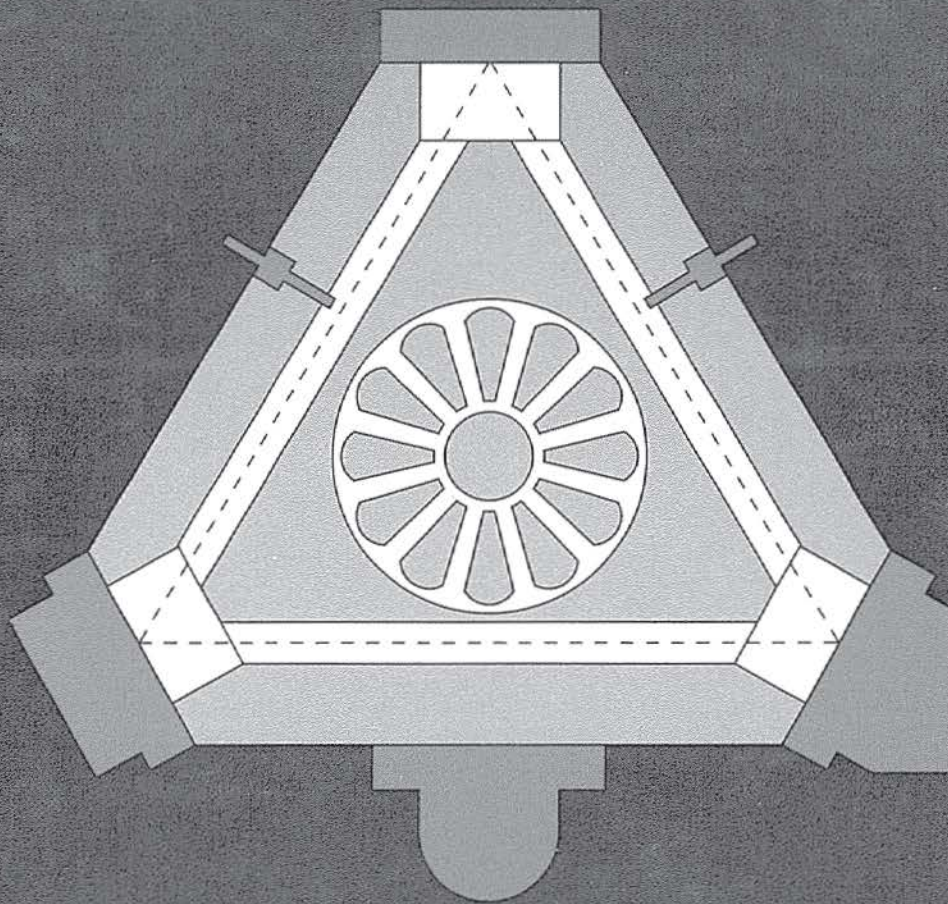


Anthony Lawrence

Modern Inertial Technology

Navigation, Guidance, and Control



Springer-Verlag

Anthony Lawrence

Modern Inertial Technology

Navigation, Guidance,
and Control

With 129 Illustrations



Springer-Verlag

New York Berlin Heidelberg London Paris

Tokyo Hong Kong Barcelona Budapest

629.X1351 L431m
Anthony Lawrence
2 Chapel Place
Pepperell, MA 01463 USA

c.1
NOV 2 '93

The cover drawing shows a ring laser gyroscope, