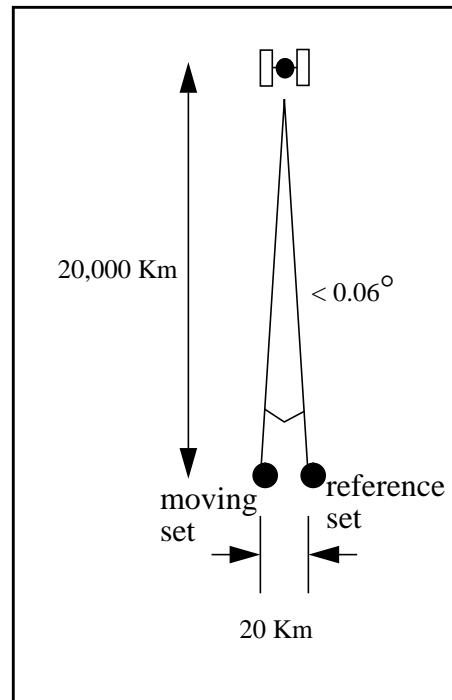


Figure 36 Spatial Variation of System Error

While spatial variation is insignificant, variation over time is a major concern. The finite time required to transmit and process pseudorange errors imply that the moving receiver will always have a slightly dated estimate available. The S/A range error rate is such that pseudorange error data is useless after one minute of elapsed time. Accordingly, the update rate of the secondary data link is a critical parameter. For a stationary receiver near the reference, 2drms error will vary over time as indicated in Figure 37.

3.5.4 Coded Differential GPS

This technique uses a second fixed receiver as the origin of coordinates. If the earth centered, earth fixed location of this reference is known to high accuracy, then high accuracy absolute positioning of the moving receiver is possible. If not, then only high accuracy positioning relative to the reference is possible. This second case is sufficient for most robotics applications. This mode of GPS use is particularly important, so some space will be devoted to acquainting the reader with the issues.

Since the position of the reference receiver is known in the navigation coordinate system, it can compute the difference between this and the GPS reported position in real time. Differential techniques rely on a second communications channel which allows the reference to broadcast these differences to the moving receiver at a high rate. If omnidirectional radio is used to transmit the position error, many mobile receivers can benefit from a single reference⁶⁹. Tethered vehicles can transmit differential corrections through the tether. The general situation is indicated in Figure 35.

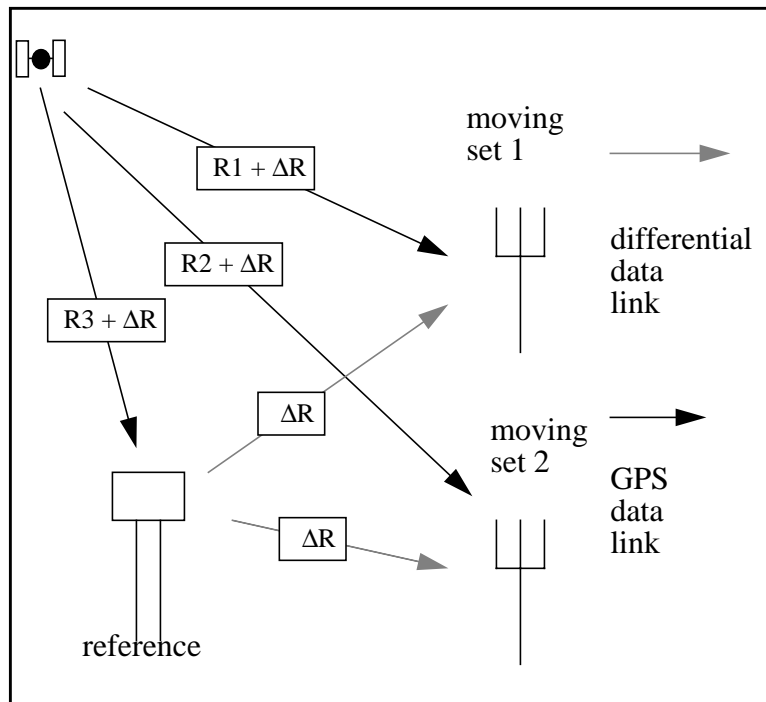


Figure 35 Differential GPS

In implementation it is not advisable to transmit cartesian positioning error from the reference since the moving sets may use a different navigation solution, or may even use a different group of satellites. Rather, because the reference has no knowledge of which satellites will be used, it must compute the pseudorange error for all satellites in view and transmit these to the moving receivers.

69. Local commercial differential GPS transmitters will be operating in urban centers soon. The standard protocol for differential service is known as RTCM-SC-104.

3.5 Modes of Operation

Several modes of use are available in mobile robotics applications. All applications can be classified first according to whether the GPS receiver is used as the real-time motion control sensor, or as a redundant sensor for site surveying applications. In the latter case, off-line data reduction potentially combined with long measurement dwell times can be exploited to remove most of the random error in measurement. This is a highly specialized application, so the reader is referred to the bibliography for further reading. This section will consider real-time positioning accuracy only.

Receivers can compute a positioning fix at a maximum rate of about once per second, so this is the closest to real-time performance available. This implies that certain highly dynamic applications will require at least a rudimentary dead reckoning system to augment the GPS receiver. Below this maximum update rate, the whole spectrum of averaging times can be used to eliminate random error provided vehicle dynamics permit it. However, note that civilian SPS service users cannot remove the relatively low frequency selective availability error unless dwell times on the order of hours are used. Only the higher frequency random errors can be removed quickly enough for real-time positioning.

Having eliminated long dwell times from consideration, the following operational modes remain.

3.5.1 Absolute GPS

This mode involves simply adopting an earth centered, earth fixed coordinate system for high level motion control. Receivers report position in this system most naturally. In this mode, nominal SPS performance applies to civilian users. It has the advantage that this accuracy applies over the entire surface of the earth.

3.5.2 Relative GPS

It may seem that for robotic applications, the system's inherently higher relative accuracy can be exploited by simply adopting the initialization point as the origin of coordinates. However, this technique can only remove bias that is independent of time over the mission duration, and independent of position over the mission excursion. S/A errors vary slowly over time, and they are the dominant component of error, so relative accuracy will degrade to levels close to the absolute accuracy over a time period of only a few minutes.

3.5.3 Repetitive GPS

It may also seem that the system's inherent repeatability can be exploited in applications where the vehicle can simply repeat the acquisition of previously surveyed waypoints. Again, the selective availability error limits the improvement in accuracy achievable with this approach because GPS will report significantly different positions for the same physical point on the earth over a period of just a few minutes.

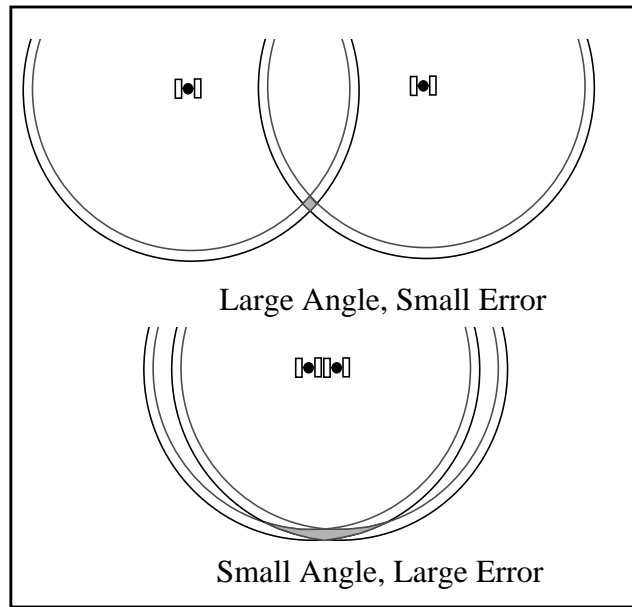


Figure 34 Geometric Dilution of Precision

In the context of GPS, five terms are defined:

- TDOP - time dilution of precision (range equivalent)
- PDOP - position dilution of precision (3D)
- HDOP - horizontal dilution of precision
- VDOP - vertical dilution of precision
- GDOP - geometric dilution of precision

These are related by:

$\text{GDOP} = \sqrt{(\text{PDOP})^2 + (\text{TDOP})^2}$	Equation 42
$\text{PDOP} = \sqrt{(\text{HDOP})^2 + (\text{VDOP})^2}$	Equation 43

For all users averaged over the globe, and averaged over time, typical values of PDOP, HDOP, and TDOP are 2.6, 1.5, and 1.2. Good receivers will compute the GDOP for particular combinations of satellites and choose the best set.

since water is a perfect radio reflector under certain conditions, but it is usually less than 1 meter. As indicated in Figure 33, the time delay for ground reflection depends on altitude. Multipath error is more pronounced when the receiver is close to the surface. Therefore, it is usually advisable to mount GPS antennae as high as possible.

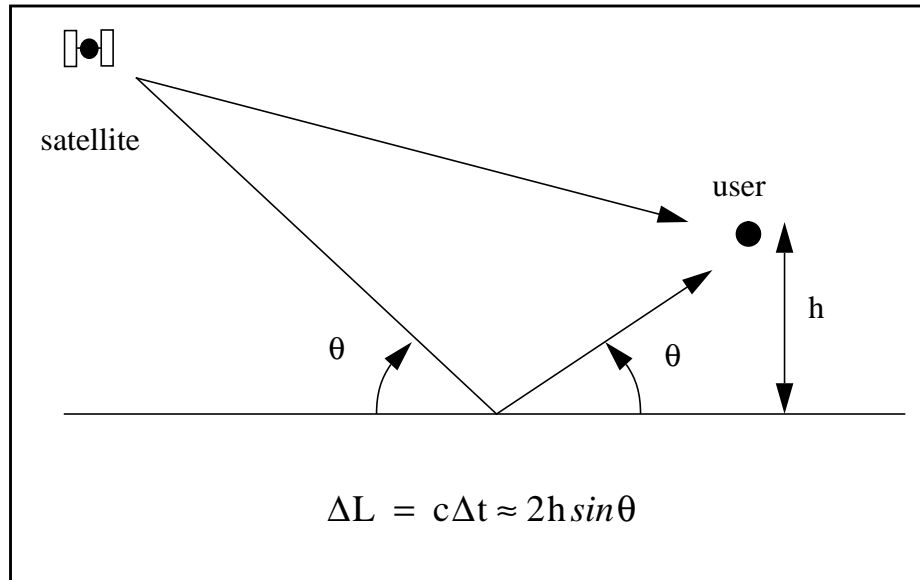


Figure 33 Multipath Ground Reflection

Other less important sources of error include the unintentional component of satellite ephemeris and clock errors, imperfections in receiver circuitry, computational approximation and truncation, and vehicle dynamics. The effect of vehicle motion can be substantial if receiver specifications are exceeded.

3.4.2 Geometric Dilution of Precision

The final GPS positioning accuracy is strongly dependent on the geometry of the satellite lines of position used. The transformation from four pseudorange errors into positioning errors is accomplished by multiplying by a geometry factor known as the geometric dilution of precision. This GDOP occurs in all triangulation problems and is defined as:

$$\text{DOP} = \frac{\text{FixError}}{\text{RangeError}} \quad \text{Equation 41}$$

where the errors are represented as standard deviations. The DOP scales the range error into the fix error, so small values indicate favorable geometry.

The concept is easy to visualize in two dimensions. Consider Figure 34. Range errors, shown as shaded regions, combine to form a roughly rectangular region of uncertainty in the fix. The greater the angle between the satellites used, the smaller is the region of uncertainty. It can be formally shown that there is a direct relationship between uncertainty and the volume of this region.

Error Source	Nominal Value (rms)
Selective Availability	8 meters
Atmospheric Delays	4 meters
Satellite Clock & Ephemeris	3 meters
Multipath	1 - 3 meters
Receiver Electronics & Vehicle Dynamics	1.5 meters
TOTAL	10 meters

Notes:
a) pseudorange errors only, GDOP (of 4 - 6) is not included
b) multiply by 2 for 2drms, by 0.8 for CEP
c) errors combined in root sum square sense

Figure 32 Pseudorange Error Sources

It should be clear from this table that the GPS design reserves the ability to remove the two most important errors for exclusive military use. With the exception of S/A, the deliberate signal degradation discussed earlier, atmospheric delay is the most important pseudorange error source. It is also called **group delay**. It varies with time and place. It can result in as much as 30 meters equivalent range error when the satellite is at **zenith**.

Atmospheric delay consists of two components: ionospheric and tropospheric. The ionospheric delay is caused by diffraction due to charged particles. It can vary by a factor of 5 from day to night, being highest during the day, and by a factor of 3 due to the elevation angle of the satellite. This means that each satellite has a different, time varying pseudorange error. It depends on solar magnetic activity and geomagnetic latitude, being particularly affected during magnetic storms and it is usually greatest at the poles, and at the equator.

Tropospheric delay is caused by water and other atmospheric constituents giving rise to local changes in the index of refraction. It can be modelled very well by a simple mathematical formula. It also varies with time and place, giving about 2.3 meters delay vertically and ten times this at the horizon.

Multipath error arises when some of the radio signal reaches the receiver through reflection in addition to the portion which arrived along the direct line of sight. The result of both signals arriving at the antenna is destructive interference. This error source can be substantial above water,

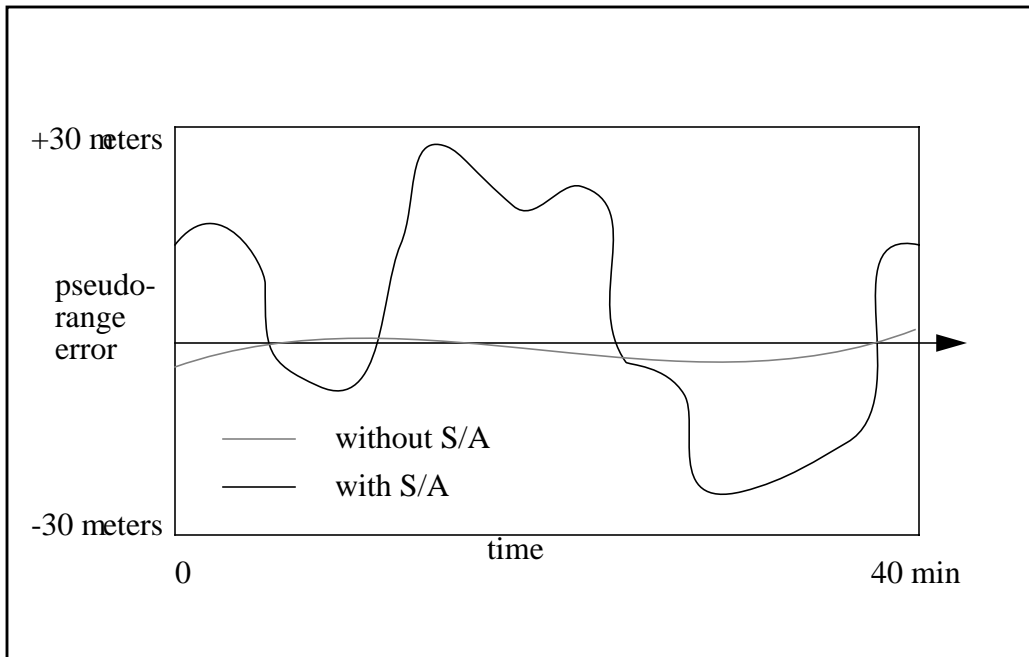


Figure 31 Selective Availability

3.4 Sources of Error

In order to assess the possible accuracies available in an application, an understanding of the sources of error and when they apply is necessary. The total error in the positioning fix can be separated into two components: the errors in the satellite pseudoranges, and the projection of these errors onto the navigation coordinate system. The first component involves physics; the second is purely a mathematical transformation.

Pseudorange error sources are additive. The **geometric dilution of precision**, or GDOP is multiplicative. GDOP can be as high as 4 or much more under certain circumstances⁶⁸, so it is important to know whether precision estimates include GDOP or not, and if they assume a particular value. It is also important to ascertain whether a particular precision estimate is horizontal or 3D.

3.4.1 Sources of Pseudorange Error

The significant sources of pseudorange errors and their nominal values are summarized in Figure 32

68. During the initial years when the GPS satellite constellation is incomplete, GDOP values as high as 20 are possible.

3.2 Measures of Accuracy

There are two measures of accuracy that are commonly used to describe GPS performance. The military commonly uses the Circular Error Probable (CEP) in two dimensions, and the Spherical Error Probable (SEP) in three dimensions. These measures are defined to be the median value of accuracy. Hence 50% of measurements can be expected to fall within the SEP of the true value and 50% will fall outside.

The scientific community prefers the standard deviation⁶⁷ as a measure of accuracy. In GPS circles, the 2drms value or twice the standard deviation is often used. Between 95% and 98% of measurements can be expected to fall within the 2drms value of the true position, depending on the ellipticity of the distribution. Both measures of accuracy are related by the approximate formula:

$$2 \times \text{drms} \approx 2.5 \times \text{CEP}$$

Equation 40

3.3 Selective Availability

The nominal precision of GPS service is precisely regulated and deliberately controlled by the Department of Defense. Two levels of positioning service are distinguished for GPS. The **standard positioning service** (SPS) is designated for civilian use and has a nominal precision of 120 meters 2drms in 3 dimensions. SPS is based on reception of the C/A PRN code. Velocity measurement is intended to be accurate to 0.3 meters/second, and time is accurate to 300-400 nanoseconds.

The **precise positioning service** (PPS) is designated for US and NATO military use and has a nominal precision of 10 meters 2drms in three dimensions. PPS is based on the simultaneous reception of the encrypted P code on both carriers. Only those users who have knowledge of the encryption will have access to the higher PPS precision.

Through a technique known as **selective availability** (S/A) the physically achievable 40 meter precision of SPS is deliberately degraded to the 120 meter nominal value. This is achieved by artificially introducing errors into the broadcast ephemeris data, and by dithering the satellite clock. As a result, S/A errors have both a slowly varying and a rapidly varying component. For a fixed receiver, measured pseudorange will vary apparently randomly over time as shown in Figure 31.

67. The use of standard deviation or CEP as a measure of “accuracy” is a misnomer. Both measure the spread of the data, not the difference between the sample and population means. CEP and 2drms really measure precision. Only in the case of a single measurement in the absence of measurement bias does the data spread relate to accuracy.

3 Performance

The performance of GPS can be characterized along three independent dimensions: the precision and accuracy of the fix, the rate at which it is updated, and the range over which the accuracy can be achieved. For robotic surveying applications, it is reasonable to allow the vehicle to dwell at a point of interest for a considerable time so that a more accurate fix is achievable. It may also be feasible in certain applications to postprocess positioning data off-line for a further improvement in accuracy, provided a redundant positioning system is used for control of the vehicle itself.

If GPS is used to determine vehicle position in real time for navigation purposes, high accuracy must be traded against the update rates necessary for real time control. Further, if the vehicle's range of motion is restricted by the application, differential techniques can be used to improve accuracy considerably, at the cost of a second receiver, a second communications link, and the limited range.

The accuracy achievable at any update rate is also a function of the signal combinations used and the processing performed on them. The inherent resolutions of the C/A, P, and carrier signals vary from meters, to decimeters, to millimeters. It is not the intention here to provide a quantitative assessment of GPS accuracy in any of its modes of use. Rather, this section will equip the reader to understand the factors that affect accuracy to a degree that will make informed design decisions possible.

3.1 Range Resolution

The fundamental range resolution available in any mode of operation is related to the signals used. Specifically, achievable range resolution is a fraction of the signal wavelength. The signal wavelengths are summarized below:

Table 1: Signal Wavelengths

Signal	Frequency	Wavelength
L1	1575.42 MHz	20 cm
L2	1227.60 MHz	24 cm
C/A code	1 MHz	300 m
P code	10 MHz	30 m

2.4 Ground Segment

The GPS ground segment consists of a series of five ground stations spaced in longitude around the globe. One of the stations is designated the Master Control Station (MCS). The function of GPS Master Control is to track the positions of all satellites very precisely and to maintain the overall system time standard.

Space vehicle positions are very predictable due to the negligible effect of atmosphere at orbital altitude. Nevertheless small perturbing forces exist which are caused by the oblateness of the earth, the presence of the moon, and even the pressure of solar radiation accumulated over time. The effect of these forces is to perturb the satellite orbits from their ideal Keplerian elliptical shape.

Inverting the ranging process, considering MCS to be the user and the ground stations to be the satellite, makes it possible to determine the position of each satellite very accurately. MCS updates the satellites on their orbital parameters at least once per day so that precise information can be later transmitted to the receivers.

Satellites must be synchronized in time to a precision of nanoseconds to avoid the range errors that arise from timing errors. To do this, MCS maintains an overall system time standard known as **GPS system time** through the use of highly accurate cesium atomic clocks⁶⁶. Also, each satellite incorporates two cesium and two rubidium atomic clocks for redundancy.

Atomic clocks are used since the more common crystal oscillators are not stable enough for this application. MCS regularly transmits to each satellite the offset of its clock from GPS system time. Satellite clock offsets are then retransmitted to receivers so that the precise GPS system time of signal emission is known. Overall constellation control is indicated in Figure 30.

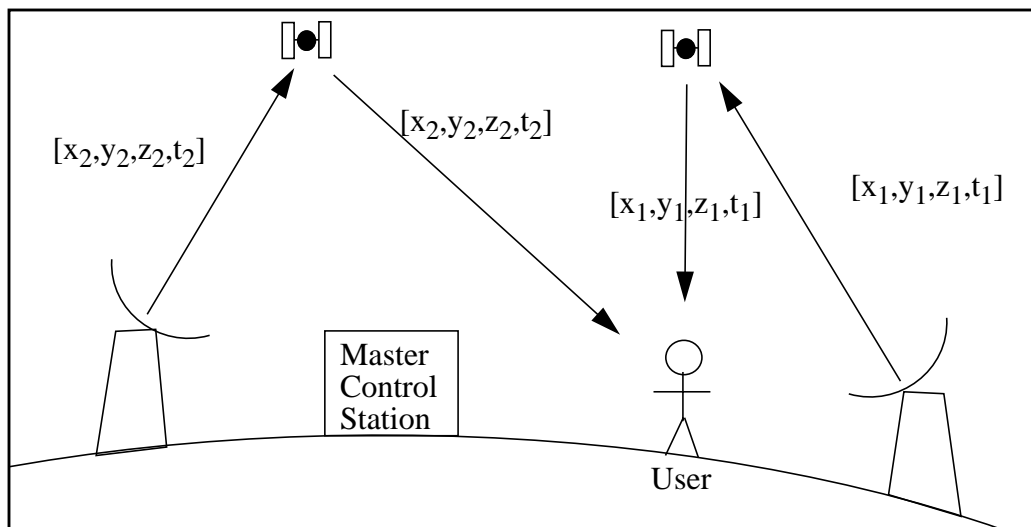


Figure 30 GPS System Master Control

66. The atomic clock is not based on the same processes as nuclear energy. Rather, the ability of the atom to resonate as it changes its quantum energy states is exploited. In the case of cesium, the magnetic dipole of the sole outermost electron is switched from being parallel to and antiparallel to the dipole of the nucleus. The principles used are similar to those of a laser or a microwave oven.

2.3 Space Segment

The GPS space segment consists of a constellation of 18 earth satellites⁶⁴ and three spares that circle the earth in each of six orbits as shown in Figure 29. The six orbital planes are each inclined at 55° to the equatorial plane, and separated by 60° in longitude. Orbits are nearly circular, 11,000 miles in altitude and repeat exactly twice per **sidereal day**. This implies that the satellites are in exactly the same place four minutes later every day⁶⁵.

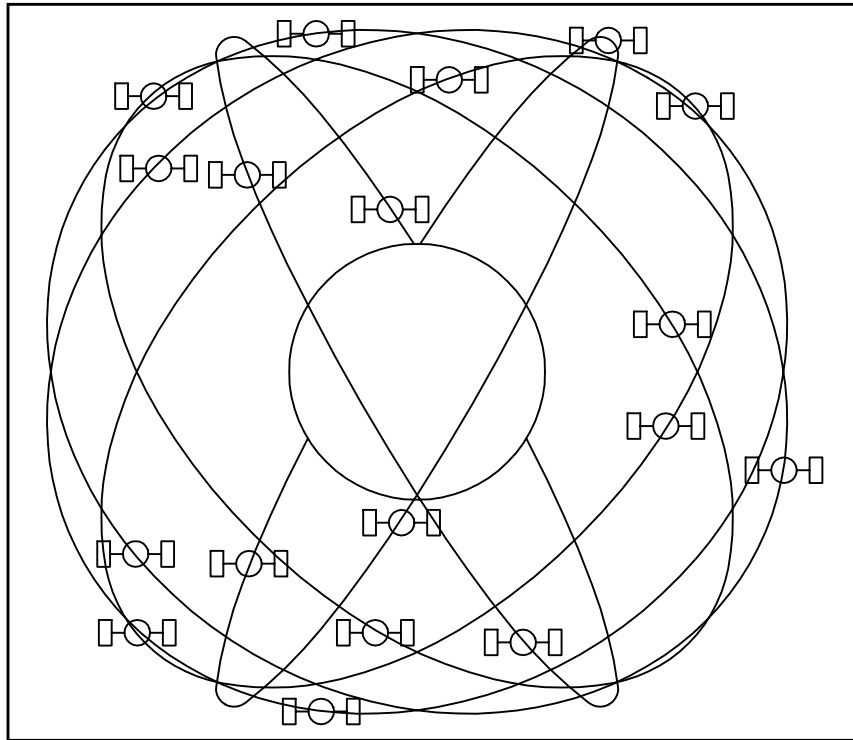


Figure 29 Orbital Configuration

The constellation is designed to ensure that at least four satellites are visible at any time. Satellites are considered visible above the nominal **mask angle** of 5° - 10° . When the elevation angle above the horizon exceeds the mask angle, the signals are strong enough for good reception. However, foliage or other attenuation may increase the mask angle significantly.

Use of GPS requires a clear line of sight, and the signals cannot penetrate water, soil, or walls very well. Hence, GPS cannot be used underwater, in thick forest, or in mines and tunnels. Signals can be obstructed by tall buildings, or by parts of a rotating vehicle if the receiver antenna has not been placed carefully.

64. According to NORAD, there are currently about 7000 detectable objects in earth orbit.

65. It is possible to have a satellite appear to be fixed over the equator by matching its orbital period with the earth's rotation. The orbital period of a circular orbit is a function of only the altitude. A geosynchronous altitude is much higher than that of the GPS satellites. Geosynchronous space is rapidly filling up with satellites.

2.2 User Segment

The GPS user segment consists of the receivers in use on the surface and in the near space region. The basic functions of the receiver are to perform code correlation in order to measure propagation time and clock bias, to perform compensation for known errors, and to implement the solution to the navigation equations. Receivers internally generate exactly the same codes as the satellites. Time of arrival measurement is accomplished by shifting the internal codes in time and computing the correlation with the externally received codes as shown in Figure 28.

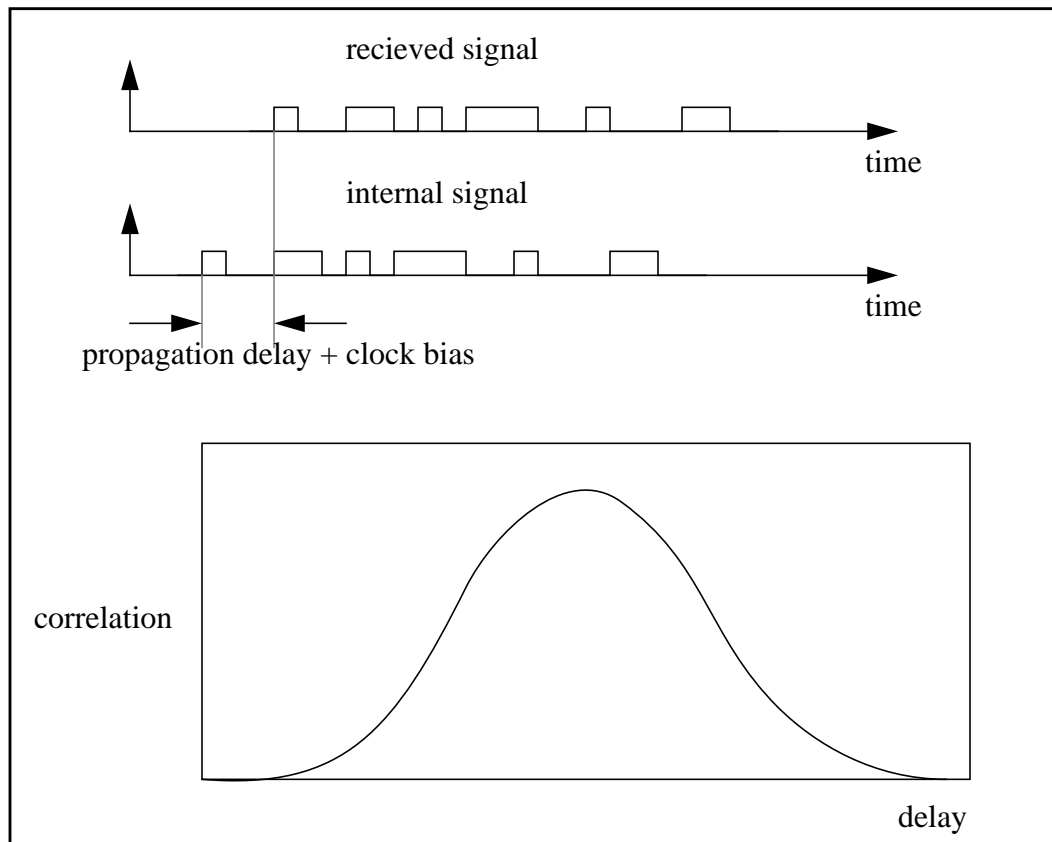


Figure 28 Receiver Code Correlation

The C/A code repeats once every millisecond while the P code period is one week long. The C/A code is short, so the amount of shifting required is limited, and a match can be found quickly. A match occurs when the correlation function peaks. Once a match is computed, the code is said to be acquired. Once acquired, the time of signal propagation is given by the time shift that was necessary. This time includes the clock bias which is removed later. The next step is to derive the wave speed.

Receivers can compensate for atmospheric delay to compute an accurate wave speed by using either mathematical model coefficients transmitted by the satellites, or by using two carriers. Multiplying this by the propagation time gives the pseudorange to the satellite. When this process is repeated for four different satellites, the receiver can solve the navigation equations and determine its position.

PRN codes have two other advantages. They are particularly suited to the multiple access requirement. Each satellite transmits a different code, so the receiver can distinguish between the signals of all satellites which arrive at its antenna simultaneously. Also, PRN codes are particularly impervious to deliberate or unintentional jamming, so they support the need for secure military access.

There are two modulation signals used. The civilian signal, called C/A for clear access or coarse acquisition, has a 300 meter pulse wavelength. The military signal is designated P for precise, and is called Y when it is encrypted. It has a pulse wavelength of 30 meters. Pseudorange resolution is directly related to the wavelength of the code, so the civilian code has a lower intrinsic precision. The civilian C/A code is mixed with the L1 carrier while the P code is mixed with both carriers and broadcast redundantly.

The use of two codes, one of which is encrypted, gives the military access to a completely separate, secure, positioning system. A second reason for using two codes is that, since the P code is difficult to acquire, the C/A was designed for easy acquisition. Once the C/A code is acquired, the P code is easier to acquire. Code acquisition is discussed in the next section.

2.1.3 Navigation Message

In addition to the PRN codes, a second block of information is digitally broadcast on both carriers. This **navigation message** is transmitted at the very slow rate of 50 bits per second, and is updated and repeated every 12.5 minutes. The navigation message contains a large amount of information provided to assist receivers in their operation.

The system time of the week, known as the **handover word**, is provided to assist in acquiring the P code after the C/A code is acquired. The transmitting satellite provides precise ephemeris data on its own orbit for use in the navigation solution. Also, less accurate ephemeris data for all other satellites, known as the **almanac**, is transmitted to permit receivers to predict when new satellites will rise above the horizon and become available. The ephemeris data is described by the classical ellipse parameters discovered and defined by Kepler.

Atmospheric delay model coefficients are provided since the magnitude of the delay varies significantly throughout the day. Satellites also provide information on their own health and the precision of range measurements that can be expected if their signals are used. In this way, receivers can pick an optimal set of satellites in view for their computations. The complete set of GPS modulation signals is summarized in Figure 27:

	C/A	P	NAV
L1	✓	✓	✓
L2		✓	✓

Figure 27 GPS Signals

2 Implementation

The navigation signals used by GPS satellites will seem overly complex to the reader unless the reasons for this complexity are presented first. The GPS signals owe their complexity to the requirements that the system was designed to meet. For example, there are two carrier signals and two different modulation signals. Further, one of the modulation signals appears on both carriers while the other appears only on one. A third modulation appears on both carriers but contains very different information than the other two. This section will explain the reasoning behind this design.

2.1 Signal Characteristics

GPS was intended to meet both a military and a civilian demand for global all weather navigation. Satellite radio sources are particularly suited for this purpose since they can radiate their signals to every point on the surface of the planet. It is intended to be a multiple simultaneous access positioning utility, available to as many users as there are receivers. Further, receivers are required to be small enough to carry in one hand. The signals are intended to provide highly accurate measurements of both position and velocity while at the same time being easy to acquire quickly.

2.1.1 Carriers

GPS signals are, like most radio signals, a modulated carrier signal. The carrier signals, being predictable sinusoids, cannot carry information themselves. Information is carried by the modulation signals which are mixed with the carriers.

It is known that the radio wave delay through the atmosphere varies with time and position by as much as 100 meters in equivalent range. Mathematical models do not remove most of this error, so the design includes two separate L band carriers, denoted L1 and L2. Atmospheric delay varies with the square of the carrier frequency. With two carriers, the differential delay of both can be used to measure the absolute delay.

The choice of L band carriers represents a trade-off between the limited available bandwidth in UHF and the excessive atmospheric absorption, called **space loss**, of C band. The L band frequency is high enough to accommodate the 20 MHz bandwidth required and the centimeter wavelength enables precision velocity measurement.

2.1.2 Modulation Signals

The requirement for small receivers translates into a requirement for small antennae⁶³ on the receiver, which in turn means that they must work at very small signal levels. In order to achieve this, both modulation signals are based on a particular kind of digital pseudorandom noise code called a Gold code. These **PRN codes** allow receivers with antenna only a few inches across to extract very low power signals from the noise by correlating them with expectations.

63. Contrast this with the large parabolic reflectors needed to concentrate satellite TV broadcasts. GPS signals are roughly the same strength as satellite TV broadcasts. Even if a GPS receiver used such a dish, it would need four of them to simultaneously track four satellites. The Starfix satellite positioning system uses a 48 inch antenna and is hence not viable in many applications supported by GPS.

solving simultaneous nonlinear equations called Newton-Raphson, and if more than four satellites are used, a least squares technique is used.

Since receivers can measure their clock offsets to great accuracy, some GPS receivers operate solely as very precise clocks or as frequency standards.

1.3 Wave Speed Measurement

The measurement of wave speed, or equivalently atmospheric delay, is accomplished in receivers in one of two different ways. Some receivers use a mathematical model of the atmosphere to compute the delay. The coefficients in this model are broadcast regularly by the satellites themselves.

Since the effect depends in a known way on frequency, some receivers are able to actually measure the difference between each of the two navigation signal carriers transmitted by each satellite. This technique is far more accurate than the use of a mathematical model. Atmospheric delay is one of the most significant sources of error.

1.4 Velocity Measurement

The principle used to measure velocity is the doppler frequency shift. The doppler shift for the frequency of each satellite is a direct measure of the relative velocity of receiver and satellite along the line between them. Each satellite has a very high velocity relative to a stationary receiver because of both the orbital motion of the satellite and the rotation of the receiver with the earth. The velocity solution is obtained by differentiating the four dimensional navigation solution of Equation 39. These are the equations solved in the receiver to determine velocity.

1.5 Orientation Measurement

The principle used to measure orientation of a rigid body is to measure the differential three dimensional positions of different points on the body. In practice, measurements of the radio carrier phase are used to achieve the positioning accuracies required for acceptable orientation measurement. Very accurate real-time orientation measurement for rigid vehicles has been achieved.

The accuracy of angular measurement increases with the length or **baseline** distance between the points positioned. Using four satellites, the user clock error and the 3D position of each point can be calculated. If baselines are known accurately enough, less satellites are required. At least three points must be measured to determine all three angles of rotation.

The time offset of the receiver is called the **user clock bias**. The effect of this bias is shown in Figure 26. The satellites are synchronized, so the user clock bias gives rise to ranges to all of them which are in error by the same amount. These erroneous ranges are called **pseudoranges**.

If the user clock is slow, considering the two dimensional case, the three ranges will create a roughly triangular region as shown. Offsetting the sides of the triangle by fixed amounts on all sides, the receiver can determine the clock bias as an extra unknown.

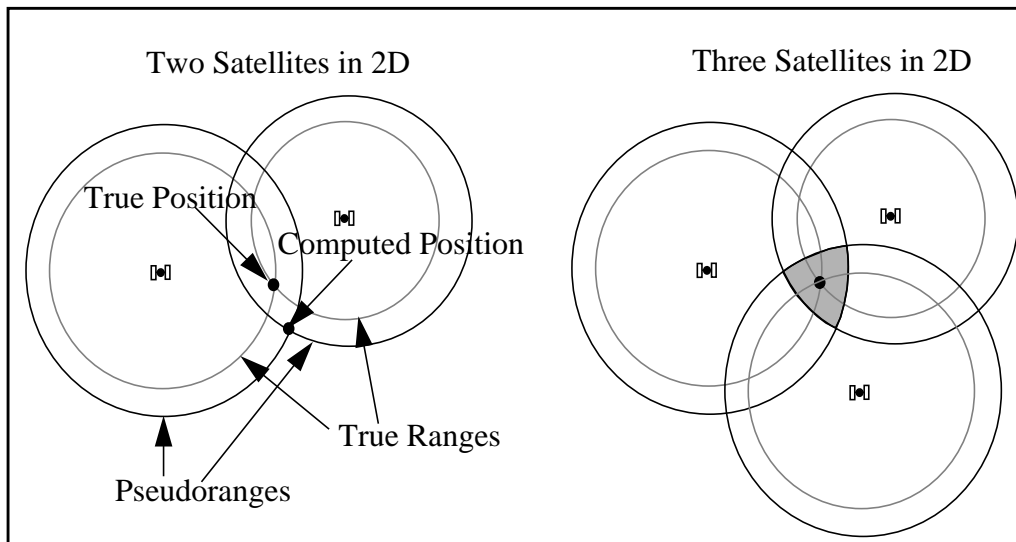


Figure 26 Computing the User Position and Clock Offset

In three dimensions, the receiver position is given by the simultaneous solution(s) (x,y,z) to the following four equations.

$$\begin{aligned}
 r_1 &= \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2} + c\Delta t \\
 r_2 &= \sqrt{(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2} + c\Delta t \\
 r_3 &= \sqrt{(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2} + c\Delta t \\
 r_4 &= \sqrt{(x - x_4)^2 + (y - y_4)^2 + (z - z_4)^2} + c\Delta t
 \end{aligned}
 \tag{Equation 39}$$

Although the equations are written in terms of geocentric cartesian coordinates, the answer can be converted into latitude, longitude, and altitude on any reference ellipsoid model of the earth⁶², or into grid coordinates on a map. The equations are solved in practice by a standard technique for

62. There are many models in use which represent the earth as an ellipsoid. Hence, the latitude, longitude, and altitude of a point depend on which model is in use.

The receiver position is given by the simultaneous solution(s) (x,y) to the following two equations:

$$\begin{aligned} r_1 &= \sqrt{(x - x_1)^2 + (y - y_1)^2} \\ r_2 &= \sqrt{(x - x_2)^2 + (y - y_2)^2} \end{aligned} \quad \text{Equation 38}$$

where the satellite positions (x1,y1), (x2,y2) and ranges r1 and r2 are considered known quantities.

The three dimensional case is similar except that three satellites are used and the lines of position are spheres centered at the satellites. Two spheres intersect in a circle and another sphere intersects this circle in at most two places. The ambiguity can be resolved with another satellite or an initial estimate. For vehicles whose altitude is accurately known, only two satellites are required.

The satellite positions are known as the **ephemeris data**. This information is broadcast to the receivers from orbit. The second required piece of information, the ranges, are determined by measuring the time required for the propagation of the signal from each satellite to the receiver, and multiplying by the radio wave speed⁶¹. To make this possible, each satellite also broadcasts the time along with its position. Then the receiver can subtract the time that the message was sent from the time that it was received.

The receiver performs the very difficult task of measuring a range of about 50,000 kilometers to the nearest meter or so. Therefore, an error of only one part in fifty million in either the propagation time or the speed of light will cause an error of one meter in the observed range. Prediction of the wave speed and tremendously precise measurement of time are both key design features of satellite navigation.

1.2 Time Measurement

In order for a receiver to compute the propagation time directly, it must have its clock synchronized with the clocks used on the satellites, and all of the satellites must be similarly synchronized among themselves. This is because every nanosecond of time error causes one foot of range measurement error. It seems at first glance that all satellites and all receivers everywhere must all agree on the time to the nearest nanosecond.

For the satellites, this synchronization is actually achieved in practice through the use of very precise atomic clocks. However, it is impractical for a receiver to use an atomic time standard. Cheap crystal oscillators are used instead. These clocks are precise enough but not accurate enough, so they may be offset in time from the satellites. The solution to the problem is to consider time a fourth unknown in the ranging process, and to use a fourth satellite to provide an extra constraint equation. If altitude is known accurately (as it is on the sea) one less satellite is needed.

61. Nominally, this is the speed of light. While this is a well known constant in vacuum, the wave speed through the atmosphere varies slightly with time and atmospheric constituents.

1. Principles of Operation

A constellation of satellites is maintained in earth orbit and radio receivers at or near the surface of the planet are used to decode the transmissions of the satellites to compute from them the motion and position of the receiver. GPS was developed primarily as a military navigational aid, so it was designed as a one way broadcast system. Receivers do not have to transmit any signals back to the satellites that might give away their position.

This aspect of the design has a number of other implications. Satellites receive no signals, so there is no satellite capacity limit on the number of users. Receivers transmit no signals, so anyone with knowledge of the communications protocol can use the system without being detected. Civilian use of GPS has been provided for in the design and there is no charge associated with satellite signal reception. The eventual number of civilian users is estimated to be in the millions.

1.1 Position Measurement

Positioning is based on the principle of range triangulation. The receiver needs to know the range to the satellites that are being used and the positions of these satellites. It can determine its own position from only this information. As in all triangulation situations, the number of independent observations required depends on how many dimensions of position are required (2D or 3D) and how much information is already known.

The two dimensional case is indicated in Figure 25. Two satellites are required in general⁵⁹. If the positions of the satellites are known, and the ranges from the receiver to the satellites can be measured, the receiver position is constrained to be one of the two points of intersection of the circular lines of position⁶⁰. After the two simultaneous equations are solved, a rough initial estimate, or a third satellite can be used to resolve which of the two solutions applies.

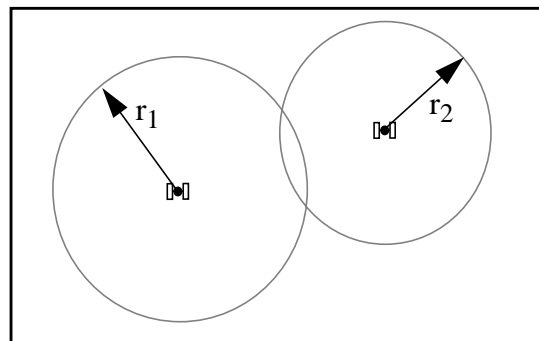


Figure 25 Two Dimensional Range Triangulation

59. The receiver can also wait until one satellite moves to a new position, but see the later section on GDOP.

60. If the receiver is on a line between the satellites, only one point of intersection occurs.

PART IV: Satellite Navigation Systems⁵⁸

Satellite navigation was developed to answer the military need to precisely determine the position of air, sea, and land vehicles. It is rapidly becoming a definitive navigation tool since it provides continuous, high accuracy positioning anywhere on the surface of the planet and the near space region, 24 hours a day, under all weather conditions. The receiver systems are small, lightweight, and easy to use, and have made handheld global positioning systems a reality.

This technology is important for mobile robots for the same reasons it is important to the military. For a mobile field robot operating over large distances, or near the poles, it is one of only a few available good solutions to the position estimation problem.

The military has used the TRANSIT constellation of satellites since 1964 and the resource industries have used the Starfix system since 1986 for such applications as locating drilling rigs, seismic surveys, and pipeline construction. Both of these systems are now obsolete by two more recent systems, the Global Positioning System developed by the United States Department of Defense, and its Soviet counterpart, called Glonass. The next generation of INMARSAT satellites is also planned to radiate GPS like signals for positioning.

All of these satellite navigation systems are based on identical principles, notably range triangulation, and GPS and Glonass are virtually identical from the perspective in this report. For this reason, only GPS will be discussed here but the reader is cautioned that virtually all of the information in this chapter applies equally well to Glonass.

Since GPS provides a positioning fix, it complements inertial navigation or other dead reckoning techniques. Integration of both techniques achieves performance which is an improvement over what would be possible using either system alone.

There are many navigation signals available from the satellites and many different ways of processing them. These processing techniques give rise to trade-offs between the accuracy of the fix, how quickly it can be updated, and the range over which the accuracy can be maintained. Accuracies vary from tenths of a kilometer to tenths of a centimeter depending on the signals used and how they are processed. The systems can be used to measure position, velocity, attitude, and very precise time.

58. This chapter has benefitted from the kind assistance of the following individuals:
Eric Anderson, Ashtech Inc, Sunnyvale, Ca.
Kevin Dowling, Research Scientist, Field Robotics Center, Pittsburgh, Pa.
Mike Leary, Trimble Navigation, Sunnyvale, Ca.
Mike Slaven, Trimble Navigation, Sunnyvale, Ca.

worldwide operation below 75 degrees latitude. Warm-up may require elevated power consumption up to 500 watts.

Newer systems integrated with GPS can function in a **moving base alignment** mode where successive GPS readings are used to align the real or computational platform as the vehicle moves.

7.8 Accuracy

The accuracy of the INS can be specified in several ways, depending upon the application it is suited for. For cruise vehicles, the rate at which the position error builds up over time is specified. For ballistic and space vehicles, the position error at the target or the velocity error at propulsion cut-off is given. Land vehicle systems often specify accuracy as a percentage of distance travelled.

Today's commercial cruise systems can achieve accuracies on the order of 0.2 nautical miles of error per hour of operation. Pitch and roll are often accurate to 0.05° and true geographic heading to 0.5° . In some cases, position accuracy along the trajectory (**alongtrack**) and both normal directions (**crosstrack** and vertical) are distinguished.

Land vehicle navigation systems achieve 0.2% to 2% of distance travelled. Pitch and roll can be measured to 0.1° and heading to 0.5° . While the inertial components typically perform better than this by an order of magnitude, systems require shock and vibration isolation mounting. These isolation systems necessarily permit relative motion between the system housing and the vehicle chassis which significantly reduces the accuracy of the orientation output.

At present, strapdown systems do not compete with gimbaled ones in terms of accuracy, but this is likely to change as component technologies improve.

- initial azimuth misalignment of platform - gyrocompassing error
- dynamic misalignment of platform - servo droop in platform controller
- accelerometer misalignment - with respect to platform axes
- gyro misalignment - with respect to platform axes

7.3.4 Computational Stabilization (equivalent to gyro drift)

- truncation - finite precision computations
- roundoff - finite precision computations
- commutation - small angle assumptions in forming the direction cosine matrix
- pickoff - finite resolution of sensor interfaces

7.3.5 External Errors

- errors in geodetic information
- errors in damping signals

7.4 Configuration Of Inertial Systems

In addition to the complete inertial navigation system supplying direct navigation information, some vendors supply strapdown IMUs as stand alone units. These units output body coordinate referenced velocity and orientation information which must be converted by the user into a more useful form. Specifications for these systems do not differ significantly from those applying to the more complete systems.

Many systems admit velocity and/or odometry damping inputs which are processed internally by the INS in order to Kalman filter them with the computed motion. Hybrid INS/GPS systems are also available.

7.5 Physical

Systems can weigh as much as 50 lbs, occupy as much as 1 cu ft. of volume, and consume up to 100 watts of power. Strapdown systems and the use of ring laser gyros are becoming popular. The lack of gimbal structure in these systems affords reduction in size, weight, power consumption, and cost.

Strapdown systems often use miniature gyro and accelerometer components which weigh a few grams and occupy only a fraction of a cubic inch of volume. The bulk of the volume of the strapdown system is taken up by the processing electronics.

7.6 Data Rates

The rate at which systems can supply positioning data can vary significantly. Up to several hundred Hz update rates are possible, depending on the processing electronics and the data interface.

7.7 Initialization

The time a system takes from power-up to completion of alignment can vary from a few minutes to 30 minutes depending on ambient temperature and local latitude. Vendors often quote

the external measurements are combined to provide the best overall estimate of the navigation quantities.

The damping of the vertical channel can be accomplished in many ways. Measurements of position, velocity, and attitude can be used singly or in any combination. The particular solution adopted is application dependent.

For flying systems, altitude is typically available from other sources. Barometers can be used for only coarse measurement of altitude since air pressure varies with weather conditions as well as altitude. Altitude can be determined to a few hundred feet in this way. Radar altimeters are often accurate to a few percent of altitude. However, many flying systems are less concerned with the accuracy of altitude than with horizontal position. Doppler radar velocity inputs are also commonly used for airborne systems.

Seaborne systems can take advantage of the extensive radio navigation systems that have been available for many decades. Space systems can use star sights, and when in earth orbit, horizon measurement and navigation satellites. The latter are also particularly useful for all terrestrial systems that have clear view of the satellites.

Land navigation systems can use satellite systems as well as landmarks and map matching to fix position. Periodic return to the same position allows position errors to be zeroed by telling the system that it has returned to the base position. Similarly, the technique of **zero velocity update** notifies the system when it has stopped moving. Odometry damping is also a common technique in land navigation systems.

7.3 Sources of Error

In practice, the most significant sources of errors in an INS are limited to nonidealities in the accelerometers and gyros, and several kinds of alignment errors. Together, these often amount to 90% of the total system error. A more complete list of error sources is as follows:

7.3.1 Accelerometer Errors

- Bias - the output of the sensor for zero input
- Scale Factor - deviation of the slope of the device calibration curve from ideal
- Cross Coupling - rebalance servo droop
- Vibropendulosity - steady state output for vibratory input

7.3.2 Gyro Errors

- bias - parasitic torques caused by restraints and magnetic and thermal effects
- proportional - torques caused by mass unbalance and anisoelectricity
- anisoelectric - deformations normal to applied loads
- torquing error - errors in the torques supplied to rotate the platform
- random drift - uncompensated nondeterministic errors

7.3.3 Platform Alignment Errors

- initial level misalignment of platform - levelling error

7 Performance

The relevant performance parameters of an INS include accuracy and data rate, initialization time, frequency of required external updates, and the latitude band in which it can be used. These factors and others are discussed below from the point of view of providing a general appreciation of the spectrum of systems available.

7.1 Inertial Grade Componentry

The realization of inertial navigation had to await the historical development of instrumentation technologies which could supply the required performance parameters. An investigation of the typical requirements placed on gyros and accelerometers provide an insight into the severe difficulty of the problems that have been solved - and the remarkable performance of today's systems.

Positioning accuracy in dead reckoning systems is particularly sensitive to angular drift error⁵⁷. For a locally level gimbaled system, consider that one minute of arc of angular platform error corresponds to one nautical mile of position error on the earth's surface. In order to achieve the modest accuracy requirement of one mile of position error per hour of operation, gyroscope drift must be held within one arc minute per hour. This corresponds to one thousandth the rotation rate of the earth or one milli earth rate unit, known as a **meru**. These specifications are three orders of magnitude more stringent than those imposed on gyros used in conventional aircraft instrumentation.

For accelerometers, since bias is multiplied by time squared in deduced reckoning, requirements for one thousandth of a g of sensitivity are common. Yet at the same time, the device must also be capable of measuring the earth's gravity. Hence wide dynamic range and extreme sensitivity are required for practical systems.

7.2 Aided Inertial Systems

While an inertial navigation system is conceptually a self-contained device, it is typically the case, for reasons discussed above, that serious degradation of performance arises unless redundant external measurements are used to remove the errors that build up over time in a manner typical of all deduced reckoning systems.

The use of external measurements in this manner is referred to as **damping**, and the systems employing the technique are called **aided** or **augmented** inertial systems. In this view of an INS, its real utility is in its ability to provide continuous positioning information between external updates.

Classically, INS outputs were adjusted to agree perfectly with external measurements at the point in time when the external measurement was made. The modern technique for incorporating external measurements is optimal estimation, often incorporating the Kalman filter. Using this technique, estimates of the errors in both the INS generated navigation quantities and the errors in

57. A good illustration of this is that a drift rate of a mere 1/10 of a degree per hour will cause a transatlantic flight to end up in Paris rather than London.

6.5.4 Levelling

During levelling, for a gimballed system, the platform is controlled by a special levelling servo which drives the platform to rotate into the gravity vector until both horizontal accelerometers cease to measure a component of gravity. At this point, the vertical accelerometer is aligned with the local vertical and the platform is level.

6.5.5 Gyrocompassing

Gyrocompassing is accomplished by rotating the now level platform about the vertical until one gyro senses no component of the earth's rotation⁵⁶. This rate goes to zero at the poles, so such systems have difficulty during initialization near polar latitudes. For systems operating in the air, alignment can be performed in-flight using an externally supplied velocity signal. Space systems can use star trackers and, if in orbit, scanning of the earth's horizon to align themselves. Sighting of known stars can also be used.

Strapdown systems can compute the vehicle pitch and roll from the appropriate accelerometer outputs given knowledge of the initial position on the earth since this fixes the direction of the gravity vector with respect to the earth. While the vehicle is stationary, the accelerometer outputs measure only the components of the gravity vector and its resolution onto the body axes constrains the vehicle orientation.

For strapdown systems, since the system is stationary, the angular rate vector measured by the gyros represents the resolution of the earth rate onto the vehicle axes, so the vehicle heading can be computed.

The gyrocompassing step is complicated by the low signal to noise ratios encountered when measuring the relatively small earth rate. Systematic gyro drift must also be eliminated for the technique to be practical. A further reduction in signal to noise results from the variation of the horizontal component of the earth spin rate by the cosine of latitude. This is the second aspect of the pole problem for inertial systems.

56. The earth's magnetic field is not a reliable indicator of north for such a process, so the obvious alternative is not used.

6.5 Initialization

When an INS is started, a sequence of operations is performed which takes at least several minutes of time before the system is operational.

6.5.1 Start-up and Warm-up

Mechanical setup involves spinning up the gyros and waiting for the components to reach operating temperature.

6.5.2 Trim

Drift trimming is the process of calibrating out the constant components of gyro drift error by trimming the biases of the torquers. When the system is known to be stationary, rotational motion of the platform is known based on the earth's rate. Any rates measured above that expected because of the earth's rotation can be removed in this way.

6.5.3 Alignment

The determination of the orientation of the accelerometers with respect to the earth requires the knowledge of their orientation with respect to earth fixed vectors. If a single vector were used, platform orientation would not be constrained in rotation about that vector. Hence, two noncolinear vectors are required. These vectors may be obtained from star sights, surveyed landmarks, the gravity and magnetic fields, and the earth's spin axis. In general, systems which report relative position still require precise alignment in order to avoid incorrect compensation for the apparent forces due to the earth's spin.

In applications where the start point is always the same, precision surveyed landmarks provide an excellent alternative. Systems that are required to align themselves while launched from another flying vehicle or on orbit can make use of star sights and the earth's horizon as landmarks, or can transfer alignment from the host vehicle. The variation of the earth's magnetic field makes it an unreliable heading reference unless it is known to a few seconds of arc at the start point. For terrestrial mobile robots, self alignment using only the sensors of the inertial system itself is a convenient method of alignment.

In **self alignment**, the local vertical is determined through the gravity vector and the earth's spin provides a reference for north. For an earth stabilized system, the process occurs in two steps called **levelling** and north alignment or **gyrocompassing**⁵⁵.

55. Although the term is used commonly, it does not mean that the gyro operates according to the principle of the gyrocompass. It means that it does the same thing as the gyrocompass, namely, finds north.

6.4.2 Shape of the Earth

The earth is not perfectly spherical⁵³. Its polar radius is about 3432 nm and its equatorial radius is about 3444 nm. So it is flattened by about 12 miles. In mathematical terms, its shape can be approximated by an oblate spheroid. This has many implications to earth fixed systems:

- the length of the latitude is not precisely constant over the globe
- the radius of the earth is not constant over the globe
- the theoretical gravity field is slightly different than for a spherical earth

There have been many models of the earth in use throughout history, and for historical reasons, different countries continue to use slightly different models. The definition of latitude varies slightly depending on which **reference ellipsoid** is used by the country in which the system operates. Some 20 slightly different ellipsoids have been suggested as various times to represent the shape of the earth. They differ by about 1 kilometer in earth radius at the equator or 0.01%. The DOD regularly charges the World Geodetic System Committee to establish a standard earth model for mapping purposes. WGS-84 and WGS-72 are commonly used standards.

6.4.3 UTM Coordinates

A coordinate system called Universal Transverse Mercator or UTM coordinates has been in common use by the army for some time. This system is commonly used to report position in land navigation inertial systems because most were originally developed for army use. The utility of UTM coordinates is that it allows the system output to be correlated simply with position on a flat map.

The nearly spherical earth is, in mathematical terms, a nondevelopable surface. This implies that any means of flattening it out onto a plane must involve some distortion. While it is not commonly appreciated, the Mercator projection which is used to generate a flat map of the world gives rise to massive distortions in the shape of the earth's surface⁵⁴. The Mercator projection is constructed by projecting the earth's surface onto a cylinder which is tangent to the earth at the equator.

The transverse Mercator projection uses a cylinder which is tangent along a meridian. UTM coordinates in the north and east directions are referred to as **northing** and **easting**, and are often reported in meters.

53. It was Isaac Newton who first showed that the earth must bulge at the equator due to the centrifugal force of its rotation.

54. A remarkable example is that Greenland appears to be larger than South America on an everyday map of the world. In fact, Greenland is only 1/10 the size of South America. The distortions of the Mercator projection are largest near the poles, so Greenland is particularly illustrative of this.

6.4 Position Coordinates

While the stabilized coordinate system is the system in which the vector equations of the first section are represented and manipulated internally by the computer, the system output is often converted to other coordinates for reasons of convenience. Two commonly used systems are **latitude** and **longitude** and the military **grid** or UTM system.

6.4.1 Latitude and Longitude

The generic INS mechanization equations of the previous section provided the earth referenced position vector of the vehicle as output. For local vertical systems, it is more common, in practice, to compute latitude and longitude by integrating the velocity, so that the position vector is never explicitly computed. For example, for local vertical systems, this is particularly easy to do.

Let Λ be the vehicle longitude and λ its latitude. The latitude, longitude, and altitude of the vehicle relate to the vehicle velocity by:

$$\frac{d\lambda}{dt} = v_{\text{north}} \cdot R$$

$$\frac{d\Lambda}{dt} = v_{\text{east}} \cdot R \cdot \cos\lambda$$

$$\frac{dR}{dt} = -v_{\text{down}}$$

Equation 37

Latitude and longitude are angular measurements, in fact, spherical polar coordinates. Several points about these coordinates must be remembered⁵². Lines of constant latitude, called **parallels**, are normal to the direction in which latitude changes. Lines of constant longitude, called **meridians**, are normal to the direction in which longitude changes. The distance between parallels is, ignoring the slight ellipticity of the earth, constant everywhere on the globe. The distance between meridians varies from a maximum at the equator to zero at the poles.

52. As a convenient mnemonic, one nautical mile is slightly larger than the statute mile and is defined as the arc length subtended at sea level of an angle of one arc minute measured from the earth's center. This makes many navigation calculations simpler.

6.3.3 Tangent Plane, Locally Level System

This system is defined in the following figure. In this case, while the platform moves with the vehicle, its orientation with respect to the earth remains fixed at its initial value. It takes advantage of the fact that constant platform rates can be generated much more accurately than variable ones. These rates can be precisely determined at start-up. Such systems are usually employed when computation of latitude and longitude is not required and when the mission excursion is limited. For excursions beyond a few miles from the base point, compensation for gravitation must be computed for the horizontal axes.

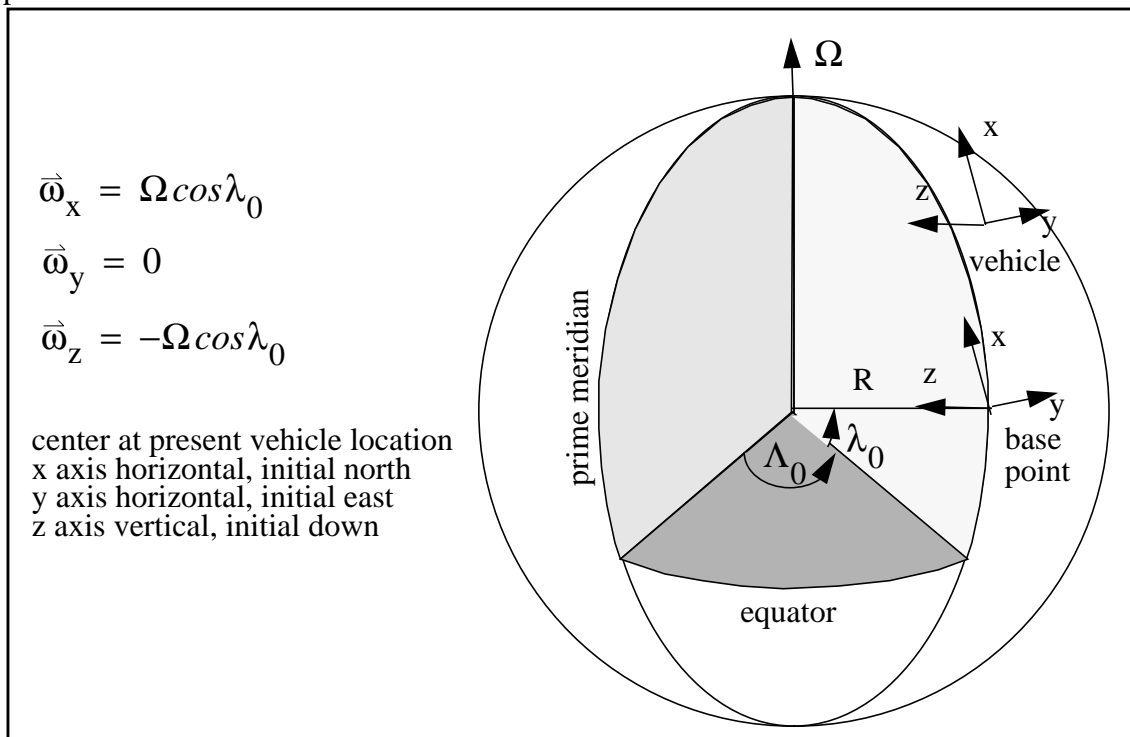


Figure 24 Tangent Plane, Locally Level System

6.3.2 Free Azimuth, Locally Level System

This system is defined in the following figure. It solves the high polar platform rate problem by providing no inertial platform rate about the vertical axis. The torquing error of the z gyro is ordinarily worse than the two other axes, so it is advantageous to remove it in this way. The azimuth is allowed to wander with respect to north and computations keep track of the true direction of north. In a sense, such systems are hybrid gimballed - strapdown systems. The angle between the x axis and true north is called the **wander angle**. **Wander azimuth** is a variant where the z component of the inertial platform rate is equal to the z component of the inertial earth rate.

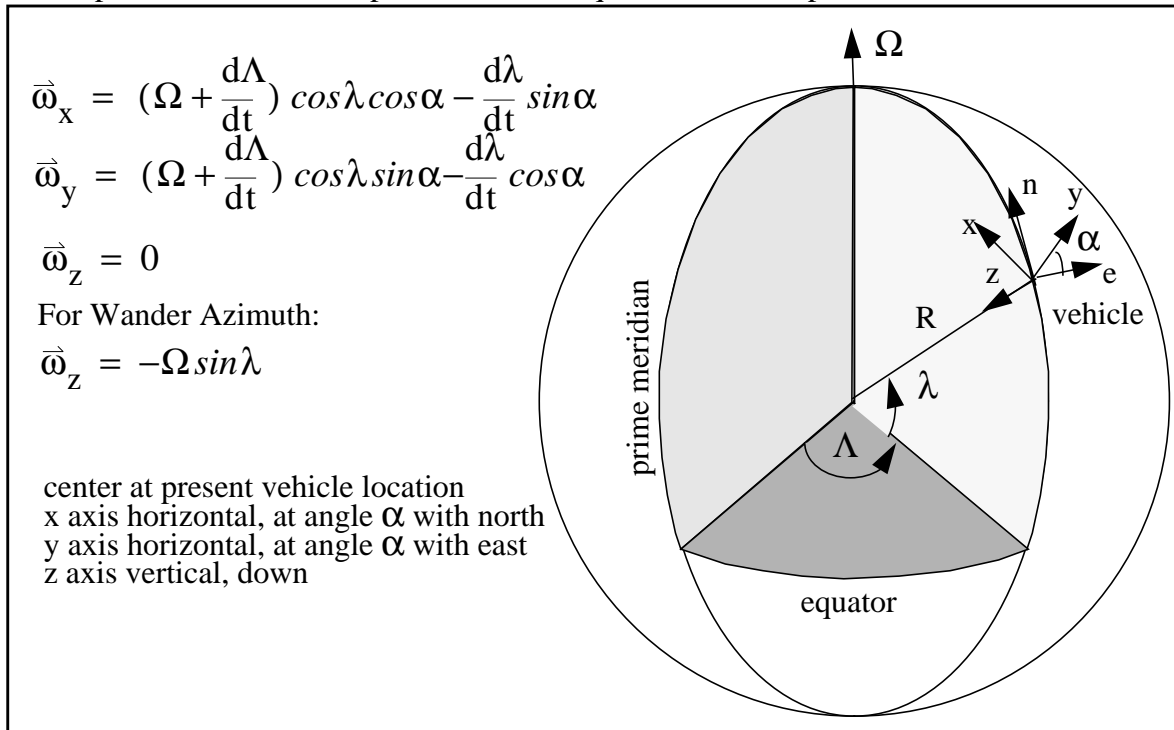


Figure 23 Free Azimuth, Locally Level System

6.3 Common Mechanized Coordinate Systems

The three most common terrestrial INS mechanizations differ only in the way that the platform⁵¹ rotates with respect to the earth. For strapdown systems, the same options are available and the direction cosine computations are driven by appropriate rates to implement a pseudo stable platform. In the notation of earlier sections, they differ in the prescription of the vector $\vec{\omega}$. These systems will be summarized below.

6.3.1 North Slaved, Locally Level System

This system is defined in the following figure. It has the advantage shared by all locally level systems that compensation of the horizontal channels for the gravity vector is not necessary, since it always points along the z axis. Like all earth stabilized systems, gyro torquers are required.

It permits easy calculation of latitude and longitude and is commonly used in long term cruise systems. It indicates the vertical so it can be used easily to assist military weapons systems and it can be easily connected with many types of navigation aids which can be used to reduce errors in the inertial system without transforming coordinates.

One serious drawback of this system is that a singularity exists at the poles, because the x axis must rotate very rapidly around the azimuth near either of the poles. In gimballed systems, this manifests itself as high platform azimuth rates. In strapdown systems, the high rotation rates limit the accuracy of the computations. This is one of two aspects to the **pole problem** for the INS. The second problem is discussed later.

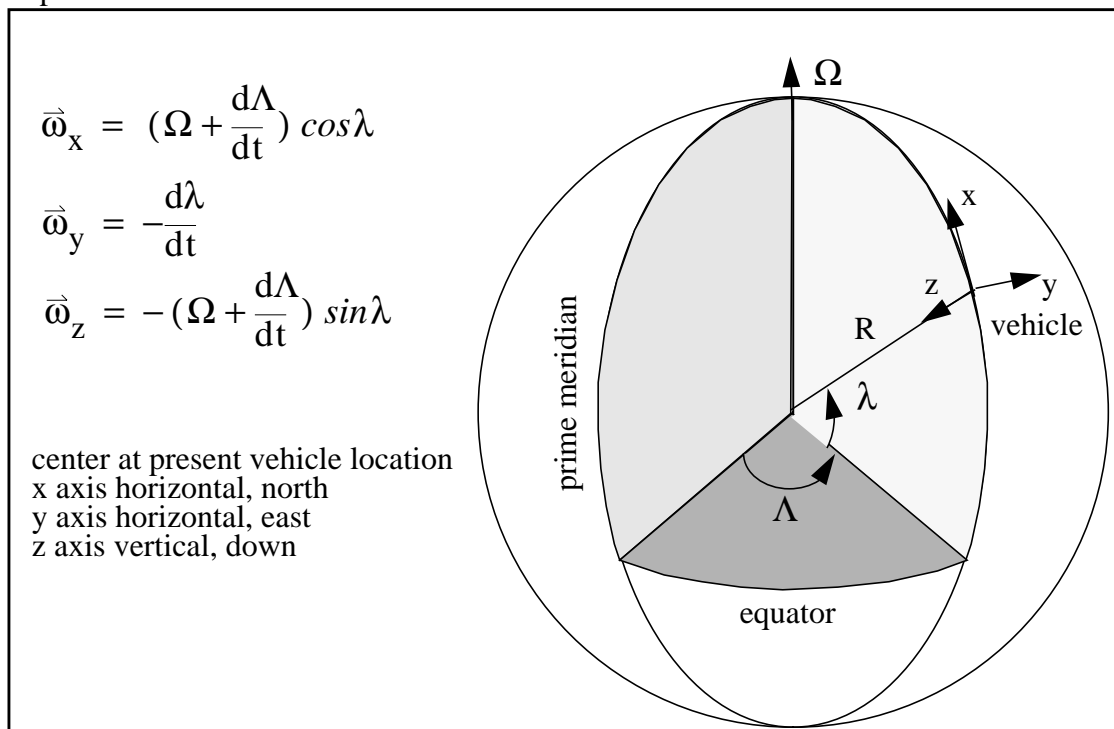


Figure 22 North Slaved, Locally Level System

51. The platform frame in gimballed configurations is equivalent to the vehicle frame in strapdown systems.

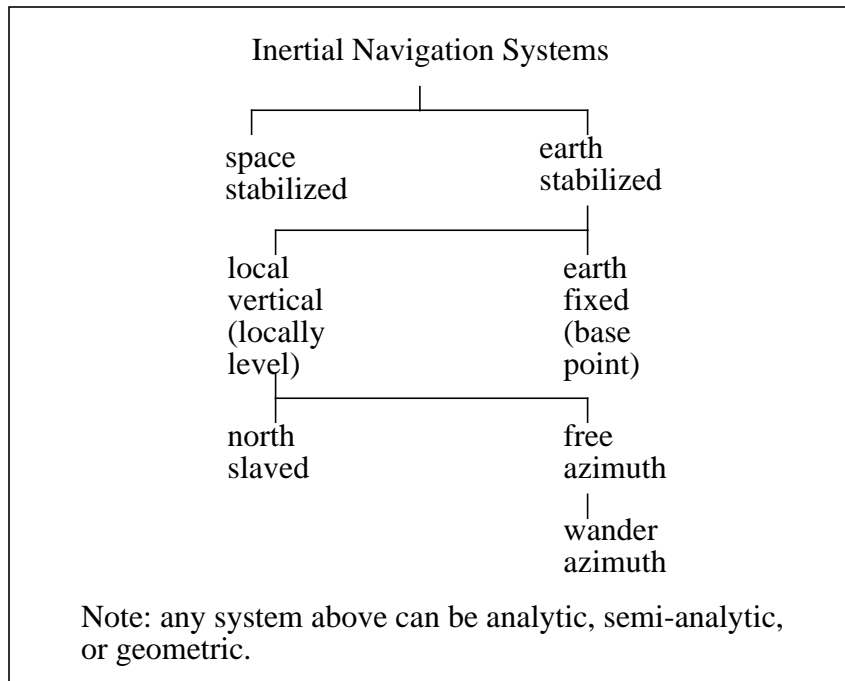


Figure 21 Taxonomy of Inertial Systems

6.2 Navigation Coordinate Systems

Perhaps the most important distinguishing characteristic of inertial navigation systems is the coordinate system in which the navigation equations are mechanized. The two categories here are space stabilized, and earth stabilized.

6.2.1 Space Stabilized Systems

Space stabilized systems employ an inertially fixed coordinate system. As a result, the navigation equations take a particularly simple form. Such systems are very inconvenient for use in terrestrial applications. Space applications often use hybrid stellar-inertial navigation systems employing **star-trackers** since star sighting is both easy and a very accurate method of determining orientation.

6.2.2 Earth Stabilized Systems

Earth stabilized systems maintain orientation referenced in various ways to the earth. **Earth fixed** systems employ a coordinate system fixed with respect to the rotating earth. These are also known as **base point** or launch point systems since the origin of coordinates does not move with the vehicle but rather remains fixed at the point of initialization.

Local vertical systems are the most common earth stabilized systems and employ coordinates referenced to the local vertical direction with the other two axes horizontal. In this case, because the coordinates rotate and move with the vehicle they are called **locally level**.

Both kinds of earth stabilized systems require that an earth's rate signal be generated by the system to drive the platform or direction cosine transform to rotate with the earth and can ignore the gravity vector provided the platform is sufficiently stable. For local vertical systems, the compensation signal must also account for vehicle motion, for, depending on latitude, each nautical mile of travel corresponds to up to one minute of arc of change in the local vertical.

Among local vertical systems, there are three common variants. **North-slaved** systems maintain one coordinate axis pointed north at all times. Axes are typically called north, east, and azimuth⁵⁰. **Free azimuth** and **wander azimuth** systems allow the level coordinate axes to rotate with respect to north about the vertical axis as a function of latitude and computations are employed to keep track of the angle of rotation.

A summary of this taxonomy of inertial systems is provided in Figure 21:

50. Azimuth refers to the angle made by an object with respect to an arbitrary reference. The angle is measured in the local earth tangent plane. Elevation is the angle an object makes with the horizon.

6.1.2 Semi-Analytic Systems

The second class of systems are called **semi-analytic**. These are gimballed systems which control the orientation of a platform in order to instrument the navigation frame only. Computations transform the sensor outputs from the inertial frame and those systems which do not actively maintain a north reference must also transform coordinates from the platform axes to local north and east directions. Computations are also used to determine the latitude and longitude of the vehicle.

Gimballed systems employ gyro torquers which drive the platform to rotate as required. Untorqued gyros naturally implement an inertial reference. The total torquer signal may include compensation for the gyro drift rate, the earth's sidereal rate and angular velocity of the vehicle with respect to the earth, called the **vehicle rate**.

6.1.3 Strapdown Systems

The third class, known as **strapdown** or **analytic** systems, are said to be computationally stabilized. Strapdown systems get their name from the fact that they strap the sensors directly to the vehicle chassis, eliminating the platform gimbals, at the cost of causing the components to suffer the dynamic rotation of the vehicle. Strapdown mechanization places more stringent requirements on the sensory components used, and on the computational throughput, so these systems have appeared relatively recently in the history of development of inertial navigation. Such systems promise to replace semi-analytic systems in most applications.

These systems are analysed mathematically in an identical manner to gimballed systems with the exception that the accelerometer readouts are first converted to a common coordinate system using a matrix transformation, often called the **direction cosine matrix**. There are many advantages to the strapdown approach, other than the obvious elimination of the mass, volume, and mechanical complexity of a gimballed platform.

These systems use rate gyros which can be as small as one cubic inch in volume and admit precessional rates up to 300 times that of gimballed systems. High precessional rates are necessary to faithfully track the motion of the vehicle to which they are attached. Torquers are required to be much more powerful. Accelerometers can be as small as 5 grams mass and 0.05 cu inch in volume. The operating principle is to use very high speed **rebalance loops** which continually torque the gyros back to their null vehicle aligned positions. The effort required to do this is proportional to the angular rate of the vehicle. The rebalance loops report current level of effort to the computational stabilization algorithm which maintains the current vehicle orientation in inertial space.

The rest of the strapdown system is identical to a gimballed system. Strapdown systems are aligned and trimmed in a manner analogous to gimballed systems. Strapdown systems are becoming particularly suited to the relatively low accuracy requirements of land navigating mobile robots.

6 Implementation

This section will briefly discuss some of the important aspects of complete inertial navigation systems. Today's commercial systems can achieve accuracies on the order of 0.2 nautical miles of error per hour of operation.

All inertial navigation systems use gyroscopes to measure vehicle rotation in inertial space, and accelerometers to measure inertial specific force. Two or three accelerometers are used to measure specific force along as many axes and three single degree of freedom or a pair of two degree of freedom gyroscopes complete the sensor package. This package is often called the **inertial measurement unit (IMU)** or **inertial reference unit (IRU)**.

Computers perform a large number of computations in order to convert the IMU outputs into the required earth-relative motion of the vehicle. In modern systems, digital computers are used and the integrations are performed using digital quadrature algorithms. In many cases, integrating accelerometers and/or integrating gyroscopes are used which perform one of the integrations through the addition of a damping term in their transfer functions.

Although simple computational models of the gravitation of the oblate spheroidal earth are available, many systems do not employ them at present. Modern systems employ sophisticated high speed integration algorithms and real-time Kalman filtering to improve performance.

Real INS's are typically designed to report position in a particular frame of reference and coordinate system. For terrestrial applications, an earth fixed frame is called for, while for extra-terrestrial applications, a space reference is more useful. Devices such as missiles, even though they operate substantially in space, must be aware that their target will have moved in inertial space since the time they were launched. All devices which navigate close to the earth experience errors reflecting the famous Schuler period, and vertical error increases without bound over time unless stabilized by periodic external updates.

6.1 Gimballed and Strapdown Mechanizations

The implementation of a particular navigation coordinate system involves either actively controlling or passively keeping track of the orientation of the accelerometers as the vehicle moves. A trade-off exists between mechanical complexity and computational complexity which, over time, has tended to drive system designs toward computational complexity as computer systems have improved in performance. Three classes of systems are distinguished based upon their relative position on this spectrum.

6.1.1 Gimballed Systems

The first class, known as **gimballed** or **geometric** systems, employ a stabilized platform which is actively servoed to the required orientation. This type of system was the first practical class to be developed historically since navigational information was available directly from the gimbal angles. Up to five gimbals are used to instrument both an inertially nonrotating reference frame and a local navigation frame. These systems require only minimal computational capacity. In practice, they have been replaced by another class of systems, called semi-analytic.

Let g_0 and R_0 be the gravitational acceleration of and radius to the local region, which have been assumed to be approximately constant. If the accelerometer errors are assumed to be constant biases, the solutions to these equations for zero initial conditions are:

$$\begin{aligned}\Delta x &= \frac{\Delta a_x}{g_0/R_0} \left[1 - \cos \left(\sqrt{\frac{g_0}{R_0}} t \right) \right] \\ \Delta y &= \frac{\Delta a_y}{g_0/R_0} \left[1 - \cos \left(\sqrt{\frac{g_0}{R_0}} t \right) \right] \\ \Delta z &= \frac{\Delta a_z}{2g_0/R_0} \left[\cosh \left(\sqrt{\frac{2g_0}{R_0}} t \right) \right]\end{aligned}\quad \text{Equation 36}$$

5.4 Error Dynamics of the Free Inertial System

Hence, the horizontal position channels oscillate at the Schuler frequency with an amplitude proportional to the accelerometer bias. The analysis is mechanization independent, so this oscillation arises in all systems which operate near the earth. The main conclusion is that while it is unfortunate that gravity needs to be modelled to use an INS, the very existence of a gravitational field inherently stabilizes the process of dead reckoning from an accelerometer. Although the analysis has assumed only accelerometer errors, many other sensor nonidealities cause similar behavior.

The vertical channel is divergent with a scale factor proportional to the accelerometer bias and a time constant related to the Schuler period. Many local vertical systems possess no vertical computation channel at all for this reason. If they do, the channel must be augmented by other measurements to stabilize it. The next section discusses the available methods of stabilization.

Notice that for periods of time which are small compared with the 84 minute Schuler period, the Taylor approximations for the cosine and hyperbolic cosine in Equation 36 both indicate a parabolic growth of the accelerometer bias with time as would be expected without gravity damping. This implies that while the error may be bounded in the horizontal case, the bound may be a large number depending on the size of the sensor error. Hence, gravitational feedback can be viewed as having the effect of converting three parabolic error profiles into two oscillatory ones and one exponential one.

These results are independent of mechanization since no particular mechanization was assumed. The fundamental message of the equation is that errors in accelerometer indication give rise to errors in computed position. These errors in position give rise to errors in computed gravitation. The errors in gravitation tend to be opposite in direction to the errors in acceleration which caused them. The errors may arise from physical deflections or considerations of a more computational nature. That is, errors oscillate in Schuler tuned systems for exactly the same reason that a pendulum oscillates in a gravitational field - the mathematics are identical.

5.2 Linearized Gravity Model

The gravitational force is a function only of position. A Taylor series expansion assuming a spherical homogeneous earth allows rewriting the gravitation error in terms of the error in the position as follows:

$$\begin{aligned}\vec{g} &= -\frac{GM}{r^3}\vec{r} = -\frac{GM}{(\vec{r} \cdot \vec{r})^{3/2}}\vec{r} && \text{Equation 32} \\ \Delta\vec{g} &= \left[\frac{\partial\vec{g}}{\partial\vec{r}} \right] \Delta\vec{r} = -\frac{GM}{r^3}\Delta\vec{r} + 3\frac{GM}{r^5}(\vec{r} \cdot \Delta\vec{r})\vec{r} && \text{by the product rule of differentiation}\end{aligned}$$

Substituting this into Equation 31 yields:

$$\frac{d^2}{dt^2}\Delta\vec{r} = \Delta\vec{a} - \frac{GM}{r^3}\Delta\vec{r} + 3\frac{GM}{r^5}(\vec{r} \cdot \Delta\vec{r})\vec{r} \quad \text{Equation 33}$$

5.3 Solution Near the Earth's Surface

Further analysis of this equation requires that a coordinate system be adopted. Let the origin be placed at the center of the earth, and the three cartesian axes be oriented arbitrarily. Writing this out in component form:

$$\begin{aligned}\Delta\ddot{x} + \frac{GM}{r^3}\Delta x - 3\frac{GM}{r^5}(x\Delta x + y\Delta y + z\Delta z)x &= \Delta a_x \\ \Delta\ddot{y} + \frac{GM}{r^3}\Delta y - 3\frac{GM}{r^5}(x\Delta x + y\Delta y + z\Delta z)y &= \Delta a_y \\ \Delta\ddot{z} + \frac{GM}{r^3}\Delta z - 3\frac{GM}{r^5}(x\Delta x + y\Delta y + z\Delta z)z &= \Delta a_z\end{aligned} \quad \text{Equation 34}$$

Any particular solution to these equations requires knowledge of the trajectory followed by the vehicle. Let the start point for the system be along the z axis on the surface of the earth, and let the vehicle trajectory remain close to this point so that $x = y = 0$ and $z = r = R$. Under this assumption, the cross coupling terms in the equations cancel and they reduce to:

$$\begin{aligned}\Delta\ddot{x} + (GM/R^3)\Delta x &= \Delta a_x \\ \Delta\ddot{y} + (GM/R^3)\Delta y &= \Delta a_y \\ \Delta\ddot{z} - (2GM/R^3)\Delta z &= \Delta a_z\end{aligned} \quad \text{Equation 35}$$

5 Inertial Error Dynamics

A more refined analysis of the phenomenon of gravity feedback leads to the conclusion that *most* errors in an INS oscillate with the Schuler period. Such an analysis would consider the particular mechanization of the system - since many errors are most easily described in the mechanized coordinate system, and would consider the effects of all significant errors simultaneously. Unfortunately, such an analysis consumes more space than is available here. However, it is possible to illuminate the basic ideas by considering a single error source - accelerometer bias, and developing the equations describing its propagation throughout the system in a mechanization independent manner.

5.1 Perturbative Analysis

The performance of the Schuler tuned system in the presence of sensor nonidealities can be investigated by a technique known as perturbation analysis. To do this, consider the basic navigation equation - inertial acceleration expressed in terms of the specific force indicated by the accelerometers \vec{a} , and **gravitation** \vec{g} :

$$\vec{a}_{\text{inertial}} = \frac{d^2\vec{r}}{dt^2} = \vec{a} + \vec{g} \quad \text{Equation 29}$$

In perturbation analysis, a perturbative error is applied to the sensed specific force and the effect of this on the system output is investigated. Let the indicated specific force include an error denoted $\Delta\vec{a}$, and cause errors in the computed position and gravitation denoted by $\Delta\vec{r}$ and $\Delta\vec{g}$. This is accomplished through the substitutions:

$$\begin{aligned} \vec{a}_i &= \vec{a}_t + \Delta\vec{a} \\ \vec{r}_i &= \vec{r}_t + \Delta\vec{r} \\ \vec{g}_i &= \vec{g}_t + \Delta\vec{g} \end{aligned} \quad \text{Equation 30}$$

Where the subscripts i and t represent indicated and true quantities. Substituting this back into the original equation and cancelling out the original equation yields.

$$\frac{d^2}{dt^2}\Delta\vec{r} = \Delta\vec{a} + \Delta\vec{g} \quad \text{Equation 31}$$

This is the differential equation which describes the propagation of errors from the accelerometer to the position and gravity computations.

4.4 Foucault Effect

The Schuler period can be viewed as the oscillation of a pendulum whose length is the radius of the earth. It is the period of the pendulum itself. This conceptual earth-radius pendulum is also subject to an oscillation of the plane of its rotation just as is the Foucault pendulum. This period is 24 hours at the equator and increases to infinity at the poles. For a period T , earth rotation rate of Ω and latitude λ , the equation is:

$$T = \frac{2\pi}{\Omega \sin \lambda}$$

There is an inherent minus sign in the process as illustrated below using a simple spring model of an accelerometer:

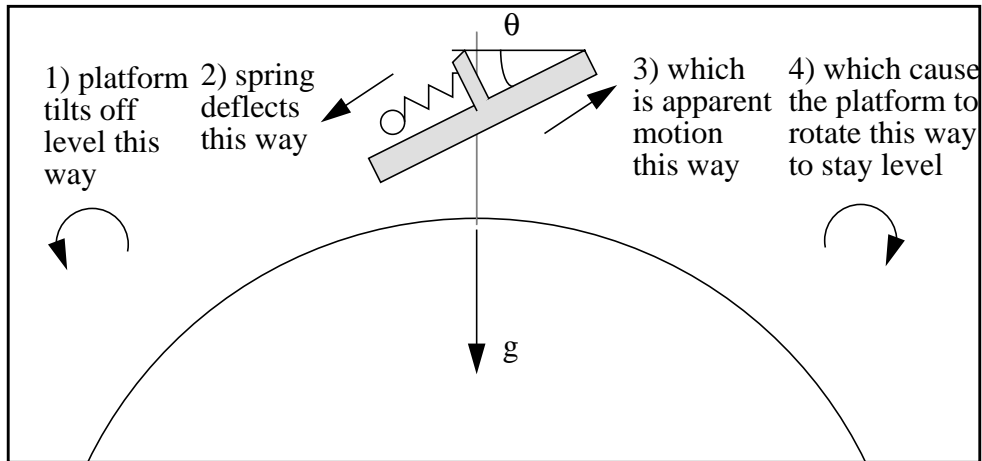


Figure 20 Accelerometer Negative Feedback

The dynamics of this loop can be investigated easily through the Laplace transform techniques of linear systems analysis. However, for the benefit of readers unfamiliar with such techniques, the time domain differential equations of the systems will be used. The accelerometer output can be written in terms of itself by following the feedback loop around backwards as follows:

$$a - g\theta = a - g \int \omega dt = a - \frac{g}{R} \int v dt$$

Equation 26

$$a - g\theta = a - \frac{g}{R} \iint [a - g\theta] dt$$

Next the common term of a is canceled, and the result is differentiated twice to obtain:

$$g\ddot{\theta} = \frac{g}{R} (a - g\theta)$$

Equation 27

$$\ddot{\theta} + \frac{g}{R}\theta = \frac{a}{R}$$

This is the differential equation describing the motion of the stable table. It is the characteristic equation of forced simple harmonic motion, which implies that the table will naturally oscillate at a period of:

$$T = 2\pi \sqrt{\frac{R}{g}} = 84 \text{ minutes}$$

Equation 28

The Schuler pendulum, having a radius equal to the earth's radius, can be accelerated arbitrarily around the surface of a sphere without being disturbed by the motion of the vehicle. In practice, one cannot construct such a pendulum, but any dynamic system with the correct natural frequency will exhibit this same stability property. In the context of the INS, it is the natural frequency which is important.

4.3 Schuler Loop

A stable table for a locally level system can be constructed on the simple principle that, when the table is not level, the horizontal accelerometers will measure a component of gravity. Consider a servo, called a **Schuler loop**, which computes the angular velocity of the vehicle with respect to the earth and rotates the table appropriately. Nominally, the required rate of rotation of the table is equal to the earth's rate with respect to inertial space plus the vehicle rate over the earth's surface. The accelerometer output is integrated to give a velocity signal, and this signal is then divided by the radius to the earth's center in order to get the angular velocity. This signal is used to torque the platform to keep it level. The velocity signal is also integrated a second time to give position. These ideas are summarized in the functional block diagram below.

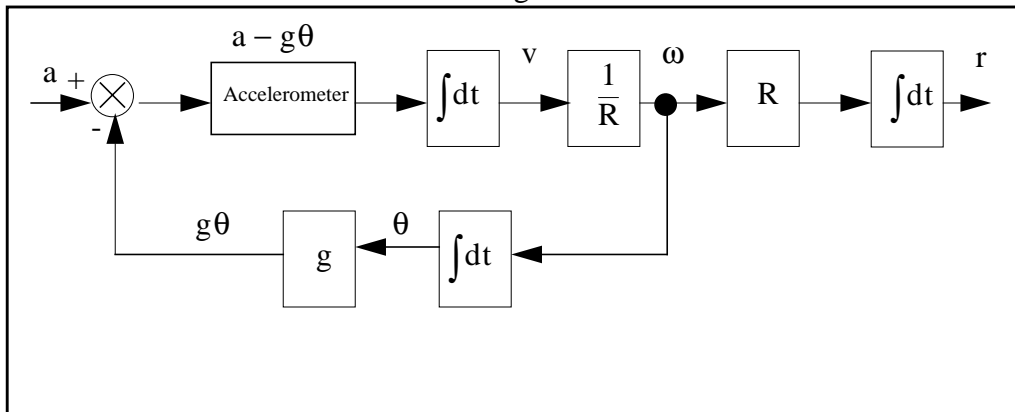


Figure 19 Schuler Loop

The feedback path is constructed by recognizing that the accelerometer specific force includes any component of gravity due to an off-level condition. This feedback is inherent in the accelerometer - there is no electrical connection required. If the table is off level, the accelerometer will measure the gravity component:

$$a = g \sin \theta \approx g \theta \text{ (for small angles)}$$

Equation 25

4 Schuler Dynamics

The gyrocompass oscillates like a pendulum with a characteristic period that is caused by the rotation of the earth. This principle of pendulous oscillation is central to the understanding of inertial systems and is developed further here.

A gyrocompass mounted on a moving ship is subject to motions much more pronounced than the earth's rotation. In an important paper published in 1923, Max Schuler, who was investigating this problem of the gyrocompass, pointed out that if the gyrocompass period was about 84 minutes, then it will be unaffected by motions of any vehicle.

4.1 Schuler Period

Many system variable errors in an inertial navigation system exhibit a period called the Schuler period⁴⁹ in the two level directions. The Schuler period is given by the following relation where R_e is the radius of the earth and g is gravity:

$$T = 2\pi \left(\frac{R_e}{g} \right)^{1/2} \approx 84.4 \text{ min}$$

The correct application of the Schuler principle is an essential component of precision inertial navigation systems. The Schuler oscillation is the natural horizontal oscillation in the error in computing gravitation.

4.2 Schuler Tuning

While precision components is one of the reasons why inertial navigation is possible, the primary reason for the feasibility of terrestrial inertial navigation is that all systems which navigate close to the earth experience horizontal oscillatory errors reflecting the famous **Schuler period** of 84 minutes. Horizontal error, being oscillatory, is inherently bounded.

Schuler tuning can be described as the construction of a device whose natural frequency is the same as that of a pendulum whose length is the radius of the earth. A pendulum attached to a vehicle moving on the surface of the earth will be deflected because of the vehicle acceleration as shown below. In fact, this is the principle of the pendulous accelerometer.

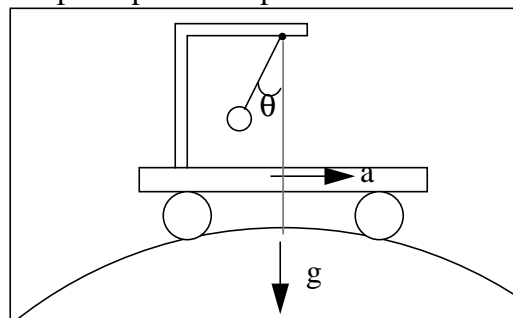


Figure 18 Accelerated Pendulum

49. Whereas a geosynchronous orbit has a period of exactly 24 hours, low earth orbits repeat several times a day. It turns out that the Schuler period is also equal to the period of an orbit at treetop height.

3.7 Direction Cosine Matrix

Once the Euler angles are known, the **direction cosine matrix** is given in [27] as:

$$\mathbf{R}_b^n = \begin{bmatrix} (c\psi c\phi - s\psi s\theta s\phi) & -s\psi c\theta & (c\psi s\phi + s\psi s\theta c\phi) \\ (s\psi c\phi + c\psi s\theta s\phi) & c\psi c\theta & (s\psi s\phi - c\psi s\theta c\phi) \\ -c\theta s\phi & s\theta & c\theta c\phi \end{bmatrix} \quad \text{Equation 23}$$

3.8 Position Solution

Finally, the position solution is given by:

$$\bar{\mathbf{r}}^n = \int_0^t \bar{\mathbf{v}}^n dt + \bar{\mathbf{r}}_0^n \quad \text{Equation 24}$$

3.3 Specific Force

The accelerometer specific forces are intrinsically referred to the body frame thus:

$$\bar{\mathbf{t}}^b = \begin{bmatrix} a_x & a_y & a_z \end{bmatrix}^T \quad \text{Equation 17}$$

These are referred to earth coordinates in the usual way:

$$\bar{\mathbf{t}}^n = \mathbf{R}_b^n \bar{\mathbf{t}}^b \quad \text{Equation 18}$$

3.4 Gravity

Gravity is intrinsically referred to earth coordinates:

$$\bar{\mathbf{g}}^n = \begin{bmatrix} g_x & g_y & g_z \end{bmatrix}^T \quad \text{Equation 19}$$

3.5 Velocity Solution

Finally, the initial velocity must be referred to earth coordinates, and all subsequent computed velocities will also be referred to earth coordinates through the time integration of:

$$\bar{\mathbf{v}}^n = \int_0^t \left[\mathbf{R}_b^n \bar{\mathbf{t}}^b - (\mathbf{R}_b^n \bar{\boldsymbol{\omega}}^b + \bar{\boldsymbol{\Omega}}^n) \times \bar{\mathbf{v}}^n + \bar{\mathbf{g}}^n \right] dt + \bar{\mathbf{v}}_0^n \quad \text{Equation 20}$$

3.6 Attitude

From [27], the rates of the Euler angles describing the orientation of the vehicle are nonlinearly related to the angular velocity vector $\bar{\boldsymbol{\rho}}^n$ by⁴⁷:

$$\dot{\bar{\mathbf{Y}}} = \begin{bmatrix} \dot{\theta} \\ \dot{\phi} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} c\phi & 0 & s\phi \\ t\theta s\phi & 1 & -t\theta c\phi \\ -\frac{s\phi}{c\theta} & 0 & \frac{c\phi}{c\theta} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \Gamma \bar{\boldsymbol{\rho}}^n \quad \text{Equation 21}$$

This expression can be directly integrated in order to determine the Euler angle expression of the attitude of the vehicle with respect to the earth. Thus⁴⁸:

$$\bar{\mathbf{Y}} = \int_0^t [\Gamma (\bar{\boldsymbol{\omega}}^n - \bar{\boldsymbol{\Omega}}^n)] dt + \bar{\mathbf{Y}}_0 \quad \text{Equation 22}$$

47. In [31], the matrix Γ is called $\boldsymbol{\Omega}$.

48. Recall earlier comments on the distinguished nature of 3D rotations. This quantity is not a vector in the sense of a vector space of linear algebra.

3 Complete Solution in Geocentric Cartesian Coordinates

One of the reasons for the particular expression of the navigation equations given in a earlier section is that the vector $\vec{\omega}$ is very close to the direct output of the gyroscopes. In the strapdown case, for an orthogonal arrangement of gyros, each has its sensitive axis aligned with one of the designated axes of the vehicle, and each responds directly to the component of the *inertial* angular velocity which is aligned with it.

3.1 Inertially Referenced Angular Velocity

Let a cartesian coordinate system called the navigation coordinate system, indicated by the letter n, be assigned to the center of the earth in accordance with the conventions set out in [27]. Let a set of axes called the body coordinate system be assigned to the vehicle, and indicated by the letter b. In accordance with the same conventions, z points up, y forward, and x out the right side. In the body coordinate system, the inertial angular velocity *referred to body coordinates* can be expressed as:

$$\vec{\omega}^b = \begin{bmatrix} \omega_x & \omega_y & \omega_z \end{bmatrix}^T \quad \text{Equation 13}$$

The rotation matrix which converts coordinates from the body frame to the navigation frame will be called the **direction cosine matrix** and denoted by R_b^n . It can be used to convert coordinates to the earth frame as follows:

$$\vec{\omega}^n = R_b^n \vec{\omega}^b \quad \text{Equation 14}$$

This quantity represents the angular velocity of the vehicle with respect to inertial space, referred to earth coordinates. The earth's sidereal rotational rate referred to earth coordinates is given by:

$$\vec{\Omega}^n = \begin{bmatrix} 0 & 0 & \Omega \end{bmatrix}^T \quad \text{Equation 15}$$

These two terms give the dual angular velocity term of the navigation equations.

3.2 Earth Referenced Angular Velocity

The gyros indicate the angular velocity of the vehicle with respect to inertial space. In order to determine the attitude with respect to the earth, the earth rate $\vec{\Omega}^n$ must be removed. Recall the definitions of the three angular velocities and then:

$$\vec{\rho}^n = \vec{\omega}^n - \vec{\Omega}^n = R_b^n \vec{\omega}^b - \vec{\Omega}^n \quad \text{Equation 16}$$

2.6 Temporal Error Propagation

It is instructive to ask whether any of the terms in the equations are insignificant and to ask what the result of neglecting it would be. For illustration purposes, let the vehicle rotation rate match the earth rate so that the vehicle is always level and $\vec{\omega} = \vec{\Omega}$. The earth's sidereal rate is 360° in 24 hours or 7.27×10^{-5} rad/sec and the radius to the equator is 7000 Km. For a vehicle at the equator, moving eastward at a velocity of 10 meters per second, and accelerating at 0.1 g, the following table gives the magnitude of each term:

Term	Nominal Value
Specific Force \vec{t}	0.1 g
Gravitational \vec{g}	1.0 g
Centrifugal $\vec{\Omega} \times \vec{\Omega} \times \vec{r}_e$	3.8×10^{-3} g
Coriolis $2\vec{\Omega} \times \vec{v}_e$	1.5×10^{-4} g

Figure 17 Terms in the Navigation Equation

While it may seem that the Coriolis term may be neglected, remember that the process of integration multiplies acceleration by the square of time, so that after 1 hour, neglecting the Coriolis term accounts for almost 10 kilometers of accumulated error. Note that this error occurs even when the system is stationary for that time period - so it is truly time dependent.

2.5 Generic Mechanization

The function of the INS is to solve, or mechanize, the above equations in order to determine the positional variables representing the motion of the vehicle over the surface of the earth. To do this, all terms in the vector equations must first be expressed in a common coordinate system. This matter will be deferred until the next section. For now, consider the problem in its coordinate system independent form.

The mechanization equations amount to a need to do four things:

- add gravitation to specific force
- remove centrifugal force due to the vehicle offset from the center of the earth
- remove coriolis force due to the vehicle's motion on the surface of the earth
- perform two integrations, incorporating initial conditions

The process of solving the equations in a computer consists of initializing the positional variables to the initial conditions, reading the accelerometers, computing gravitation, coriolis and centrifugal terms using the position and velocity from the last cycle, transforming specific force to earth referenced acceleration, and finally doing the integration to generate position and velocity for the current cycle. The solution of the equations can be visualized with the functional block diagram shown below:

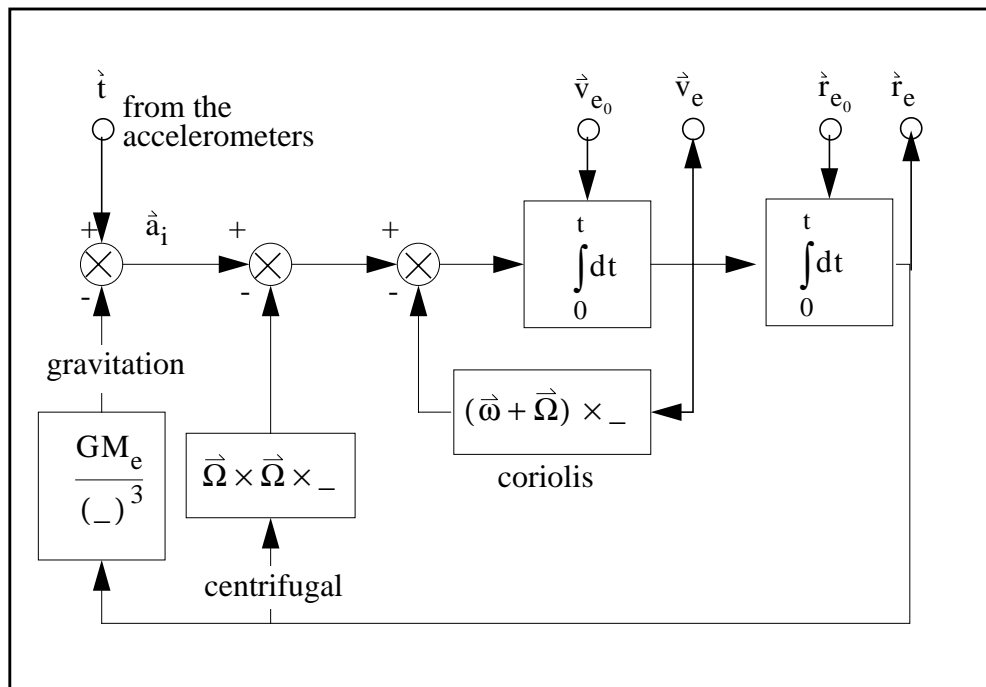


Figure 16 Generic Mechanization

Notice that the system provides a continuous output of the vehicle position, velocity, and acceleration with respect to the center of the earth.

2.4 Vector Integration

Solving for the quantities of interest:

$$\begin{aligned}\vec{v}_e &= \int_0^t [\dot{\vec{t}} - (\vec{\omega} + \vec{\Omega}) \times \vec{v}_e + \vec{g}] dt + \vec{v}_{e_0} \\ \vec{r}_e &= \int_0^t \vec{v}_e dt + \vec{r}_{e_0}\end{aligned}\quad \text{Equation 12}$$

Now these expressions are *vector integrals*, and the vectors must be converted into a common coordinate system before they are integrated. It is the position of the vehicle with respect to the earth which is required, so the integrands must be referred to an earth fixed (e frame) coordinate system before they can be used.

It is incorrect to integrate them in a coordinate system fixed to the vehicle frame. One way to see this is that the velocity vector is often of constant orientation in the vehicle frame - even though the vehicle rotates in the earth frame. Integration of the forward velocity without accounting for rotation leads to the conclusion that motion is always rectilinear, which obviously cannot be right.

These equations can be solved for the unknown position, and velocity, given the following information:

- a model of the earth's acceleration due to gravitation as a function of position
- the earth's sidereal rate of rotation
- the vehicle rotation rate with respect to the earth
- the specific forces from the accelerometers
- the initial position
- the initial velocity

2.3 Vehicle Referenced Acceleration

The first term on the right hand side can be referred to the vehicle frame by another application of the theorem of Coriolis:

$$\left(\frac{d\vec{v}_e}{dt}\right)_i = \left(\frac{d\vec{v}_e}{dt}\right)_v + \vec{\omega} \times \vec{v}_e \quad \text{Equation 7}$$

Substituting Equation 7 and Equation 5 into Equation 6 yields:

$$\vec{a}_i = \left(\frac{d\vec{v}_e}{dt}\right)_v + (\vec{\omega} + \vec{\Omega}) \times \vec{v}_e + \vec{\Omega} \times (\vec{\Omega} \times \vec{r}_e) \quad \text{Equation 8}$$

Substituting this into Equation 2:

$$\left(\frac{d\vec{v}_e}{dt}\right)_v = \dot{\vec{t}} - (\vec{\omega} + \vec{\Omega}) \times \vec{v}_e + \vec{w} - \vec{\Omega} \times (\vec{\Omega} \times \vec{r}_e), \quad \text{Equation 9}$$

This is one of the most convenient forms of the equation of motion of the vehicle. The last two terms on the right hand side are both functions of only the position and are often grouped together and called **gravity**. Gravity is part gravitation and part centrifugal force:

$$\vec{g} = \vec{w} - \vec{\Omega} \times (\vec{\Omega} \times \vec{r}_e) \quad \text{Equation 10}$$

This gives the vehicle referenced acceleration as:

$$\left(\frac{d\vec{v}_e}{dt}\right)_v = [\dot{\vec{t}} - (\vec{\omega} + \vec{\Omega}) \times \vec{v}_e + \vec{g}] \quad \text{Equation 11}$$

In precise terms, this quantity is the derivative of the velocity of the vehicle with respect to the earth, where the derivative is computed by a vehicle fixed observer. It is not the derivative of the velocity of the vehicle with respect to the vehicle, which would be zero. Further, it is also not the derivative of the vehicle velocity with respect to the earth that would be measured by an earth fixed observer, because it includes terms for the rotation of the vehicle with respect to the earth.

The reason for the use of such a quantity is that, in the most general case, the accelerometers are strapped down to the vehicle chassis. By the Theorem of Coriolis, the rotation of the vehicle with respect to the earth adds an extra amount to the the rotation of the moving frame.

2.2 Inertial Acceleration to Earth Referenced Acceleration

In the conversion of the basic equation to an earth referenced form, it is necessary to recognize that all inertial frames of reference are equivalent from the point of view of measuring acceleration. Hence, the previous result can be considered to be referenced to any inertial frame. It can be shown that a geocentric frame of reference that is not rotating with respect to the stars is inertial.

In the following result, two frames of reference are defined to be situated at the center of the earth. The geocentric inertial frame is associated with the subscript i and the geocentric frame which rotates at the earth's sidereal rate is associated with the subscript e. Another frame of reference is associated with the moving vehicle, and has subscript v.

These three frames are in relative rotation. Let the constant rotation of the earth with respect to inertial space be given by $\vec{\Omega}$, and the rotation of the vehicle frame with respect to the earth be given by $\vec{\rho}$. If $\vec{\omega}$ denotes the inertial rate of the vehicle frame, it can be expressed as:

$$\vec{\omega} = \vec{\Omega} + \vec{\rho} \quad \text{Equation 4}$$

Let \vec{r}_x , \vec{v}_x , and \vec{a}_x be the position vector, velocity, and acceleration of the vehicle measured in the frame x. Then, since the position vector is the same in both the inertial and earth frames, the inertial velocity of the vehicle can be expressed as:

$$\vec{v}_i = \left(\frac{d\vec{r}_e}{dt} \right)_i = \left(\frac{d\vec{r}_e}{dt} \right)_e + \vec{\Omega} \times \vec{r}_e = \vec{v}_e + \vec{\Omega} \times \vec{r}_e \quad \text{Equation 5}$$

The result of differentiating the above one more time in the inertial frame is:

$$\vec{a}_i = \left(\frac{d\vec{v}_i}{dt} \right)_i = \left(\frac{d\vec{v}_e}{dt} \right)_i + \vec{\Omega} \times \left(\frac{d\vec{r}_e}{dt} \right)_i \quad \text{Equation 6}$$

Let the vehicle exert a force \vec{T} on the spring because the vehicle is moving, and let the gravitational force be \vec{W} and the inertial acceleration be \vec{a}_i of the mass m . The device registers the specific force, so while the mass responds to the sum of these two forces, the spring responds only to the force \vec{T} . After the spring has finished extending, the motion of the mass is the same as that of the vehicle to which it is attached. The fundamental equation of inertial navigation is Newton's second law applied to the proof mass:

$$\sum \mathbf{F} = \vec{T} + \vec{W} = m\vec{a}_i \quad \text{Equation 1}$$

The specific force is the quantity \vec{T}/m . The inertial acceleration can be expressed in terms of the specific and gravitational forces as follows:

$$\vec{a}_i = \frac{\vec{T}}{m} + \frac{\vec{W}}{m} = \vec{t} + \vec{w} \quad \text{Equation 2}$$

At this point, it is possible to convert the specific force into inertial acceleration, so that it can be integrated. Notice that explicit knowledge of the gravitational field strength is required at every position of the vehicle. Inertial navigation is only viable when this field is known or known to be insignificant.

Since most inertial systems operate near the earth, the gravitational effects of the other heavenly bodies are relatively small, so it is the earth's gravitational field which is required. A practical gravitation model must account for the earth's oblate spheroidal shape and other anomalies⁴⁶, but for purposes of illustration, consider the model of an ideal spherical, homogeneous earth which is given by Newton's law of universal gravitation applied to a point mass at the earth's center. Let G be the universal gravitation constant, M_e be the mass of the earth and \vec{r}_e be the radius of the vehicle from the center of the earth:

$$\vec{w} = -\frac{GM_e}{|\vec{r}_e|^3} \vec{r}_e \quad \text{Equation 3}$$

46. The practice of mapping the earth's gravitation is called geodesy.

2 Generic Mechanization

It is the engineering solutions to the problems of the last section which cause the modern INS to take the form that it does. The following discussion will explain how each of these three problems are corrected in the following order:

- convert specific force to inertial acceleration
- convert inertial motion to earth referenced motion
- express earth referenced motion in the navigation coordinate system.

2.1 Specific Force to Inertial Acceleration

Recall that the accelerometer functions by coupling a proof mass to the instrument case through a calibrated restraint. The deflection of the device is a measure of the force or torque experienced as a result of its accelerated motion. Higher quality devices employ rebalance loops which prevent the mass from moving, and the effort required to do this is a measure of specific force.

Such a device can be modelled conceptually as a mass attached to a spring which is constrained to move only in a single direction within a frictionless tube as shown in Figure 15. Suppose both the tube and the end of the spring are attached to the vehicle and the mass is free to move while the vehicle accelerates upward.

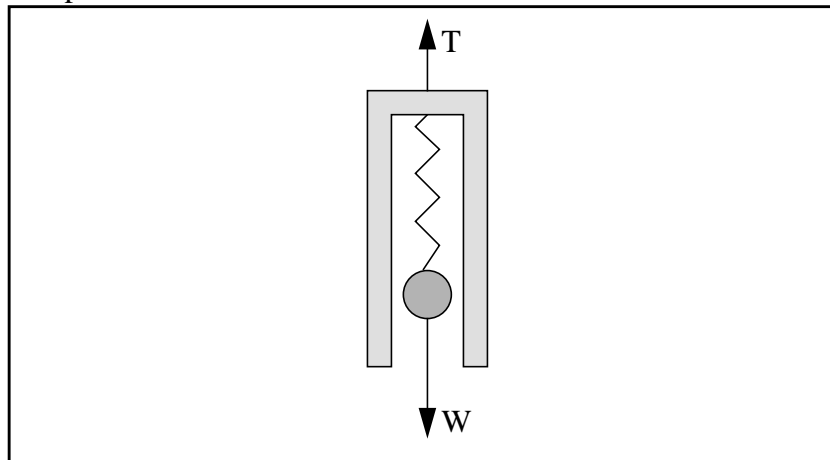


Figure 15 Accelerometer

1.2.3 Gravitation

Recall that accelerometers are adversely affected by gravitational fields, because they actually measure a quantity called **specific force**, and not even inertial acceleration. The solution to this problem is to account for the effect of the gravitational field of the earth in the mechanization equations.

1.2.4 Summary

In summary, the naive implementation of an inertial navigation system suffers from the following set of fundamental problems:

- the accelerometers measure the wrong quantity
- they measure it in the wrong reference frame
- they represent it in the wrong coordinate system

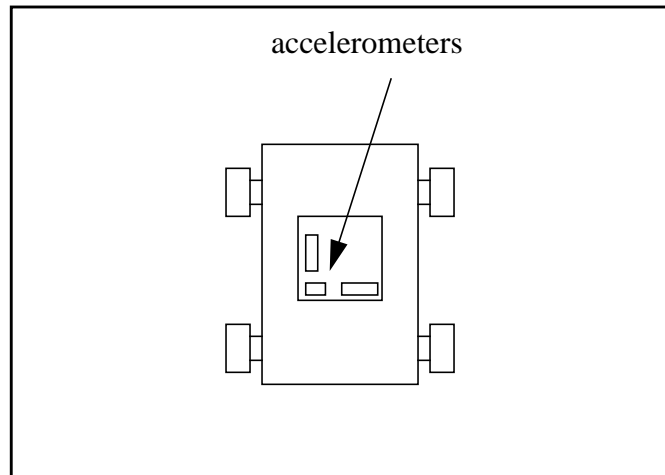


Figure 14 Naive INS Concept

1.2.1 Coordinate Systems

Deduced reckoning theoretically consists of adding up a number of differential displacement vectors. After the accelerometer reading is integrated over a time step twice, the result is a differential displacement of the vehicle with respect to the earth. In this case, since the axes of the accelerometers move with the vehicle, the displacement vectors are inherently expressed in a coordinate system which rotates with the vehicle.

In order to add vectors of any kind, they must all first be expressed in a common coordinate system⁴⁵. Further, that coordinate system must be fixed with respect to the reference frame of interest. Therefore, all displacement vectors must first be expressed in a coordinate system fixed with respect to the earth, not the vehicle. In order to do the transformation, a sensor, or collection of sensors, is needed which measures the total rigid body orientation of the vehicle - pitch, yaw, and roll. The solution to this coordinate system problem is to incorporate a set gyroscopes to measure the orientation of the accelerometer triad, and to physically or computationally resolve the differential displacements into an earth fixed coordinate system before adding them.

1.2.2 Reference Frames

Acceleration is inherently an inertially referenced vector, so accelerometers basically measure acceleration of the vehicle with respect to inertial space. Also, gyroscopes inherently measure inertial rotation, that is, rotation with respect to the stars, not with respect to the earth. These properties result from the physical principles upon which the devices are based. The solution to this reference frame problem is to use the transformation laws of mechanics to convert the inertially referenced quantities to earth referenced quantities. In the case of the accelerometers, the coriolis and centrifugal forces caused by the rotation of the earth must be accounted for. In the case of the gyros, their orientation reference must be somehow altered so that it rotates with the earth.

45. The reader is encouraged to review the discussion of the distinction between coordinate systems and reference frames in an earlier section.

globe, and the atmosphere, provided the local gravitational field is known. They are feasible alternatives in many places where other types of systems cannot be used, including interplanetary space, underground, and undersea. In this respect, they are the most generally applicable navigation technology in existence.

Some people now envision a day soon when inertial navigation systems will be robust enough and cheap enough to become standard equipment on automobiles. To the ancient mariner, these little black boxes which radiate nothing and do not need to see, touch, or otherwise sense the environment would have likely seemed nothing less than sorcery.

The inertial navigation system (INS) can be distinguished from other types of navigation instruments by the following characteristics:

- requiring no external information, are jamproof
- radiating nothing, exhibit perfect stealth
- cannot sense accelerations of unpowered space flight
- can operate only in a known gravitational field
- provide continuous or sampled position, velocity, and acceleration
- achieve highest accuracies after warm up and drift trim
- most errors exhibit Schuler oscillation
- most errors are time, not speed or distance related
- require input of initial position, velocity, and attitude

1.2 The Naive INS

The basic principle employed in inertial navigation is deduced reckoning. A set of accelerometers are used to measure acceleration along some set of orthogonal axes, and their outputs are integrated twice to determine position. The integration is usually performed numerically in a computer. Although the above statement of the principle is an accurate one, it does not address a large number of other practical engineering considerations. The complexity of a modern inertial navigation system, incorporating gimbals, accelerometers, gyros, and a stable table results directly from the need to **mechanize** the above principle in real hardware.

Rather than stating the generic design of an INS, the discussion will start by explaining why all of the components are necessary. Most of the issues can be illustrated by examining the problems which arise when accelerometers are simply strapped physically to a vehicle as shown in Figure 14. Although it is customary to mount the accelerometers along orthogonal axes, it is not necessary. Suppose three accelerometers are fixed such that their sensitive axes are oriented with one along the vehicle longitudinal axis, one transverse to the vehicle longitudinal axis, and one vertical. Further, suppose that the accelerometer outputs are read directly and integrated twice.

While it may seem that this is a basically sound approach, the reader may be surprised to discover that it is fundamentally flawed in several ways. This naive implementation of the basic principle suffers from all of the problems listed below.

PART III: Inertial Navigation Systems⁴⁴

The name inertial navigation or inertial guidance derives from the use of inertial principles by certain types of navigation systems. These systems incorporate a large number of components including accelerometers, gyroscopes, computers, and sometimes gimballed stable platforms - all of which are engineered to perform together to supply navigation information.

Early versions of the INS were used by the Peenemünde group in Germany in World War II to guide the A3 series rocket and later the V2 rocket. The navigation systems were enabled during the powered ascent, but were shut off with the rocket engines, since the rest of the vehicle trajectory became ballistic and therefore predictable. The design of complete systems in their modern form first occurred in the 1940's and is credited to Charles Draper and his colleagues at MIT.

1. Basic Principle of Inertial Navigation

1.1 Characteristics of the INS

These devices exhibit a large number of ideal characteristics. For the military, they are a favorite navigation aid because they radiate no energy that may assist an adversary in locating the vehicle. Also, they require no information from the environment, so they are indifferent to weather conditions, and the attendant poor visibility of landmarks, and they cannot be jammed or interfered with. They can even be used on vehicles travelling at high enough speeds to ionize the surrounding air, and effectively jam their own navigation radar.

They can be used easily in unsurveyed areas where landmark positions are unknown, in featureless terrain where there are no landmarks, and in situations where artificial landmarks cannot be installed for any of a variety of reasons. They are ubiquitous components in missiles, tanks, satellites and long haul commercial aircraft.

Inertial navigation systems (INS) are important to mobile robotics because they provide the most accurate form of dead reckoning currently available. Also, the use of inertial principles implies that these devices can be used literally anywhere in the universe, including the entire surface of the

44. This section has benefitted from the kind assistance of the following individuals:
Yash Amin, Teledyne Ryan Electronics, San Diego, Ca
Kevin Dowling, Research Engineer, Field Robotics Center, Pittsburgh, Pa.
Jim Farrell, Westinghouse, Severna Park, Maryland.
Ed Luxford, Senior Program Manager, Allied Signal, Teterboro, NJ.
Gerry Roston, Graduate Student, Field Robotics Center, Pittsburgh, Pa.
Larry Valot, Honeywell, Minneapolis, Minnesota.
John Yamomoto, Litton Systems Inc., Woodland Hills, Ca.

5.6 Optical Fibre Gyroscopes

Unlike the RLG, the fibre optic gyroscope (FOG), employs the Sagnac effect directly. It achieves the required sensitivity increase by increasing the path length. Recasting the earlier Sagnac length difference result in terms of a phase difference for a single round trip between the counterrotating beams:

$$\Delta\phi = \frac{2\pi\Delta L}{\lambda} = \frac{8\pi^2 r^2 \Omega}{c\lambda}$$

Now by using n coils of a fibre optic light pipe, this phase difference is multiplied by n . Substituting L for $2\pi r$:

$$\Delta\phi = \frac{8n\pi^2 r^2 \Omega}{c\lambda} = \frac{4n\pi L r \Omega}{c\lambda}$$

which shows that the phase shift grows linearly with the number of turns of the fiber. Current bendable optical fibers allow several miles of extremely thin cable to be coiled into a compact package. FOGs may be simpler, more reliable, and less costly than the RLG. They do not suffer from lock-in, require no mechanical dither, and require no precision machined blocks. They require no critical mirrors, and therefore do not suffer from mirror aging. They promise to replace the RLG even as it continues to replace the mechanical gyro today.

5.7 Practical Optical Fibre Gyroscopes

The feasibility of the FOG had to await the development of the optical fiber, and its development closely follows optical communications technology. The first experimental FOGs were produced in 1976, and research has been feverish since. The first all single mode FOG was reported in 1981. As new problems are discovered and quantified, solutions are found quickly and performance improves. In one case, the Kerr effect, identified in 1982, two solutions to the problem were published only a few months afterward.

Current designs are only useful for modest applications, and have been proposed for automobiles. These employ a ring interferometer constructed of optical fiber coiled into a compact loop, and a semiconductor laser, similar to those used in optical disks. A beam splitter splits the beam to send it in both directions simultaneously around the loop. At the end, both beams are recombined and sent to a photodetector.

An effect called polarization instability is now believed to be the substantial limiting factor preventing real performance from approaching the theoretical. The polarization of the beam in current fibers is unstable, so interferometer performance is degraded. In some designs, polarization maintaining fibers are used in a special loop before the main loop, to stabilize the polarization of the beams. With this enhancement, the research community now believes that the major problems encountered by the original systems are understood, and the FOG is now entering the development and engineering phase. Their use in high performance inertial sensors is now considered to be inevitable.

One of the practical problems of building these devices is that the two beams tend to compete for excited atoms, and, if left alone, one beam or the other will eventually win the competition, and half of the laser will be shut off. One method available to counter this effect is to use two different lasing isotopes which are individually tuned to one beam or the other. The neon laser is a popular choice for RLGs because of the near ideal separation of the two isotopes neon-20 and neon-22.

A more important practical problem is the **lock-in** phenomenon, which is a property of all oscillators. Imperfections in the mirrors used in the laser cause a portion of the clockwise beam to be reflected back along the path of the counterclockwise beam and vice versa. This, in turn, causes the two beams to approach each other in frequency when the frequency difference is small, as it will be for small rotational rates. The net effect is a large deadband in the response of the device. Rotational rates below a threshold will not be sensed at all because the beams will lock and the beat frequency will be zero. Left unchecked, this mode locking gives a deadband that is 10,000 times that required for acceptable angular rate sensitivity.

The classical approach to any deadband problem is to introduce bias. That is, to artificially push the device away from operation near zero input through some means. For the RLG, the bias takes the classical form of mechanical **dither**. The device is mechanically shaken at high frequency and imperceptible amplitude, so that the true rotational rate is measured as an increase over the rate artificially induced by the dither. The practical RLG loses its theoretical designation as a solid state device because of this, and research is underway to discover a way to create optical bias.

Unlike earlier modular versions, modern RLGs sacrifice the maintenance convenience of modularity for the structural stability of a single monolithic machined block for all three axes. Cavities are precision bored to hold the lasing medium, the mirrors, and the electrodes into a single compact package. A typical configuration is a cube where three sets of four faces each are used to implement three mutually orthogonal closed loops. One such loop is indicated in the figure below:

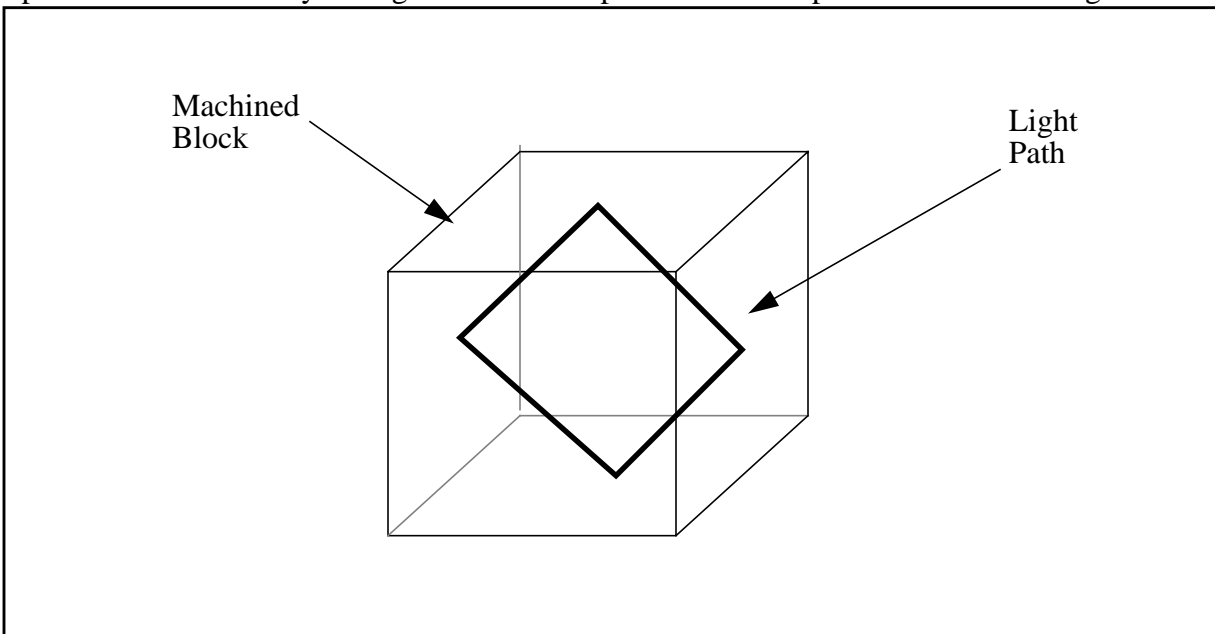


Figure 13 Typical Monolithic Ring Laser Gyro Configuration

5.3 Lasers

The term **laser** stands for light amplification by stimulated emission of radiation. The laser is basically an optical oscillator. As such, it incorporates all of the features common to an electrical oscillator, namely filtering, and positive feedback. The filtering is provided by the fact that the resonance cavity sets up standing waves whose wavelength must divide the cavity length by an integer. Positive feedback arises from the use of a lasing medium, a gas, liquid, crystal ions, or any of a number of other possibilities. The lasing medium has the property that, under the conditions of operation, excited atoms can be stimulated to release new photons of radiation of precisely the same phase, frequency, and directional characteristics as the energy stimulating the atom. Where there was one photon, a new identical copy appears as well. Energy is continually supplied, or **pumped**, to the system to excite the lasing medium atoms.

5.4 Ring Laser Gyroscopes

The ring laser gyro (RLG) combines the features of a ring interferometer and a laser into the same device. The lasing medium is placed directly into the path of the counterrotating beams and coherent light is made to circulate in the ring.

Now, since the laser light wavelength is related to the cavity length, and since the apparent cavity length is different for both beams by the Sagnac effect, it follows that the frequency of the two beams will slightly differ. For a path length of L , an integer m , and wavelength λ , the resonance condition can be expressed as:

$$m\lambda = L$$

Rewriting this in terms of the resonant frequency ω :

$$\omega = \frac{mc}{L}$$

Using the previous result for the Sagnac effect, the frequency difference for the two beams can be approximated by:

$$\Delta\omega = \omega_+ - \omega_- = \frac{mc}{L_+} - \frac{mc}{L_-} \approx \frac{mc\Delta L}{L^2} = \frac{\omega\Delta L}{L}$$

Since light frequencies are extremely high, this frequency difference is larger than the Sagnac path differential by the considerable factor of ω/L . This amplification, combined with the relative ease of measuring the beat frequency, typically in the kHz range, makes the RLG a practical device. It is important to recognize that the RLG measures the frequency difference, not the path length difference given by the Sagnac effect.

5.5 Practical Ring Laser Gyroscopes

The development of the RLG had to wait for the laser to be developed. In 1962, a mere two years after the helium-neon laser was developed, Macek of Sperry Gyroscope company, produced a one meter square prototype that measured the earth's rotation of 1/4 rad/hour. Sagnac's original interferometer had a sensitivity of about 2 rad/sec, so Macek's was over 10,000 times more sensitive.

5.2.2 Apparent Path Length Difference

Expressions for these can be written in terms of themselves by noting that the apparent path length difference is given by the arc length travelled by the interferometer during the transit time, or $r\Omega t$.

$$t_+ = \frac{2\pi r + r\Omega t_+}{c} \qquad t_- = \frac{2\pi r - r\Omega t_-}{c}$$

Solving for the times and taking the difference gives:

$$\Delta t = t_+ - t_- = \frac{4\pi r^2 \Omega}{c^2 - r^2 \Omega^2}$$

For practical values of r , $(r\Omega)^2 \ll c^2$, so the expression simplifies to:

$$\Delta t = \frac{4\pi r^2 \Omega}{c^2}$$

Thus, the round trip path difference is:

$$\Delta L = c\Delta t = \frac{4\pi r^2 \Omega}{c}$$

Real devices also typically use a square or triangular loop, but this does not significantly affect the formula. For arbitrary shaped devices, the circular area is replaced by the polygonal area and the rotation rate is projected onto the normal to this area. Unfortunately, the path lengths given by this formula are much less than an optical wavelength, which should be clear when contemplating how fast light traverses a distance of a few meters⁴³.

For a loop area of one square meter, and a rotation rate of one degree per second, the device measures only a few angstroms of deviation. Real applications require much smaller sized devices and sensitivities many thousand times this one degree per second figure. For this reason, the Sagnac effect was a mere scientific curiosity for decades after it was discovered. However, as a result of two recent technical developments, lasers, and fibre optics, two different ways of significantly increasing the sensitivity of the effect have become possible.

43. As a convenient mnemonic, light travels a foot in a nanosecond.

5.2.1 Ring Interferometer

The ring interferometer, here composed of four mirrors, splits the beam into two counterrotating components and they are recombined after a complete transit of the loop to observe interference fringes. Sagnac showed that rotation of the entire loop gives rise to an apparent lengthening of the path length for one beam, and an apparent shortening for the other.

This path length difference gives rise to motion of the interference fringes which is a direct measure of the rotation of the device in inertial space. Where Michelson showed that it is impossible to measure inertial translation using light because the speed of light is the same for all observers, whatever their state of motion, Sagnac showed that optical measurement of inertial rotation was possible.

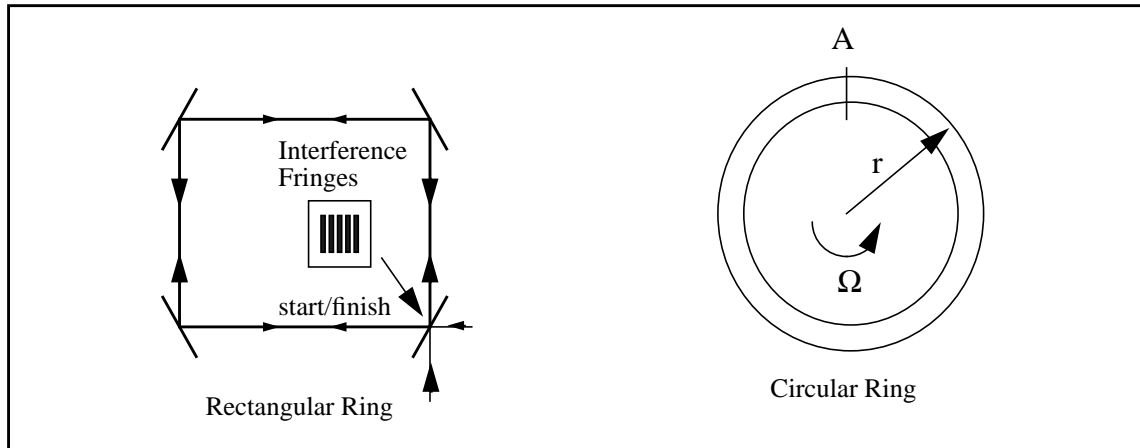


Figure 12 Ring Interferometer

There are two intuitive interpretations of the effect, based on different propagation path lengths, or different times of transit for both beams, and both lead to the same conclusion, namely a relative phase shift of both beams.

A rigorous derivation of the effect can only be done using general relativity, because in the presence of a gravitational field, even special relativity cannot be used. However, there is a classical derivation due to Aronowitz, which gives the right answer and is intuitively clear. To simplify things, consider a perfectly circular ring interferometer of radius r . Let the beam enter at point A, be split into a clockwise and a counterclockwise beam, complete one circuit, and be recombined again at A. Let the device rotate at a rate Ω and let c be the speed of light. Let t_+ and t_- denote the round trip transit times for the counterclockwise and clockwise beams respectively.

5 Optical Gyroscopes

With the recent arrival and rapid improvement of optical gyroscopes, the complete disappearance of mechanical gyroscopes in most applications now seems inevitable within the next ten years. For this reason, these devices which were once considered impractical, have now become required knowledge in the field of inertial navigation. This section will briefly overview the principles of operation of these devices.

Current commercial optical gyros are designed to be strapped down directly to the vehicle - requiring no stabilized platform at all. In this respect, they complement the general trend toward strapdown technologies. Optical gyros are fundamentally solid state devices, and because of this, are more reliable, more rugged, and are able to withstand high g fields beyond the capabilities of mechanical devices. They can be designed for harsh environments, exhibit large dynamic range⁴⁰ and exhibit linearity of parts per billion over that range. These devices are currently in use on the latest civil transport aircraft, including the Boeing 757, and 767, and the Airbus A310, and have now logged millions of flight hours.

5.1 Michelson Interferometer

In the period shortly before relativity theory was developed by Einstein, physicists believed that light propagated through an invisible, odorless, tasteless, medium that pervaded all of space called the **luminiferous ether**. In 1881, A. A. Michelson conducted an experiment which demonstrated this idea to be wrong and set the stage for the development of the special theory of relativity⁴¹. Michelson was the first American to be awarded the Nobel Prize in science.

He demonstrated this through the use of a device of incredible sensitivity which he invented, called the Michelson interferometer⁴². He showed that the “ether wind”, supposed to result from the earth’s motion through space, was undetectable by devices that were easily sensitive enough to detect it if it existed.

5.2 Sagnac Effect

Consider now a slightly different apparatus, called a **ring interferometer**, upon which optical gyroscopes are based, as indicated in Figure 12. If the device is made to simultaneously propagate two counterrotating beams of coherent light, it can be used to measure rotation according to an effect discovered in 1913 by G. Sagnac and known today as the **Sagnac effect**. All current optical gyroscopes are based on the Sagnac effect.

40. Dynamic range is the range of rotation rates that can be measured correctly.

41. Although Einstein is justly credited with the theory, many of the ideas resulting from it were stated earlier by Lorentz and also by Poincare’. Einstein’s contribution was to show that all of the fudge factors being proposed to explain Michelson’s results could be derived from a very small set of simple postulates.

42. To this day, devices based on interferometry are among the most precise instruments built. This arises from their ability to measure distance on the scale of the optical wavelength.

4.3 Interpretation as a Pendulous Oscillator

Now this is the equation of a pendulum. The variable θ oscillates like a pendulum, or equivalently, the spin axis oscillates about the local north-south direction. Addition of a damping system will cause the device to settle into pointing north.

Notice that, like a magnetic compass, the device will have trouble near the poles³⁸ because of the latitude dependence, and that the rate of oscillation is proportional to the spin rate. Also, if the earth did not rotate, the device would not work at all. In this respect, it operates as a compass because the earth is rotating. It is also interesting to note that the device can be used to determine latitude, by measuring the spin rate of the device and its period of oscillation.

The operation of the device can be understood on an intuitive level as follows. It is mounted on a spinning earth³⁹, so it is experiencing a very small torque transmitted through its fixture sufficient to rotate it once per day in inertial space. This torque is parallel to the earth's axis itself, but only that component of it which is oriented along the y axis in the figure can be transmitted through the gimbals to the rotor. The device is a gyroscope, so it will precess according to the gyroscope rigidity equation, but this precession will be normal to both the spin axis and the axis of the earth's rotation. The kinematics is such that the tendency to precess is opposite to the current direction of rotation, so the device oscillates.

4.4 Period of Oscillation

Using typical figures for the parameters involved, the period of oscillation can be calculated. The inertia ratio I_s/I_a is close to 2.0 for a flat circular object. Let the latitude be the equator, and the rotor spin rate be a typical value of 3600 rpm or 60 rev/sec:

$$T = 2\pi \left(\frac{I_a}{I_s (\omega_e \cos \lambda) \Omega} \right)^{1/2}$$
$$T = 2\pi \left(\frac{24 \times 60 \times 60}{2 \times 2\pi \times 2\pi \times 60} \right)^{1/2}$$
$$T = 47 \text{secs}$$

A very high rate of spin is required in order to get a reasonable period of oscillation. This is important because, depending on the initial conditions, the damping cannot bring the device into alignment any faster than one period, and a device which takes an hour to settle is useless. Notice that even for a rotor spin of 3600 rpm, the pendulous oscillation will be imperceptible to the eye.

38. However, the gyrocompass measures true north, not magnetic north, so it is immune to the complicated motions of the magnetic pole.

39. This is the fourth time in this report that a phenomenon that measures the earth's rotation has occurred: the Foucault pendulum, the eastward deflection of a dropped object, the gyroscope pointed at a star, and the gyrocompass.

4.2 Expression in Principle Coordinates

Let I_s denote the moment of inertia about the symmetry axis z , and I_a denote the moment about the other two axes. Performing the cross product and moving subscripts to superscripts to allow indication of components gives:

$$1: N_x = I_a \dot{\omega}_x^s + (I_s \omega_y^r \omega_z^s - I_a \omega_z^r \omega_y^s)$$

$$2: N_y = I_y \dot{\omega}_y^s + (I_a \omega_z^r \omega_x^s - I_s \omega_x^r \omega_z^s)$$

$$3: N_z = I_s \dot{\omega}_z^s + (I_a \omega_x^r \omega_y^s - I_a \omega_y^r \omega_x^s)$$

This equation looks incorrect because it is not the standard form of the scalar Euler's equation, but this arises from the use of the rotating frame as the moving frame and not the spinning frame. Consider the last equation. The z axis is gimbaled, so it can exert no torque. Also since $\omega_x^r = \omega_x^s$ and $\omega_y^s = \omega_y^r$, the bracketed term is zero. This implies that ω_z^s is constant. This constant is composed of two terms proportional to the earth's rate of rotation and the rotor's rate of rotation. The rotor spins at many orders of magnitude higher velocity, so the earth's rate can be neglected. It can be concluded that:

$$\omega_z^s = \text{constant} = \omega_e \cos \lambda \cos \theta + \Omega \approx \Omega \text{ (since } \Omega \gg \omega_e \text{)}$$

Consider the first equation. The x axis is also gimbaled, so there can be no applied torque. Also the second term in brackets is of order $(\omega_e)^2$ and can therefore be neglected. Rewriting:

$$I_a \dot{\omega}_x^s = -I_s \omega_y^r \Omega$$

$$I_a \ddot{\theta} = -I_s (\omega_e \cos \lambda \sin \theta) \Omega$$

$$\ddot{\theta} + \frac{I_s (\omega_e \cos \lambda) \Omega}{I_a} \sin \theta = 0$$

about its local vertical, and the device itself rotates at a high rate about its symmetry axis. Let the gimbal rotation be given by θ . The discussion will be concerned with four different frames of reference and their relative motion. A frame positioned at the center of the earth which is not rotating is inertial (subscript i). The spinning earth frame will have subscript e. The gyrocompass is positioned on the rotating earth and rotates about the local vertical. A frame of reference attached to the compass but not spinning around its symmetry axis will be called the rotating frame (subscript r). Finally, a frame attached to the spinning rotor will be called the spinning frame (subscript s).

The problem is solved by writing Euler's equation in the rotating frame. Although it is not spinning, this frame is chosen because the inertia tensor is still constant. Also, it is the frame whose motion will ultimately prove interesting. A cartesian coordinate system for this frame is given in the figure. The angular velocity of each noninertial frame with respect to the inertial frame will be expressed in terms of the angles defined in Figure 11 and expressed in the chosen coordinate system. In doing so, the fact that angular velocities add like vectors³⁷ will be used.

$\vec{\omega}_e$	rotation of earth frame in inertial frame
$\vec{\omega}_r$	rotation of rotating frame in inertial frame
$\vec{\omega}_s$	rotation of spinning frame in inertial frame
$\vec{\Omega}$	rotation of spinning frame in rotating frame

Now all of these are expressed in the rotating frame of reference.

$$\vec{\omega}_e = \omega_e \sin\lambda \hat{i} + \omega_e \cos\lambda \sin\theta \hat{j} + \omega_e \cos\lambda \cos\theta \hat{k}$$

$$\vec{\omega}_r = \vec{\omega}_e + \dot{\theta} \hat{i} = (\omega_e \sin\lambda + \dot{\theta}) \hat{i} + \omega_e \cos\lambda \sin\theta \hat{j} + \omega_e \cos\lambda \cos\theta \hat{k}$$

$$\vec{\omega}_s = \vec{\omega}_r + \Omega \hat{k} = (\omega_e \sin\lambda + \dot{\theta}) \hat{i} + \omega_e \cos\lambda \sin\theta \hat{j} + (\omega_e \cos\lambda \cos\theta + \Omega) \hat{k}$$

Substitution of these complicated expressions will be avoided in the following development. Consider now Euler's equation applied to the rotor. The theorem of Coriolis will be used where the inertial frame is the fixed frame and the rotating frame is the moving one. The theorem takes the form:

$$\vec{N} = \left(\frac{d\vec{L}_i}{dt} \right)_r + \vec{\omega}_r \times \vec{L}_i \qquad \vec{N} = I_i \left(\frac{d\vec{\omega}_s}{dt} \right)_r + \vec{\omega}_r \times I_i \vec{\omega}_s$$

Notice that the two angular velocities of the cross product are now different. This is because the first angular velocity is the relative rotation of the two frames and comes from the theorem of Coriolis while the second is the rotation of the body in inertial space because it is related to angular momentum.

37. While this is true, angles themselves do not add like vectors. They are not a vector space in the linear algebra sense. This is the basic reason why inelegant rotation matrices are used to represent rotational displacements, while everything else in simple mechanics is generally a vector. Angles can be made to add somewhat like vectors using a generalization of the complex number called the quaternion.

4 Gyrocompasses

Although it is typically not a component of the inertial navigation system, the gyrocompass is an ingenious device based on many of the same principles. The gyrocompass has replaced the magnetic compass on virtually all large ships and it is not based on magnetism at all. The device can be used to implement an automatic helmsman on a ship by using it as a null indicator in a servo loop. The magnetic compass has two disadvantages compared to the gyrocompass. First, the earth's magnetic field varies locally and the poles themselves move slightly over time. Second, the presence of magnetic material on a vessel disrupts the operation of the magnetic compass. The gyrocompass is immune to these two problems.

It will be shown that the equation of motion of this device is that of a pendulum where the spin axis oscillates about the local north-south direction. Damping this oscillation causes the device's spin axis to eventually settle into north-south alignment. In this way, it functions as a compass. Several means of damping have been used, but one of the simplest and best methods is a eccentrically weighted semicircular ring developed by Elmer A. Sperry³⁶.

4.1 Equation of Motion

The theory of operation of the device can be derived from Euler's equation. The device is constructed from a rotor spinning at very high speed. The rotor is constrained so that its axis must stay in a plane which is horizontal at its current position on the earth. That is, the normal vector to the plane is the local vertical. This rotor is mounted inside a set of gimbals which permit rotation about the local vertical.

To derive the equation of motion, consider Figure 11. The right figure is a view from the equatorial plane of the earth spinning on its axis. The left view is a local overhead view of the device with its spin axis displaced from the local north-south direction.

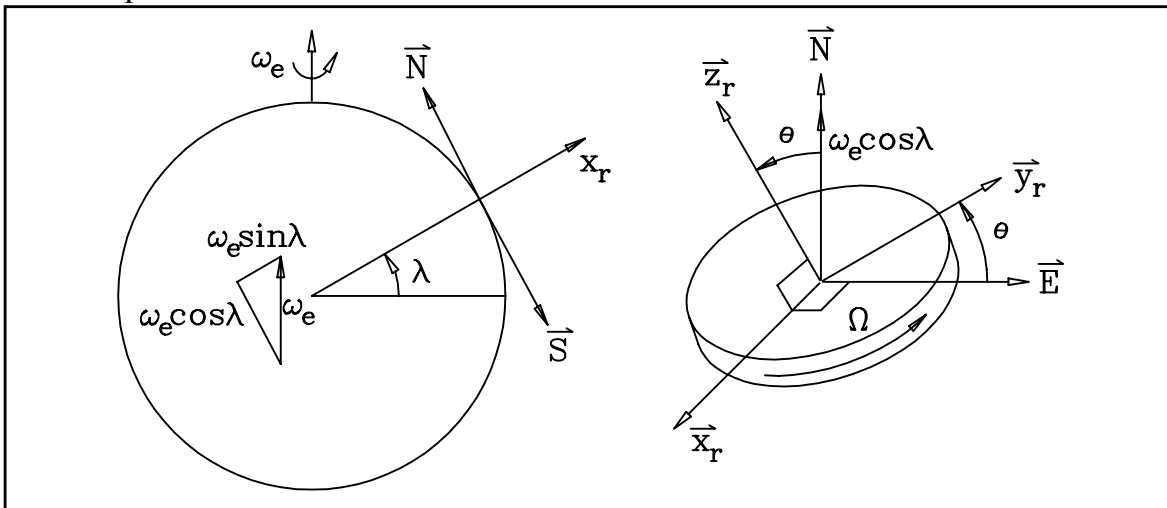


Figure 11 Gyrocompass

Let the device be positioned at some latitude λ on the rotating earth. There are three independent rotations happening simultaneously. The earth rotates about its axis, the device outer gimbal rotates

36. Of, you guessed it, Sperry Gyroscope Company.

3.2 Three Axis Stable Table

It should be clear that, because the accelerometers are mounted on the stable platform with their sensitive axes oriented along three orthogonal directions, the orientation of the stable platform defines the coordinate system in which the inertial specific force is measured. It is the function of the stable table to ensure that the accelerometers always point in the direction required by the calculations.

Practical stable tables always incorporate at least three gimbals and often more. Torque motors, called **torquers**, and rotary position sensors, called **signal generators** or **pickoffs** are mounted directly to the gimbals. For earth-fixed systems, the torquers are also used to rotate the table in order to maintain a locally level configuration.

Such tables directly implement the Euler angle model of rotation. This has two important consequences to the design of a three axis stable table. First, because the rotation axes of the gimbals are not always orthogonal, it is necessary to uncouple the control loops driving the gimbal torquers. Second, the gimbals are subject to the inherent singularity of such an arrangement. In the context of stable tables, the condition of singularity³⁵ is called **gimbal lock**, and it occurs when all three axes of rotation fall in the same plane. In this condition, the apparatus loses one of its degrees of freedom.

Gimbal lock can be avoided in two ways. For ships and similar vehicles, the required angular travel of the table with respect to the vehicle is small enough that the arrangement of gimbals can be designed to avoid gimbal lock. For guided missiles and other vehicles which require unconstrained rotation, a fourth gimbal can be added.

3.3 Complete INS

With a three axis stable table, the coordinate system problem is solved. The gyros can be mounted on the table in such a way as to provide orientation information around three orthogonal axes. The accelerometers are mounted on the same table in such a way that their sensitive axes are also oriented along three orthogonal axes. This alignment permits measurement of the specific force vector in three dimensions in an inertially stable coordinate system.

Another frequently used approach is to deliberately drive the platform to rotate as the earth does, so that an earth fixed coordinate system is used, and motion relative to the earth is measured. However, in this case, it is only the coordinate system which is earth referenced. The accelerometers will still register the coriolis and centrifugal forces because specific force is inherently an inertially derived vector, and no technology can change this fact. Simply changing coordinates cannot eliminate the effect of rotating at 1000 m.p.h. around the center of the earth. Here the importance of the distinction between reference frame and coordinate system is illustrated.

35. Singularity is a common concern in mechanisms. It can be defined in many ways: loss of a degree of freedom, mathematical singularity of the Jacobian matrix, co-linearity of axes or linear dependence of axes, and convergence of redundancy to a single point. All definitions are equivalent. Singularity is always a very serious matter since it implies loss of a degree of freedom for positioning and infinite rates for rate control. The Euler wrist commonly used in robot manipulators suffers from the same problems.

3 Stable Platforms

In the stable table or stable platform, a feedback loop is used to regulate the table in a such a way as to prevent the precession of the gyroscopes. This important principle is the basis of maintaining a stable angular reference in inertial systems.

3.1 Single Axis Stable Table

A simplified single axis stable platform is shown in Figure 10:

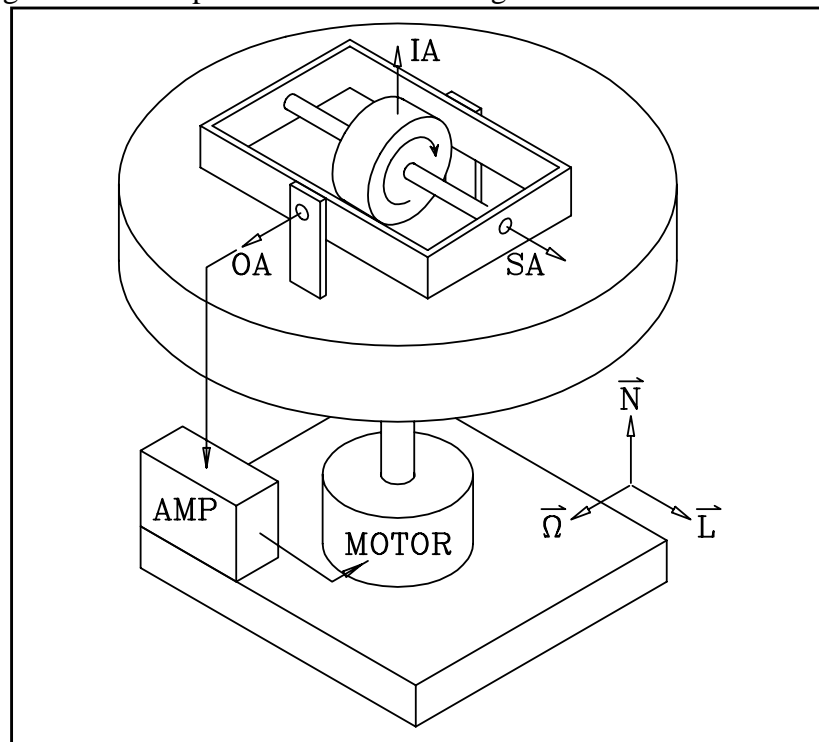


Figure 10 Single Axis Stable Platform

The figure uses the standard notation for the input (IA), output (OA), and spin axes (SA) of a gyroscope. The directions of input torque (\vec{N}), angular momentum (\vec{L}), and precession ($\vec{\Omega}$), are associated with these axes as shown in the figure.

An intuitive explanation of the operation of the platform is as follows. Suppose the rotation of the output axis is measured and used to drive an amplifier connected to a motor which is able to rotate the platform. If the vehicle to which the rectangular base is attached performs a yawing motion, this is a rotation about the input axis (IA) of the gyro. Initially, the motor will transmit this torque directly through to the circular table. This will yaw the table a small amount which will cause the circular table to rotate. This, in turn, will cause the gyro to nod about its output axis (OA). If this output signal is fed back to the motor, a suitable servo control law will drive the circular table in such a way as to null the gyro output.

The conclusion is that the platform will be forced to execute the correct motions to keep the gyro from nodding about its output axis. From the point of view of the vehicle, it performs a yawing motion, and the table stays behind. The circular platform stays where it is pointed in an inertial frame of reference.

2.3 Practical Gyroscopes

Gyros can be used as **rate gyros** or as **integrating gyros**. The former are characterized by spring restraint of the output axis and low damping and derive their name from the fact that the output angle is proportional to the input rate. These are useful as turn indicators, but are not used in precision inertial systems³³. The latter are characterized by no spring restraint and much higher damping of the output axis, and derive their name from the fact that the output angle is proportional to the integral of the input rate.

Typical requirements for a gyro are the measurement of one thousandth of a degree per hour of angular motion. The ability to manufacture precision gyroscopic instruments which are sufficiently free of unpredictable drift is the ultimate limiting factor in the performance of inertial navigation technology. Gyro drift³⁴ is the overriding performance concern and it is caused by parasitic precessional torques which are outside the system mechanization loops.

Gyroscopes develop imperfect performance for all of the following reasons:

- mass imbalance of the rotor
- thermal gradients
- anisoelectricity
- buoyancy imbalance (in floated devices)

Current technology relies on the following techniques to improve performance:

- flotation in fluid to buoy components and damp oscillations
- high incidence of cylindrical symmetry in components
- precise manufacturing tolerances
- matched coefficients of thermal expansion
- temperature control
- preloaded bearings

Some practical designs of gyroscopes include the floated gyro, originally developed at MIT, air bearing gyros, electrostatic gyros, and optical gyros. All but the latter derive their name from the technique used to isolate the rotary bearings from loads. This is because spin axis bearings are an important limiting factor in performance.

33. See, however, strapdown systems later.

34. The earth rotates 15 degrees per hour. Drift is often measured in milli earth's rate units or "merus" equal to 0.0015 degrees per hour.

2.2 Single Axis Gyroscope

Real gyros are built based on the second property of the spinning rotor, precession, and employ servomechanisms to implement a stable table³² on which the accelerometers are mounted. Consider Figure 9:

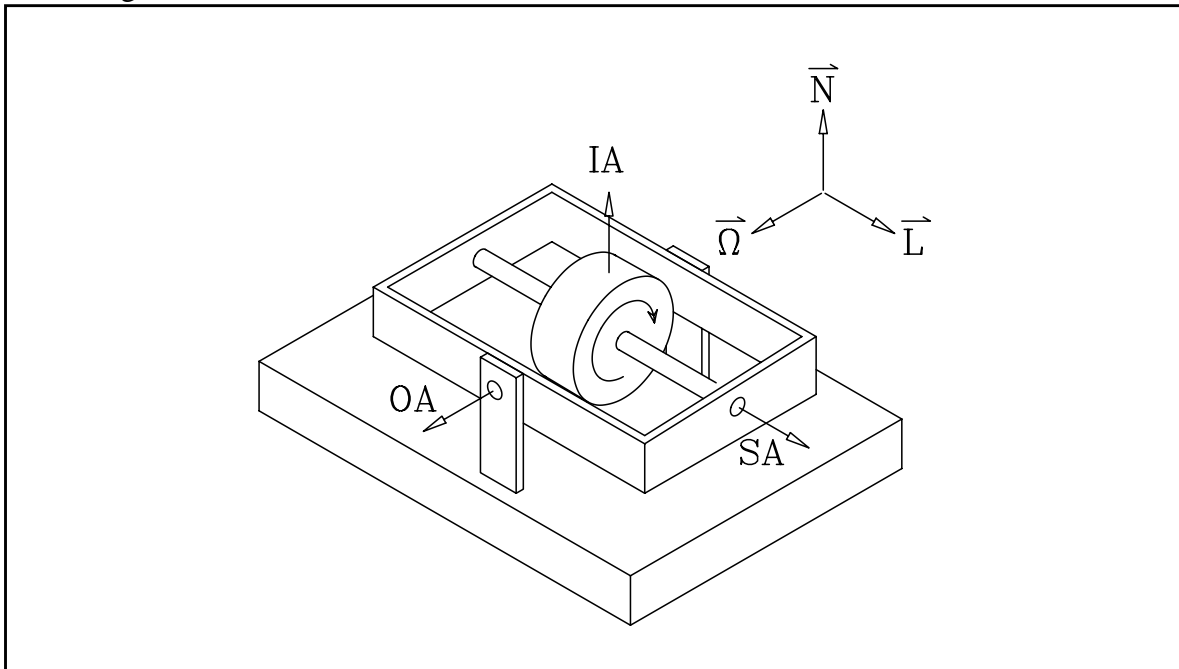


Figure 9 Single Axis Gyro Precession

Any attempt to yaw the platform on which the gyro is mounted about the input axis will give rise to a nodding of the device about the output axis. A rotational transducer such as a resolver placed on the output axis will register this rotation which is proportional to the applied torque. Using this technique, gyros can be employed as null indicating devices in the feedback loops of servomechanisms.

Single axis gyros such as this one are called **restrained gyros** since they are prevented from motion about all axes but one, and, in practice, the output angles are never allowed to become large. There are also two degree of freedom gyros in common use. These are called **free gyros**, incorporate an extra set of gimbals and are used to control angular velocity around two orthogonal axes. Gimbals may be used both in the gyro itself and in the stable table, so sometimes it is necessary to distinguish these.

32. A stable table is stable in orientation in a specific coordinate system.

2 Mechanical Gyroscopes

While the accelerometer measures translational motion, the **gyroscope**³¹ or **gyro** is a device used to measure or maintain orientation. Cartesian coordinate systems are defined by the orientation of their reference axes, so the gyroscope can be used to solve the coordinate system problem. The properties of the gyroscope are based on the spinning rotor which was first investigated by Euler and Foucault in the eighteenth century. The two primary properties of interest to inertial navigation are the **rigidity** of orientation in inertial space and **precession**.

2.1 Simple Gyroscope

A stable angular reference might be constructed based on the rigidity property of the gyro as follows. Use a symmetric body (disk or ring), spin it about its axis, and isolate it from external torques. The method used to isolate the spun mass from external torques is to use a system of hierarchical, independent, ideally frictionless pin joints called **gimbals** as illustrated in Figure 8.

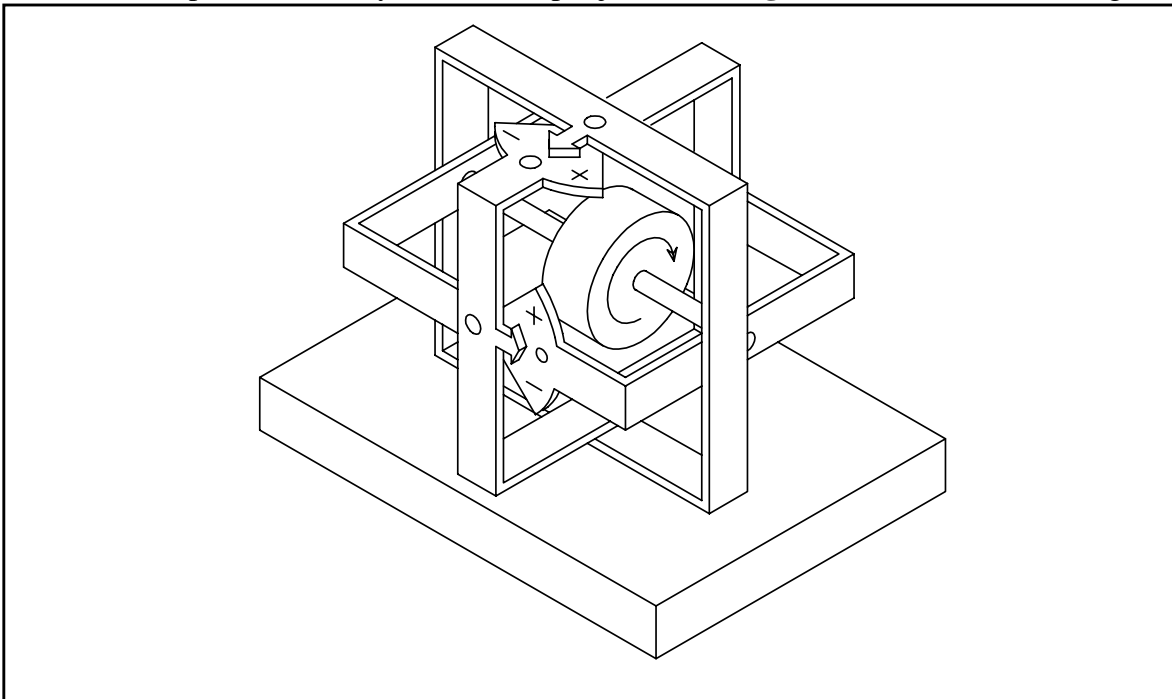


Figure 8 Mechanical Gyroscope

As the platform is rotated, the gyro will tend to remain stably pointed in inertial space. Sensors mounted on the gimbals, as indicated schematically in the figure, can measure the vehicle rotation in this way.

31. From the Greek gyros for “circle” or “rotation”, and skopein “to view”. Foucault used the gyroscope to view the rotation of the earth by pointing the device at a star.

1.5 Practical Accelerometers

All accelerometers operate similarly, though details of construction differ. A practical accelerometer is a precision instrument which couples a mass to the instrument case through an elastic, viscous, or electromagnetic restraint. Typically, the mass is allowed a single degree of freedom which may be linear or rotary. Rotary deflecting devices, called **pendulous** devices, are often used. Calibration of the restraint, whatever its form, provides a measure of the specific force that is tending to cause a deflection along the free degree of freedom.

Most devices operate in a closed loop, called a **rebalance loop**, which prevents the mass from actually moving. In this case, the effort required to retard motion is a measure of the specific force experienced. These devices are generally superior to those permitting error inducing motion. Most devices incorporate fluids to damp vibrations or to support critical components through buoyancy.

The design of inertial grade accelerometers is concerned with achieving:

- null repeatability, 5×10^{-4} g is typical
- low threshold (sensitivity), 10^{-4} g is typical
- linearity, 0.10% over 10 g is typical
- large dynamic range
- time constant, few milliseconds is typical
- low sensitivity to vibration

Accelerometers may be designed to be mounted on a stable platform. These incorporate their own small gyroscopes and perform integration as part of their function. The string accelerometer maintains mass between two strings under tension. Beat frequencies between the vibrating strings provide the output. There are many exotic types including, nuclear, electrostatic, cryogenic, gaseous, particle stream, solid state, and vibrating reed instruments.


gravitation is significant, as is the gravitation of the sun or moon or planets when operating close to them. However, on the earth, the sun and moon give rise to the tides, so they are not completely insignificant forces.

It may seem that these effects have already been discounted by noting that the earth is in free-fall with respect to them, but this simply established the earth an inertial frame. In any inertial frame, real forces arising from material bodies must be considered in the equations. The next step is to account for all those real forces, and gravitation of the sun and the moon and everything else in the universe are real forces. If they were not, the sun would not orbit the galactic center once every 200 million years or so.

Luckily, the acceleration due to the sun is about 0.0006 g and that of the moon is similarly small, so they can be neglected as engineering approximations when operating near the earth. The forces from all other matter in the galaxy, and of other galaxies can likewise be neglected, because, massive though they are, they are also very far away.

1.4 Basic Equation of Inertial Navigation

The frame of reference problem is finally solved in principle. The basic equation of inertial navigation is the specific force equation which relates the accelerometer readout to the gravitational force on the vehicle and its inertial acceleration. Consider a spring and mass accelerometer mounted to a vehicle chassis. The spring experiences forces T required to counteract gravity when the vehicle is stationary, and further forces as the vehicle executes accelerated motion dragging the mass along with it. Using a subscript of i to indicate a geocentric inertial frame, T to represent total force registered by the accelerometer, G to represent gravitational attraction, and m for the accelerometer test mass, Newton's second law gives:

total force	$T - G = ma_i$	
specific force	$\frac{T}{m} = a_i + \frac{G}{m}$	

It is important to avoid assuming a constant acceleration due to gravitation. The accelerometer measures the specific force on the vehicle expressed in an inertial frame in vehicle coordinates. Usually, the quantity of interest is the position on the surface of earth and hence in the rotating earth frame. Using the theorem of Coriolis:

$\frac{T}{m} = (a_e + 2\vec{\omega} \times \vec{v}_e + \vec{\omega} \times \vec{\omega} \times \vec{r}_e) + \frac{G}{m}$
--

This equation can be solved for the acceleration in the spinning earth frame a_e , and it can be integrated twice. The earth's sidereal rate of rotation is known. The position and velocity can be integrated forward from the initial conditions, and there are standard formulae for the earth's gravitation. In order to integrate, all accelerometer readings must be converted to a common coordinate system. In order to solve this coordinate system problem, the discussion will turn to the second inertial sensor, the gyroscope, in the next section.

gravitation will be distinguished.

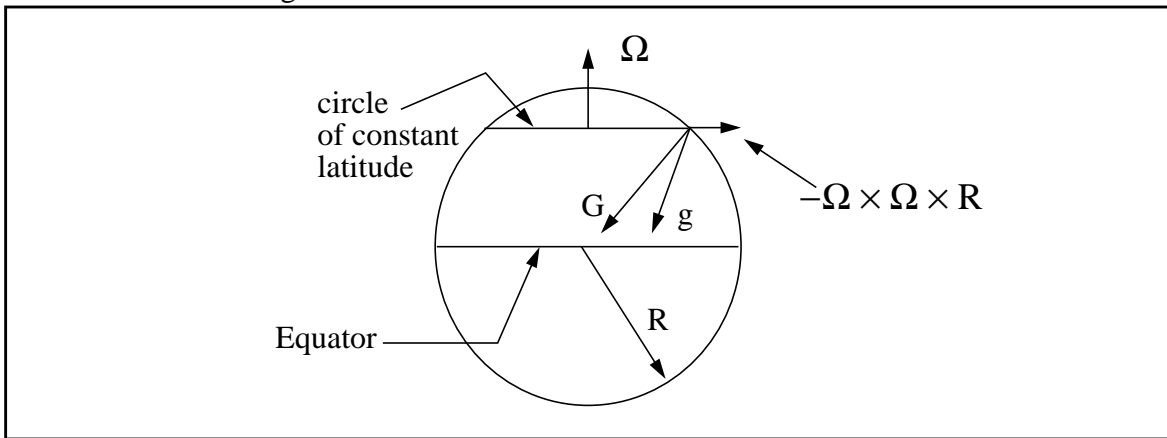


Figure 7 The Components of Gravity

1.2 Specific Force

Imagine an accelerometer constructed from a test mass on a spring. It is intuitively clear that such a device must measure the influence of gravitation because it will deflect with the weight of the mass. It is not enough to simply calibrate this constant out of the device, because, when mounted on a vehicle climbing through the atmosphere at constant speed, the mass would slowly rise, since gravitation varies with distance, compressing the spring and indicating an erroneous acceleration upward. This is yet another example of the principle of relativity, for an observer in such a vehicle cannot tell he is in a gravitational field. This is the fundamental reason why an INS can only operate in a known (or known to be insignificant) gravitational field, and why space vehicles using an INS must know the gravitational field of any planets they encounter to avoid missing the real accelerations that they cause.

Provided a vehicle always operates near the surface of the earth, the gravitational constant acceleration can in principle, be calibrated out. However, as a practical matter, real vehicles rotate so the direction of the gravitational attraction vector is not constant in the vehicle frame of reference. Some systems however, completely avoid the gravitation problem by maintaining the accelerometer sensitive axis normal to the local gravitation vector. The real issue is that the deflection of the mass is directly influenced by the total force on the mass and not by its acceleration. When stationary in a gravitational field, a mass can experience force even when there is no acceleration. When free-falling, a mass can register no force even when there is an acceleration. A quantity called the **specific force** is defined as the total force divided by the mass.

1.3 Effect of the Earth's Gravitation

It was shown in the last chapter that a nonrotating geocentric frame of reference is inertial. This allows the expression of Newton's laws in such a frame. However, it is now clear that choosing an inertial frame is only half the problem. It is also necessary to know the value of the gravitation vector field in which the device operates. In general, this field is determined by the superposition of all influences of all bodies in the universe. This can be approximated easily by considering only those bodies which exert significant gravitational force on the test mass.

Gravitational force is proportional to the mass on which it is exerted, so it is completely equivalent to speak of an acceleration due to gravitation. It is clear that when operating near the earth, its

PART II: Inertial Sensors and Stable Platforms

1. Accelerometers

The accelerometer is the basic sensor of inertial navigation. This section will examine its theoretical operation, its limitations, and its practical implementation.

1.1 Gravity

Gravity is defined as the force per unit mass required to keep a test mass in the same position relative to the earth. It has a nominal value of 32.2 ft./sec/sec or 9.8 meters/sec/sec. The practice of measuring and modelling gravity is called **geodesy**. In the last section, it was shown that an object fixed to the surface of the earth experiences a centrifugal force as viewed from a reference frame spinning with the earth. A plumb bob at the surface of the earth does not point toward its center but is rather displaced slightly toward the equator.

The natural trajectory for an object positioned at a radius of R , given some initial tangential velocity, is in an orbit around the planet. Its real trajectory is not a circle centered at the center of the earth (an orbit) but a circle of constant latitude²⁹. Some force other than gravitation, imparted by contact with the surface of the earth must be the cause of this.

It is the practice in geodesy to combine both gravitational attraction and centrifugal force together into a single force, divided by the mass, and call the result **gravity**. Local density changes in the earth give rise to local changes in the gravitational attraction³⁰, and it also varies obviously with elevation. The centrifugal force varies with latitude as shown in Figure 7. Henceforth, gravity and

29. This is why “geostationary”, or “geosynchronous” orbits must be equatorial and why the satellite pictures in your nightly newscast are always shown from above the equator. Geosynchronous orbit over London or New York is impossible without great expenditure of energy, despite the popular science fiction.

30. High density meteorites can be detected in this way.

4.2.2 Coriolis Force

The conclusion, which makes perfect physical sense, defies intuition. The observer is not in an inertial frame and observes paradoxical motion if he assumes that he is. Often, the assumption is mostly valid because the effect is relatively small.

From the point of view of the observer in his rotating frame, there is some apparent force which has caused the object to drift to the east. This force is called the Coriolis force²⁶. It is one of two such apparent forces which arise when a reference frame has an angular velocity with respect to an inertial frame. The other is called centrifugal.

For an accelerometer in an INS, which measures the inertial acceleration²⁷, the Euler apparent force is zero because the earth rotates at constant rate²⁸. The geocentric and spinning frames are associated with the same object (the earth) so they have no translational interframe position, velocity, or acceleration. The Einstein apparent force is zero. Rewriting the earlier acceleration relationship:

$$\vec{a}_f = \vec{a}_m + 2\vec{\omega} \times \vec{v}_m + \vec{\omega} \times \vec{\omega} \times \vec{r}_m$$

So, even when the vehicle is fixed on the earth, since r_m is measured from the center of the earth, the vehicle sees a centrifugal acceleration, and when it moves, it also sees a coriolis acceleration. Accounting for these apparent forces is a central issue in inertial navigation.

26. Also called sideways force. Visit the Foucault pendulum at the UN building in New York to see a demonstration of it in action.

27. Not really, see next section.

28. Ignoring precession.

4.2.1 Tangential Velocity

Establish a tangential/radial cartesian coordinate system to be oriented according to the direction of motion of the object at the instant it is released. The radius of the earth is about 4000 miles, and it rotates about once per 24 hour day. Therefore, at the equator, the tangential velocity as it spins on its axis is:

$$v_t = R_e \times \Omega_e$$
$$v_t = 4 \times 10^3 \text{ miles} \times \frac{2\pi}{24\text{hours}} \approx 1000\text{mph}$$

Given that it does not fall straight, it may seem that once the observer releases the object, it will remain stationary in the inertial frame along the tangential direction, and accelerate along the radial direction, so the planet will continue to rotate around it until it reaches the ground somewhere “behind” the observer.²⁵ It actually falls in front of the observer²⁵.

The reason is that it remains *unaccelerated* and not stationary in the tangential direction. So it continues to translate along the tangential axis at its original considerable speed. Now, because of its greater distance from the center of the earth, that speed is in fact slightly larger than that of the base of the observer’s tower, which is fixed to the ground. The object, no longer constrained by the observer’s hand, moves at a velocity determined by its slightly larger radius, and hence moves faster in the tangential direction. In the radial direction, it is accelerated, so the entire trajectory is ballistic in the inertial frame. The situation is depicted in Figure 6.

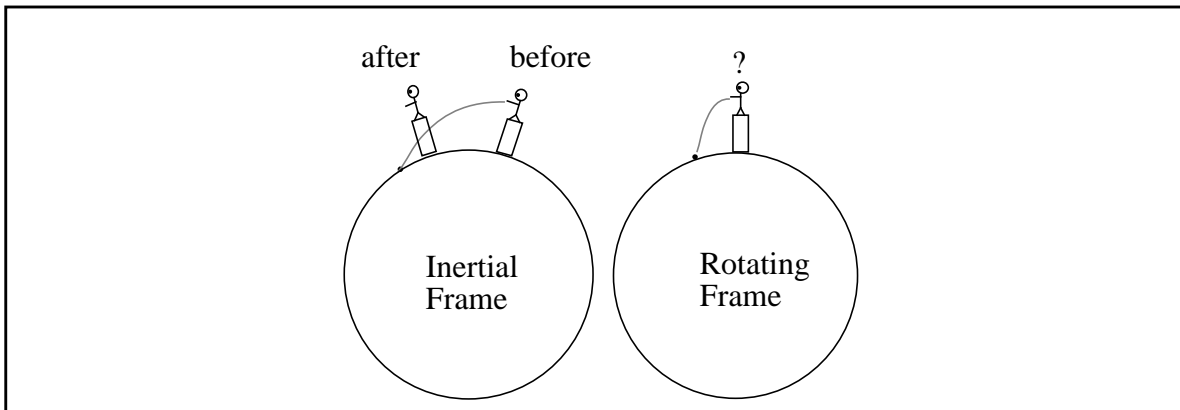


Figure 6 Galileo’s Experiment Revisited

25. The Tower of Pisa is 180 ft. high. As a result, ignoring air friction, all of Galileo’s masses would have fallen about 0.3 inches to the east of the point directly below where he dropped them. The Empire State Building is 1200 ft. high, giving rise to a 6 inch eastward deflection.

4.1.4 Motion in the Galaxy

The sun orbits the center of the galaxy at a period of one revolution every 200 million years with a tangential velocity of roughly 30,000 m.p.h.²³. This is gravitational free-fall.

4.1.5 Motion in the Cosmos

According to astronomical observations, all galaxies are flying apart at incredible speeds. Although the speeds are so remarkable that relativistic red shifts²⁴ of starlight are the rule, the accelerations are completely determined by gravitation. This is gravitational free-fall.

4.1.6 Motion of the Cosmos

To be complete, it is necessary to ask the question of whether the center of mass of all the universe can be in motion. If it can, with respect to what would it be in motion? There is nothing else by definition. Physicists have pondered this question deeply, and some believe that the paradox is tied to the existence of inertial frames. They believe that the existence of inertial frames is the only way to resolve the paradox. Hence the mass center of the cosmos is the ultimate inertial frame and the analysis is complete.

In summary, this section has shown that the only motion of the earth that disqualifies it as an inertial frame is its rotation on its axis. Frames fixed to the earth, the sun, the galactic center, and the center of the cosmos are all inertial. Looking as far as the galactic center, and ignoring the earth's spin, observers on the earth are in an elevator, in an elevator, in an elevator, and all three of these elevators are inertial frames because they are in gravitational free-fall with respect to each other.

It is not the case that a nonrotating earth frame is inertial because the acceleration of the earth with respect to the sun is very small. The qualification as inertial is independent of the magnitude of the acceleration. The earth could be spinning around the sun 1000 times a second at a radius of one meter and it would still be an inertial frame if it did not rotate on its own axis. On such a planet, all stars but the sun would be fixed in the sky.

4.2 Effect of the Earth's Rotation

In order to investigate the effect of the earth's rotation on an accelerometer, it will first be shown that objects do not, in fact, fall in straight lines on earth. Suppose an observer attempts to reproduce Galileo's original experiment by dropping masses from the top of a tower to observe their motion. Assume the observer is positioned on the equator and is facing east.

It has been shown that the object is attracted by the sun just as the earth is, so that does not enter the picture. It is not necessary to consider what happens from a sun fixed viewpoint, or from a galaxy fixed viewpoint. A nonrotating geocentric frame of reference is inertial.

In this inertial frame, and not in the rotating earth frame, the released object will move under the influence of all external forces arising from interactions with material bodies according to Newton's second law. A careful exploration of the consequences of this reveals some very startling results.

23. At 5 billion years old, the sun has only orbited the galactic center 25 times since it formed.

24. This is the doppler shift of the frequency or color of light due to relative motion of emitter and receiver. Although the speed of light in vacuum is constant to all observers, its color is not. These red shifts were an important experimental verification of the fact that the universe is expanding.

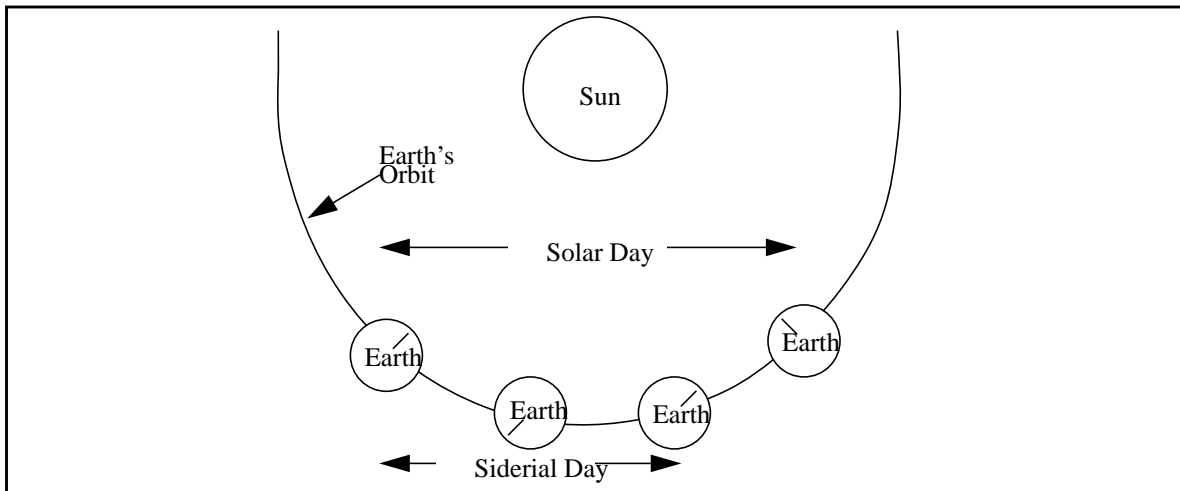


Figure 5 Sidereal Day

Another subtlety is the change in the earth's orbital speed about the sun due to its elliptical shape. Figure 5 demonstrates that the solar day varies in length for this reason. The familiar 24 hour day is the **mean solar day**. Of course, the amount of daylight per day varies with the season as a consequence of the tilt of the earth's axis at roughly 22 degrees¹⁹ to the ecliptic²⁰, but the current issue is the day when it is defined as a complete rotation of the earth.

4.1.2 Annual Revolution About the Sun

While the earth's velocity varies with the season of the year, ignoring daily rotation, this acceleration is a translational free-fall because it is completely caused by the gravitational attraction of the sun on the mass center of the earth. The earth is also in gravitational free-fall with respect to all other bodies in the universe, so a geocentric earth frame which does not rotate about the earth's axis must be inertial by the principle of relativity.

4.1.3 Precession

However, there is one subtlety. The earth is not perfectly round, so the mass attraction of the moon, sun, and planets on the bulge of the equator cause precession²¹ of the axis. These attractions try to pull the earth's axis perpendicular to the ecliptic, and they actually cause precession because of the earth's rotational motion. Precession amounts to small changes in the angular velocity vector of the earth about its own axis, and as was shown earlier, gives rise to an effect called Euler apparent force. Luckily, the precessional rate is 360 degrees every 26000 years²² and the effect is so small as to be undetectable. The discussion could stop here since an inertial frame has been found, but it will continue for reasons that will be clear later.

19. This is why the tropics of cancer and capricorn are 22 degrees from the equator.

20. The plane of the earth's orbit about the sun.

21. The so-called precession of the equinoxes was apparently discovered by the Babylonian astronomer Cidenas in about 343 BC.

22. The ice ages are tied to this cycle. They are as predictable as the seasons of the year. Buy boots before the winter of 24000 AD.

4 Motion on the Earth

The earth is not an inertial frame because objects do not fall to it under the influence of its gravitation according to Newton's laws. This section will investigate why this is so in more detail.

4.1 Earth as an Inertial Frame

The motion of the earth through space is a complicated superposition of several component motions. The most important of these are:

- daily rotations about its own axis
- annual revolution about the sun¹⁶
- precession and nutation of the axis¹⁷
- motion of the sun in the galaxy
- motion of the galaxy in the cosmos

This section will investigate the reasons why the first of these is the only one that is important to the problem of finding an inertial frame.

4.1.1 Daily Rotation

This is, of course, a rotation of 360 degrees per day. However, there is a subtlety of importance. The solar day must be distinguished from the **sidereal day**. The first is a complete rotation with respect to the sun and is 24 hours¹⁸ long. The second is a complete rotation with respect to the stars and is about 4 minutes shorter. This is a consequence of the fact that, if the earth did not rotate on its axis, the sun would still rise and fall once per year as the earth orbits it, so the sun goes around the earth slightly faster than the stars do as shown in Figure 5. If the stars are an inertial frame, it is the sidereal rotation of the earth that matters. Also, since the earth-fixed frame rotates with respect to the stars at a considerable rate, it cannot be inertial. The extent of this effect will be calculated shortly.

16. As you read this you are travelling through space at 70,000 m.p.h. with respect to the sun. Hold on.

17. The constellations of the zodiac are those constellations which appear on the horizon at sunrise. Ancient Egyptians noticed a correlation between the constellations and the Nile flood and assumed there was a causal relationship - hence astrology was born. Due to the annual circuit about the sun, these constellations appear to circle the horizon once a year. Due to precession, the zodiac also rotates once every 26,000 years, so, according to standard tables, everybody's sign is now off by one constellation and things won't be right again until the year 24000. Astrologers, naturally, don't emphasize this.

18. The hour is defined so that this is so.

perpendicular to both the spin axis and the applied torque.

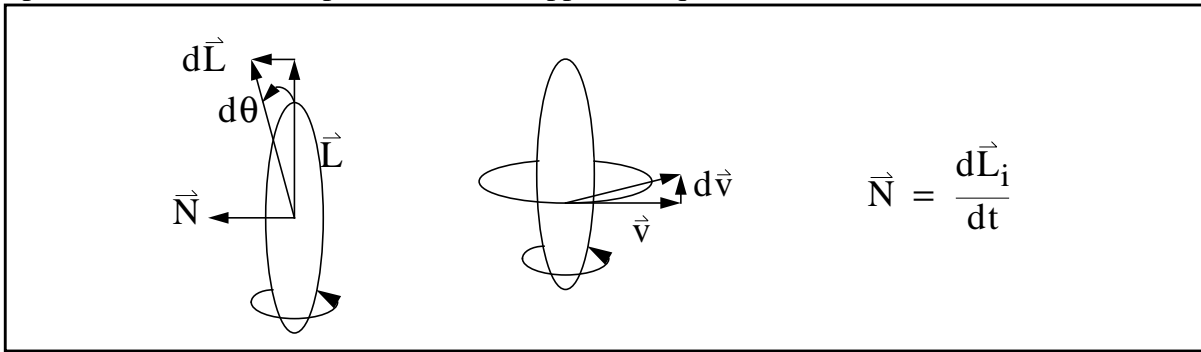


Figure 4 Precession

If the angle between the new and old angular acceleration vectors is defined as $d\theta$, and its rate is defined as Ω , then the gyroscope rigidity equation is:

$$d\theta = \frac{|d\vec{L}|}{|\vec{L}|} = \frac{|\vec{N}|dt}{|\vec{L}|}$$

$$\Omega = \frac{d\theta}{dt} = \frac{|\vec{N}|}{|\vec{L}|}$$

Remember that Ω is the rate of change of the angular velocity vector, not the rotation of the body. Considering the orthogonality of the precession, this equation is often expressed as:

$$\hat{N} = \hat{\Omega} \times \hat{L}$$

where only the direction of the vectors is considered. So the rate of change of the angular velocity is proportional to the applied torque, and inversely proportional to the angular velocity. This is the justification for the experimental observation that the faster a rotor spins, the harder it is to disturb it. This technique is called **spin stabilization** and it is commonly used for communication satellites.

conclusion is that *the unforced symmetric body has an angular velocity which describes a cone in space*. Now the obvious question is which observer sees it this way. The angular velocity in the formulation is the angular velocity of the fixed body frame with respect to inertial space. This quantity was expressed in the principle body coordinate system.

It is tempting to say that the body fixed observer sees the conical motion. However, remember that the angular acceleration vector is inertially referenced and only *represented* in body coordinates. In an inertial frame, for torque free motion, the angular momentum is constant, and the kinetic energy is constant. The latter is given by:

$$T = \frac{\vec{\omega} \cdot \vec{L}}{2}$$

Thus, the angular momentum vector and the angular velocity vector make a fixed angle, and the angular momentum itself is constant. Hence, the velocity vector rotates around the fixed momentum vector while maintaining a constant angle with it. To the inertial observer, the object spins about an axis which is, in general, not the symmetry axis, and the axis itself spins in a cone. This conical motion of the spin axis is called free **precession**^{13 14}.

3.4 Forced Precession

Consider now what happens when a rotating symmetric body is subjected to a torque. It is clear from Euler’s equation that when the applied torque is parallel to the angular momentum, the body experiences an angular acceleration which changes the magnitude of the angular velocity but not its direction. However, that is where the analogy to the translational case ends. Strange as it may seem, the equation implies that the response of the rotor to a torque applied normal to the spin axis is to rotate in a direction normal to both the spin axis and the applied torque. This phenomenon is also called precession¹⁵.

That is, if one attempts to rotate the spin axis, it does not rotate in the direction that the torque is trying to force it to go. Consider Figure 4. The vector algebra makes it seem obvious, but the reality seems counterintuitive, because the body will tilt into the torque. It can be understood intuitively by considering the incremental velocities applied to particles on the equator of the body. The torque imparts an incremental upward velocity to the particle shown in the right side figure, and an incremental downward velocity on its twin behind the body.

When added to its current velocity, the net result is a rotation of the velocity vector about an axis

13. This phenomenon is related to the rotation of the earth’s spin axis, which is also called precession. Clearly the earth itself is a spinning rotor.

14. Ancient astronomers noticed that the constellation in which the sun rises on the vernal equinox advances or “precedes”. Hence the term.

15. There is much confusion about the terms precession and nutation in the literature. Some authors call the free precession described above “nutation”, reserving precession for the gyroscopic response to torques. Others define precession as above and reserve nutation for small variations of the shape of the cone swept out by the spin axis. This is the sense of nutation as used with respect to the earth’s motion. Others reserve nutation for the nodding of the spin axis of a top. Still others define three rates of Euler angles called precession, nutation, and spin. The author has been unable to find a strict equivalence in these uses of the terms.

Now, consider expressing these equations in the coordinate system in which the inertia matrix is diagonal, called the principal axes¹²:

$$\begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} = \begin{bmatrix} I_x \dot{\omega}_x \\ I_y \dot{\omega}_y \\ I_z \dot{\omega}_z \end{bmatrix} + \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \times \begin{bmatrix} I_x \omega_x \\ I_y \omega_y \\ I_z \omega_z \end{bmatrix}$$

Writing out the cross product, the familiar scalar form of Euler's equations, in principal coordinates is obtained:

$$\begin{aligned} N_x &= I_x \dot{\omega}_x + (I_z - I_y) \omega_y \omega_z \\ N_y &= I_y \dot{\omega}_y + (I_x - I_z) \omega_x \omega_z \\ N_z &= I_z \dot{\omega}_z + (I_y - I_x) \omega_x \omega_y \end{aligned}$$

Now, for a line symmetric body, the z axis can be arbitrarily identified as the symmetry axis. Then let $I_{xy} = I_x = I_y$. The N_z term and the last term in the last equation vanish implying that ω_z is constant. Rewriting by dividing by I_{xy} :

$$\begin{aligned} \dot{\omega}_x + \left(\frac{(I_z - I_{xy})}{I_{xy}} \omega_z \right) \omega_y &= 0 \\ \dot{\omega}_y + \left(\frac{(I_{xy} - I_z)}{I_{xy}} \omega_z \right) \omega_x &= 0 \\ \dot{\omega}_z &= 0 \end{aligned}$$

Differentiation of the first two equations yields the general differential equations for free rotation of a symmetric body:

$$\begin{aligned} \dot{\omega}_x + \Omega^2 \omega_x &= 0 \\ \dot{\omega}_y + \Omega^2 \omega_y &= 0 \\ \dot{\omega}_z &= 0 \end{aligned} \quad \Omega = \left(\frac{(I_{xy} - I_z)}{I_{xy}} \omega_z \right)$$

It is easy to verify that the first two equations together describe a circle, and the last a constant. The

12. By a theorem of linear algebra, all bodies of any shape possess principle axes.

3.2 Forced Rotation of an Arbitrary Body

Another way to achieve pure spinning is to constrain all other motion physically. A body whose angular acceleration is constrained in this way must still follow Newton's laws. Consider such a body mounted in such a way that it is free to spin on one axis only. It is clear that such a body will have a constant direction of angular velocity. However, if the spin axis does not coincide with an axis of symmetry, then the inertia tensor will not be constant. Hence the net torque on the body will not be constant. These changing torques must be supplied by the constraint bearings which will therefore undergo frequent oscillatory loading¹⁰.

3.3 Free Precession of a Symmetric Body

The next to previous section had assumed that axial spin was the case and shown that it was consistent with the equation of motion. It turns out that there are other motions possible for a spinning rotor. To get these, consider again Euler's equation:

$$\vec{N} = \frac{d\vec{L}_i}{dt}$$

Identifying the fixed frame with the inertial frame (subscript i), and the moving frame with a frame fixed to the body (subscript b), the theorem of Coriolis can be applied. This gives after some manipulation:

$$\begin{aligned}\vec{N} &= \left(\frac{d\vec{L}_i}{dt} \right)_b + \vec{\omega} \times \vec{L}_i \\ \vec{N} &= \left(\frac{d(I_i \vec{\omega}_i)}{dt} \right)_b + \vec{\omega} \times I_i \vec{\omega} \\ \vec{N} &= I_i \left(\frac{d\vec{\omega}_i}{dt} \right)_b + \vec{\omega} \times I_i \vec{\omega}\end{aligned}$$

This result is an expression of Euler's equation in a very subtle form. It is important to be very careful about what the above symbols mean. The first term on the right hand side is the body fixed observer's derivative of the inertial observer's angular momentum¹¹. The first ω in the cross product is the relative rotation of the fixed and rotating frames. The second ω is the rotation of the body in an inertial frame. Only when the fixed frame is inertial and the second is body fixed do these two angular velocities become equal.

10. This is why the wheels of cars are balanced, and why it is done by adding small weights in strategic places.

11. Yes, that sounds like a very strange thing to be evaluating. But this is exactly what the theorem of Coriolis is for.

3 Rotational Dynamics

This section will investigate the properties of the spinning rotor, a body which exhibits axial symmetry and spins about that axis. These ideas form the basis of the operation of the gyroscope.

3.1 Free Rotation of An Arbitrary Body

Consider Newton's second law as applied to a purely rotating mass, and expressed in an inertial frame (i subscript) in the angular momentum form. Often called Euler's equation, it applies only with all quantities referred to an inertial frame of reference:

$\vec{N} = \frac{d\vec{L}_i}{dt} = \frac{d(I_i\vec{\omega})}{dt}$			
Net Torque	\vec{N}	Angular Momentum	\vec{L}
Inertia Tensor	I	Angular Velocity	$\vec{\omega}$

Under the assumption that a body experiences no net torques, one would expect, in analogy to translational dynamics, that the motion is one of constant angular velocity. This is not the case. The equation becomes on integration:

$\vec{L}_i = I_i\vec{\omega} = \text{const}$
--

Again this only holds in an inertial frame, which is a problem, because, for a body of arbitrary shape, the inertia tensor of the rotating body will not be constant, even under a constant angular velocity. If the inertia tensor is not constant, the angular velocity vector must also change in order for the product, the angular momentum, to remain constant. The conclusion is that the general torque free motion of a rigid body is a tumbling motion. This is why arbitrarily shaped objects tumble - even in space - instead of spinning neatly.

In the very special case of a line symmetric body, rotating about its axis of symmetry, the inertia tensor is constant in inertial space⁹. Under these conditions, it can be concluded that:

$\vec{\omega} = \text{const}$

It has been shown that a body under no net applied torque, which is rotating around its axis of symmetry, will remain rotating about its line of symmetry. This property is called **rigidity** of the spin axis. An attendant issue is stability. It can be shown that, provided the axis of symmetry is the axis of smallest inertia, the motion is stable to small perturbations.

9. Assuming, of course, uniform density.

$\vec{a}_f = \vec{a}_m + \frac{d^2 \vec{p}}{dt^2} + \frac{d\vec{\omega}}{dt} \times \vec{r}_m + 2\vec{\omega} \times \vec{v}_m + \vec{\omega} \times \vec{\omega} \times \vec{r}_m$			
acceleration seen in fixed frame	\vec{a}_f	Euler acceleration	$\frac{d\vec{\omega}}{dt} \times \vec{r}_m$
acceleration seen in moving frame	\vec{a}_m	Coriolis acceleration	$2\vec{\omega} \times \vec{v}_m$
Einstein acceleration	$\frac{d^2 \vec{p}}{dt^2}$	Centripetal acceleration	$\vec{\omega} \times \vec{\omega} \times \vec{r}_m$

acceleration. The Centripetal acceleration is also called Centrifugal. It is due to the angular velocity of the moving frame and the offset position of the object in the moving frame.

In order to see which observer sees the extra accelerations, identify the fixed frame with an inertial frame, and the moving frame with a spinning frame. The inertial observer is the only one who observes motion according to Newton's second law, so his measure of the acceleration is the one that must be used in that law. Assume that there are no forces on the object, so that its inertial acceleration is zero, and that its velocity is zero. If this is so, then the fixed observer sees a stationary object and the moving observer observes all of these extra terms of acceleration, even though there are no forces on the object. He would naturally assume that there are forces causing these accelerations. These forces are called **apparent forces**. Sometimes they are referred to as **inertial forces** - which can lead to confusion, because they are not measured in an inertial frame. They are caused by the inertial properties of matter as seen by an observer in a non-inertial frame, and not by the interaction of other bodies.

In an inertial frame, the only forces on a body required in Newton's laws are *those arising from the interaction of other material bodies*, namely those caused by contact (which is essentially electrostatic), electricity, magnetism, or gravitation. If this is so, it is natural to ask why it is that gravitation seems to have a distinguished role among the fundamental forces of nature. For Einstein did not say that frames in free-fall under the influence of an electric or magnetic field are inertial.

The simple answer is that Newton's laws say so, but this does not address the why of the question. According to Newton's laws, even though the forces on a body may result from the charged particles that it contains, another body in contact with it, the current flowing through it, or the gravitational field that attracts it, it is the mass of the body that determines its motion. There are laws of magnetic, electric, and gravitational interaction for determining forces, but there is only one law for determining accelerations, and that law uses the mass, not the electric charge or current.

Consider an electron in a television set that is being bent from its natural trajectory in order to have it hit the correct place on the screen. It is the **charge** of the electron that determines how much **force** can be exerted on it by an applied voltage. It is the **mass** of the electron which determines how much **acceleration** it will experience to bend it before it hits the screen.

- writing the vector \mathbf{v} in terms of the unit vectors of both frames
- using the chain rule of differentiation noting that the unit vectors of the moving frame are time varying

This formula will be called the **theorem of Coriolis**⁸. It is imperative to keep the reference frames straight when using this equation. The equation is used to relate time derivatives of the same vector computed by two observers in relative rotational motion. The vector is the *same* in all its instances in the equation. The derivatives, however, are different. The vector may be the position, velocity, acceleration, force, field strength, or anything else of anything with respect to anything else. If there are physical transformation laws associated with the translational motion of the two frames, these are considered outside this equation.

The intuitive meaning is that the derivative computed by the moving observer is the derivative seen by the stationary observer plus the component caused by his own rotational motion. The angular velocity is *the angular velocity of the moving frame with respect to the fixed one*, and the last instance of the vector in the equation is the original vector again.

The first step is to differentiate the position relationship in the fixed frame. This step is a straightforward vector time derivative which has nothing to do with the above derivative expression:

$$\left(\frac{d\mathbf{r}}{dt}\right)_f = \left(\frac{d\mathbf{r}}{dt}\right)_m + \left(\frac{d\mathbf{p}}{dt}\right)_f$$

Now the first term of the right hand side can be written using the Coriolis theorem as:

$$\left(\frac{d\mathbf{r}}{dt}\right)_f = \left(\frac{d\mathbf{r}}{dt}\right)_m + \vec{\omega} \times \mathbf{r}_m$$

This gives the velocity relation as:

$$\vec{v}_f = \vec{v}_m + \frac{d\mathbf{p}}{dt} + \vec{\omega} \times \mathbf{r}_m$$

The fixed observer sees an extra two components in addition to the velocity seen by the moving observer. The second term is due to the relative linear velocity of the moving frame. The third term is the motion of the vector due to the angular velocity of the moving frame. If the fixed observer observes his position vector to be fixed, the moving observer will observe circular motion due to the motion of the moving frame. Application of the theorem to the velocity relation generates:

There are four extra components of acceleration in the equation. The Einstein acceleration is due to the linear acceleration of the moving frame. It is the same as the apparent gravity sensed by the observer in the elevator. The Euler acceleration is due to the angular acceleration of the moving frame.

The Coriolis acceleration is due to the angular velocity of the moving frame and the velocity of the object as seen in the moving frame. Objects do not fall in straight lines on the earth because of this

8. Although the term applies strictly only to the acceleration formula, it will be used to describe any transformation of a vector's derivative.

The fact that astronauts training in a free-falling aircraft experience weightlessness is not an aspect of human subjectivity. Even with the best instruments they cannot tell the difference, and since in physics, things that cannot be measured do not exist, they are indeed, in a very deep sense, actually weightless. With this final idea, it is possible to find the inertial frame necessary for expressing the equations of inertial navigation.

2.3 Theorem of Coriolis

It is clear now that accelerometers on an INS that is stationary with respect to the surface of the earth will register an acceleration due to the rotation of the system with the earth. In order to investigate this more fully, the amount of this acceleration will be quantified.

Consider two frames of reference that are rotating with respect to each other at some instantaneous angular velocity ω , as measured in the first frame, and that have an instantaneous relative position of p as measured in the first frame. The first one will be called *fixed* and the second *moving* - though this choice is completely arbitrary. Quantities measured in the first frame will be indicated by an f subscript and those measured in the second will be indicated by an m subscript. Suppose an observer measures the position (r), velocity (v), and acceleration (a) of an object in the moving frame, and it is necessary to know what a second observer would measure in the fixed frame for the motion of the same object. The situation is depicted in Figure 3.

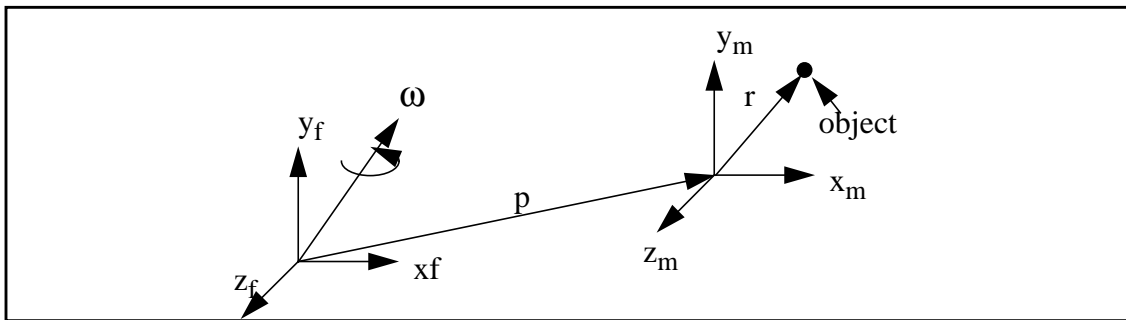


Figure 3 General Moving Reference Frames

Clearly, the positions add by simple vector addition:

$$\vec{r}_f = \vec{r}_m + \vec{p}$$

In order to compute the time derivative a very important general formula can be used to relate moving reference frames, *as applied to any vector v*. It is a tedious but straightforward exercise to show that:

$$\left(\frac{d\vec{v}}{dt}\right)_{\text{fixed}} = \left(\frac{d\vec{v}}{dt}\right)_{\text{moving}} + \vec{\omega} \times \vec{v}$$

The derivation proceeds by:

- assigning coordinate systems to the fixed and moving frames

7. That is, he cannot look out a window.

2.2 Principle of Relativity

A fundamental axiom of physics is the **principle of relativity**. This principle originally held that the laws of mechanics are the same in all of Newtons “absolute” frames. Newton meant it to apply only to mechanics:

“The motions of bodies included in a given space are the same among themselves whether that space is at rest or moves uniformly forward in a straight line.”

Another often used way of saying this is that all absolute frames are **equivalent** or that the laws themselves are **invariant**. Einstein called this the restricted sense of the principle. In his words:

“Even though classical mechanics does not supply us with a sufficiently broad basis for the theoretical presentation of all physical phenomena, still we must grant it a considerable measure of truth since it supplies us with the actual motions of the heavenly bodies with a delicacy of detail little short of wonderful. The principle of relativity (in the restricted sense) must therefore apply with great accuracy in the domain of mechanics.”

He was getting at the fact that it apparently does not apply to electromagnetic phenomena. Carefully examination of the word apparently led to Special Relativity. Careful examination of the requirement for absolute frames led to General Relativity. In his General Relativity theory, Einstein asserts that:

“we wish to understand by the general principle of relativity the following statement: all bodies of reference are equivalent for the description of natural phenomena, whatever may be their state of motion”

Here, he is getting at the fact that the principle holds in inertial, not absolute frames. For the present purpose, the point is that inertial and gravitational mass are equivalent, and therefore frames in gravitational free-fall are inertial.

Einstein argued this with a thought experiment consisting of a man in an elevator out in space and some other agent pulling with constant acceleration “up” on the entire elevator with the man in it as shown in Figure 2.

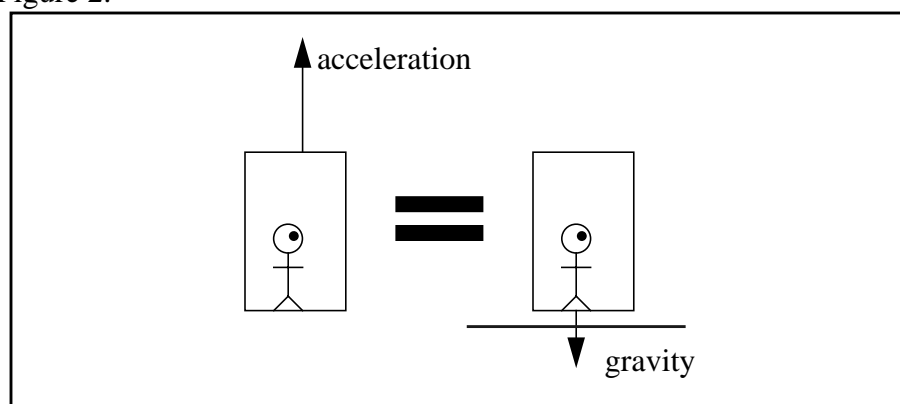


Figure 2 Einstein’s Elevator Thought Experiment

From the point of view of the man in the elevator, he seems to be experiencing a gravitational attraction pulling him toward the bottom of the elevator. He feels that he has some weight. The principle says that *no experiment, conducted in the frame of reference of the man in the elevator⁷, can tell the difference.*

2 Frames in Relative Motion

Accelerometers intrinsically measure with respect to an inertial frame of reference because they are based on Newton's Laws. This section will investigate the notion of an inertial frame and its implications to the navigation problem.

2.1 Inertial Frames

Experience with physical measurements points to the existence of certain distinguished frames of reference - the **inertial frames** - which are required for the expression of certain laws of physics. Newton believed in absolute space and time:

“Absolute space, in its own nature, without relation to anything external, always remains similar and immovable”

but general relativity has changed ideas of space and time since then. It is customary now to speak of inertial frames rather than absolute frames. There are an infinity of inertial frames and they pervade all of space just as did Newton's absolute ones.

If the existence of inertial frames and Newton's laws are taken as axiomatic, then inertial frames can be defined as those frames in which Newton's laws apply. Under this operational definition, the earth is not an inertial frame because an object dropped by an observer falls to the ground along a nonstraight trajectory. However, a spacecraft in orbit is such a frame, as is a free-falling aircraft, as is a point in space far removed from any massive object.

The simple existence of a gravitational field apparently does not disqualify a frame, for while the observer of a dropped object is not in an inertial frame on earth, the free-falling ball itself is an inertial frame⁵. If one inertial frame exists, then all other frames translating uniformly with respect to it are also inertial⁶. Any frame rotating with respect to it, cannot be inertial since rotation about any point but the center of mass requires acceleration.

Another definition of an inertial frame is one where the sum of all accelerations and gravitational attractions are zero. In the language of general relativity, the acceleration is “transformed away”. By and large, in an inertial frame, the gravitational attraction of the sun and other heavenly bodies can be ignored. These attractions can be ignored because the motions of any bodies close to the earth are affected by the sun just as the earth is, so the effect cancels out. However, as bodies move further from the earth, this is no longer so. For this reason, general relativity admits “local inertial frames”, where the laws of physics apply in a local region. Such frames are central to the description of both mechanical and electromagnetic phenomena.

5. Air friction is ignored throughout the document.

6. But there are still others, see the next section

$$v_{bs} = \frac{v_{bm} + v_{ms}}{1 - \frac{v_{bm}v_{ms}}{c^2}}$$

Notice that this law of physics permits combination of measurements taken in two different reference frames in order to predict a third measurement that was never made. Newton's second law:

$$\vec{F} = m\vec{a}$$

allows prediction of a measurement of force *taken in an inertial frame*, from a measurement of mass and of acceleration *taken in the same inertial frame*. For physical vectors, change of coordinate system does not change the reference frames involved or the quantity measured in any way.

Now the coordinate system conversion formulae of algebra solve an entirely different problem. They permit conversion between two coordinate systems for a single measurement taken in a single frame. In summary:

- The laws of physics permit prediction of measurements based on other measurements provided all measurements are taken in the appropriate reference frames for the law to apply. Different laws impose different limitations on reference frames.
- The mathematical manipulation of equations involving physical quantities require that all quantities first be expressed in a common coordinate system, but that choice of coordinate system is arbitrary.

This material is presented here because a naive implementation of inertial navigation leads to two basic problems. These problems come down to handling coordinate systems incorrectly, and handling reference frames incorrectly. The solution to the problem is twofold. First determine the laws of physics which allow a conversion from whatever an accelerometer measures to the measurement required. Second, express all such measurements in a common coordinate system. Then, a straightforward integration provides the vehicle position.

In order to discover what an accelerometer measures, it is necessary to find a particular frame of reference called an **inertial** frame of reference, and understand how it is affected by gravitational fields. In order to do that, it is necessary to confront the basic postulate of the theory of general relativity. With these tools, it will be possible to address the first half of the problem. The first step in that process is a brief discussion of the effects of motion of observers on the measurements of physical quantities.

1.6 Principle of Coordinate System Invariance

In linear algebra, when applied to physics, its vectors are often the mathematical representation of physical vectors in a particular cartesian coordinate system. In physics, the vector notation is designed to emphasize that the laws of physics hold irrespective of the coordinate system in which they are represented. This is called the **principle of coordinate system invariance**⁴. Newton's second law for instance says that force and acceleration are parallel vectors in the physics sense, whose amplitude ratio is the mass, irrespective of coordinate system, provided they are both measured with respect to an inertial reference frame and provided they are the net force on and actual acceleration of the same body. Hence the choice of different coordinate systems gives different vectors in the linear algebra sense, but the same vector in the physical sense.

1.7 Transformation Laws and Conversion Formulae

In parallel to the coordinate system, reference frame distinction, there is a related distinction between transforming vectors, and transforming their representations in a particular coordinate system. The first involves physics, the second mathematics.

Consider an experiment where a person stands on a moving train and throws a baseball toward the front of the train as shown in Figure 1:

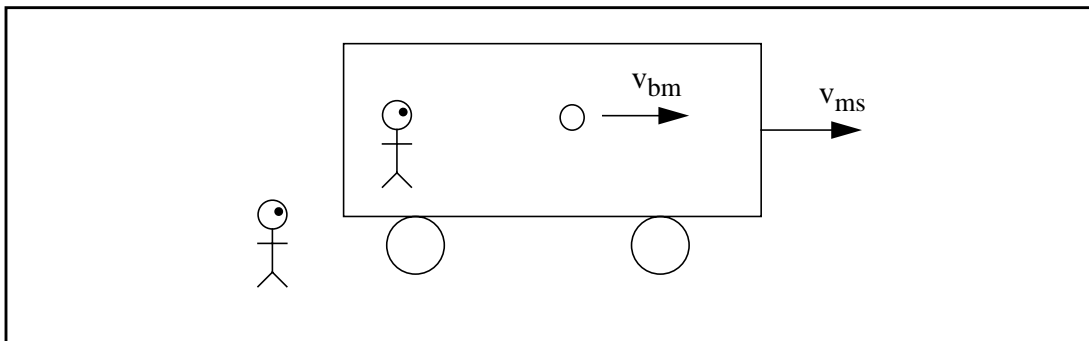


Figure 1 Einstein's Train

Let the person in the train be called the *moving observer* and the one on the side of the track be called the *stationary observer*. For the stationary observer measuring the ball's velocity, it is the laws of physics which permit a prediction of the measurement made by that observer based on the speed of the train and the speed of the ball. In Newtonian mechanics it is known that:

$$\vec{v}_{bs} = \vec{v}_{bm} + \vec{v}_{ms}$$

Now this is a law of physics which relates measurements made by observers in reference frames. It turns out that it involves the mathematical operation of addition, as defined on physical vectors. However, as velocities approach the speed of light, a different law of physics applies, namely the relativistic velocity addition equation:

4. This is why you can arbitrarily "assume" a direction of a scalar quantity, and it will come out negative if you "assumed" wrong. You actually are assuming nothing, but rather are establishing a scalar coordinate system.

1.1 Unit Systems

The unit system is the conventional meaning of a meter and a second. If the velocity were given in fathoms per decade, it would change the number 6.32 to something else, but it would still reflect the same measurement.

1.2 Number Systems

The number system is the conventional meaning 6.32. The Arabic number system is a base ten **weighted positional number system**. Other systems include the Roman system which is not entirely weighted positional and has no regular base. If the number 6.32 were given in Roman notation (assuming there is a way to represent decimals at all), it would look quite different, but it would still reflect the same measurement. Hence unit systems and number systems are coupled together and are used to express *quantity*.

1.3 Coordinate Systems

The **coordinate system** in the above example is a conventional two dimensional cartesian system based on a locally flat plane tangent to the earth. It is used to represent the direction of the velocity. Other systems include the spherical polar coordinate system, cylindrical polar coordinates, and curvilinear coordinates, and there are many others. If another coordinate system were used, it would have changed the word “northeasterly” to something else, but it would not change the measurement’s direction - it changes only *how the direction is expressed*. A coordinate system is therefore an artifice used to express direction of a measurement. For a scalar, the coordinate system is just the convention for assigning a plus or minus arithmetic sign.

1.4 Reference Frames

The idea of a **reference frame** is deeply imbedded in mechanics. Reference frames are real objects as expressed by Newton himself. Reference frames are distinguished from each other by their *state of motion*. Measurements always express a quantity with respect to two reference frames, since even such scalars as mass vary with reference frame in general. In conversation, the phrase <quantity> of <reference frame 1> relative to <reference frame 2> is often used to express these two frames of reference. Using velocity as an example, the frame in which a velocity is measured expresses the agreed upon notion of zero velocity, for that particular measurement. The frame possessing the velocity is the thing which is considered to be moving, for that particular measurement. Only things have a velocity. The velocity of empty space is meaningless.

1.5 Reference Frames vs. Coordinate Systems

One often overlooked fact which one must face squarely when doing advanced dynamics is that a reference frame and a coordinate system are different. The former is a concept of physics, the second a concept of mathematics, and they are absolutely not the same thing. There are at least four common practices today which obscure this distinction. The first is the use of the same term “vector” in both linear algebra, and in physics, when two different concepts are involved. The second is the lack of emphasis of the distinction between the tensor of second order and the matrix. The third is the hybrid term “coordinate frame” as it is used in computer graphics and robotics. The fourth is the widespread and sometimes confusing practice of the use of random expressions like “coordinate reference system”, “reference coordinate system”, “coordinate reference frame”, etc., in many texts. This may be justified, however, because the issue can do more to confuse the student than to illuminate the basics. In the work here, the issue is of central concern.

PART I: Relevant Concepts from Theoretical Mechanics

In order to present the implementation of a working INS, some aspects of physical mechanics that are central to the discussion must be covered. This section will overview ideas from the branch of physics called **theoretical mechanics**, rather than its pragmatic engineering counterpart. The reader will find that later sections will refer heavily to the material presented here.

1. Context of Measurements of Physical Quantities

Physical quantities can be classified into tensors, vectors and scalars, which are all familiar³. Consider a measurement made by an observer of a velocity of a train relative to a car of 6.32 meters per second in a northeasterly direction. When the observer announces the measurement, there is an entire *context* in which it is made, part of it explicitly stated, part understood by convention. That context involves five components:

- a unit system (meters, seconds)
- a number system (6.32)
- a coordinate system (north, east etc.)
- a reference frame to which the measurement is ascribed (the train)
- a reference frame with respect to which the measurement is made (the car)

3. Scalars possess magnitude only, vectors magnitude in a particular direction. Tensors such as solid body stress (not fluid pressure - which is a scalar) and inertia possess magnitude in all directions. Tensors possess three principal axes of magnitude, from which all others can be determined - hence the use of the inertia ellipsoid in rigid body mechanics and of Mohr's circle of stress, strain, and inertia in solid mechanics.

It turns out that the devices are not that difficult to understand at all - there just is no information available. I knew there was some deep stuff involved because I had heard that the motion of the earth and the gravitation of the planets had to be modelled, and that you needed general relativity to understand the new optical gyros, but this is overstated.

In the course of the work, it became clear that integrated INS-GPS systems are the wave of the future, and that the reason for this is the long recognized symbiotic relationship between dead reckoning and triangulation. Therefore, a complete effort required a study of the new GPS system as well.

1. Introduction

This century has seen two breakthroughs in navigation¹ technology, equal perhaps in importance to the discovery of the compass. The latest is the **global positioning system**, or GPS, which has recently come on line. Based on modified triangulation of a network of satellites in earth orbit, it allows anyone, anywhere, to fix position through the use of a small hand held receiver. The other breakthrough, arriving some years earlier, is **inertial navigation systems** (INS). These systems, based on dead reckoning², have been of considerable interest to the military and aerospace industries for many reasons. This report will discuss the basic principles underlying the operation of inertial and satellite navigation systems from both a theoretical and a practical perspective.

1.1 Commentary

One of the distinguishing features of the study of navigation is that, as old as it is, it seems that it has always been the province of a small group of specialists. Most other ancient technologies like flint tools, fire, the wheel, animal husbandry and agriculture, had an obvious utility to the ordinary person, and were intellectually accessible to the masses. Literacy, while once enjoyed only by the ruling classes, has now become available to at least most people in the industrialized world. Of all of the achievements of the ancients, navigation seems to stand alone in this regard.

This project started as an attempt to understand some of the peculiarities of the INS as it is employed on mobile robots at CMU. In particular, the issue was why it is that the Litton PADS inertial navigation system then in use on the NAVLAB II apparently could not generate a reliable estimate of the vehicle z coordinate. This has become a very important question because sophisticated map matching algorithms still consume too much time for high speed navigation. The answer to this question turned out to require nothing less than the fundamental postulate of general relativity.

I ended up doing more work than I intended, and came to answer a number of other questions which are more fundamental prerequisites. I wrote down what I learned in the belief that others would find the information useful. I had a hard time finding any real information myself. There seems to be three kinds of information out there, industrial product brochures which teach you nothing, popular pseudoeducational seminars which sanitize the math to the point of uselessness, and the academic literature which, after sixty years of development of the technology behind closed doors, is now impenetrable. What was intended as a five page analysis, has become twenty times that size because the more I looked, the more I did not understand.

Ask the community here how these devices function and you are likely to get the answer “black magic” as often as something technical. Like many other devices whose major current use is the military, there is an aura of secrecy about them. The lore is that they do really neat stuff like measuring the earth’s rotation and that they incorporate those seemingly ultra high-tech components called gyroscopes which are synonymous with space travel and decades of painstaking precision engineering. Because they use the gyroscope, I felt I could do this project for academic credit in Matt Mason’s manipulation course and get a double benefit.

1. The term navigation, which is two dimensional, is distinguished from guidance, which is three dimensional. The word is derived from the latin “navis” meaning ship and “agere” meaning to move or direct.

2. Other navigation techniques include landmark recognition and triangulation. The origin of the term is “deduced reckoning”.

List of Figures

Figure 1 Einstein's Elevator Thought Experiment	8
Figure 2 General Moving Reference Frames	9
Figure 3 Precession	16
Figure 4 Sidereal Day	18
Figure 5 Galileo's Experiment Revisited	20
Figure 6 The Components of Gravity	23
Figure 7 Mechanical Gyroscope	26
Figure 8 Single Axis Gyro Precession	27
Figure 9 Single Axis Stable Platform	29
Figure 10 Gyrocompass	31
Figure 11 Ring Interferometer	36
Figure 12 Typical Monolithic Ring Laser Gyro Configuration	39
Figure 13 Naive INS Concept	43
Figure 14 Accelerometer	45
Figure 15 Generic Mechanization	50
Figure 16 Terms in the Navigation Equation	51
Figure 17 Accelerated Pendulum	55
Figure 18 Schuler Loop	56
Figure 19 Accelerometer Negative Feedback	57
Figure 20 Taxonomy of Inertial Systems	65
Figure 21 North Slaved, Locally Level System	66
Figure 22 Free Azimuth, Locally Level System	67
Figure 23 Tangent Plane, Locally Level System	68
Figure 24 Two Dimensional Range Triangulation	78
Figure 25 Computing the User Position and Clock Offset	80
Figure 26 GPS Signals	83
Figure 27 Receiver Code Correlation	84
Figure 28 Orbital Configuration	85
Figure 29 GPS System Master Control	86
Figure 30 Selective Availability	89
Figure 31 Pseudorange Error Sources	90
Figure 32 Multipath Ground Reflection	91
Figure 33 Geometric Dilution of Precision	92
Figure 34 Differential GPS	94
Figure 35 Spatial Variation of System Error	95
Figure 36 Temporal Variation of System Error	96
Figure 37 GPS Interferometry	96
Figure 38 GPS Modes of Operation	97

2.4	Ground Segment	86
3	Performance.....	87
3.1	Range Resolution	87
3.2	Measures of Accuracy	88
3.3	Selective Availability	88
3.4	Sources of Error	89
3.5	Modes of Operation	93
3.6	GPS Receivers	97
4	Glossary	99
5	Bibliography	102

5.3	Solution Near the Earth's Surface	60
5.4	Error Dynamics of the Free Inertial System	61
6	Implementation	62
6.1	Gimballed and Strapdown Mechanizations	62
6.2	Navigation Coordinate Systems	64
6.3	Common Mechanized Coordinate Systems	66
6.4	Position Coordinates	69
6.5	Initialization	71
7	Performance	73
7.1	Inertial Grade Componentry	73
7.2	Aided Inertial Systems	73
7.3	Sources of Error	74
7.4	Configuration Of Inertial Systems	75
7.5	Physical	75
7.6	Data Rates	75
7.7	Initialization	75
7.8	Accuracy	76

PART IV: Satellite Navigation Systems77

1	Principles of Operation	78
1.1	Position Measurement	78
1.2	Time Measurement	79
1.3	Wave Speed Measurement	81
1.4	Velocity Measurement	81
1.5	Orientation Measurement	81
2	Implementation	82
2.1	Signal Characteristics	82
2.2	User Segment	84
2.3	Space Segment	85

PART III: Inertial Navigation Systems41

1	Basic Principle of Inertial Navigation	41
1.1	Characteristics of the INS	41
1.2	The Naive INS	42
2	Generic Mechanization.....	45
2.1	Specific Force to Inertial Acceleration	45
2.2	Inertial Acceleration to Earth Referenced Acceleration	47
2.3	Vehicle Referenced Acceleration	48
2.4	Vector Integration	49
2.5	Generic Mechanization	50
2.6	Temporal Error Propagation	51
3	Complete Solution in Geocentric Cartesian Coordinates	52
3.1	Inertially Referenced Angular Velocity	52
3.2	Earth Referenced Angular Velocity	52
3.3	Specific Force	53
3.4	Gravity	53
3.5	Velocity Solution	53
3.6	Attitude	53
3.7	Direction Cosine Matrix	54
3.8	Position Solution	54
4	Schuler Dynamics.....	55
4.1	Schuler Period	55
4.2	Schuler Tuning	55
4.3	Schuler Loop	56
4.4	Faucault Effect	58
5	Inertial Error Dynamics.....	59
5.1	Perturbative Analysis	59
5.2	Linearized Gravity Model	60

PART II: Inertial Sensors and Stable Platforms22

1	Accelerometers	22
1.1	Gravity	22
1.2	Specific Force	23
1.3	Effect of the Earth's Gravitation	23
1.4	Basic Equation of Inertial Navigation	24
1.5	Practical Accelerometers	25
2	Mechanical Gyroscopes.....	26
2.1	Simple Gyroscope	26
2.2	Single Axis Gyroscope	27
2.3	Practical Gyroscopes	28
3	Stable Platforms.....	29
3.1	Single Axis Stable Table	29
3.2	Three Axis Stable Table	30
3.3	Complete INS	30
4	Gyrocompasses	31
4.1	Equation of Motion	31
4.2	Expression in Principle Coordinates	33
4.3	Interpretation as a Pendulous Oscillator	34
4.4	Period of Oscillation	34
5	Optical Gyroscopes.....	35
5.1	Michelson Interferometer	35
5.2	Sagnac Effect	35
5.3	Lasers	38
5.4	Ring Laser Gyroscopes	38
5.5	Practical Ring Laser Gyroscopes	38
5.6	Optical Fibre Gyroscopes	40
5.7	Practical Optical Fibre Gyroscopes	40

Table of Contents

1	Introduction	1
1.1	Commentary	1
PART I: Relevant Concepts from Theoretical Mechanics.....		3
<hr/>		
1	Context of Measurements of Physical Quantities	3
1.1	Unit Systems	4
1.2	Number Systems	4
1.3	Coordinate Systems	4
1.4	Reference Frames	4
1.5	Reference Frames vs. Coordinate Systems	4
1.6	Principle of Coordinate System Invariance	5
1.7	Transformation Laws and Conversion Formulae	5
2	Frames in Relative Motion	7
2.1	Inertial Frames	7
2.2	Principle of Relativity	8
2.3	Theorem of Coriolis	9
3	Rotational Dynamics	12
3.1	Free Rotation of An Arbitrary Body	12
3.2	Forced Rotation of an Arbitrary Body	13
3.3	Free Precession of a Symmetric Body	13
3.4	Forced Precession	15
4	Motion on the Earth.....	17
4.1	Earth as an Inertial Frame	17
4.2	Effect of the Earth's Rotation	19

Abstract

Every mobile robot needs to know where it is so everyone who does mobile robotics wants to know how inertial and satellite navigation works. These two technologies are the most sophisticated forms of navigation systems available and they complement each other very well. Applications requiring indications of highly dynamic 3D motion, excellent relative accuracy and high update rates benefit from inertial systems. Applications requiring bounded absolute accuracy for extended excursions or position estimates relative to the earth itself benefit from satellite navigation.

This decade has seen the development of inexpensive handheld global positioning systems based on reception of the GPS satellite signals, and advances in optical gyroscopes and integrated circuit accelerometers which promise to put an inertial navigation system in every automobile before the end of the next. This report is a detailed tutorial which explains the principles, practice, and issues of using these new technologies.

Modern Inertial and Satellite Navigation Systems

Alonzo Kelly

CMU-RI-TR-94-15

The Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

May 2, 1994

©1994 Carnegie Mellon University

This research was sponsored by ARPA under contracts “Perception for Outdoor Navigation” (contract number DACA76-89-C-0014, monitored by the US Army Topographic Engineering Center) and “Unmanned Ground Vehicle System” (contract number DAAE07-90-C-RO59, monitored by TACOM).

The views and conclusions expressed in this document are those of the author and should not be interpreted as representing the official policies, either express or implied, of the US government.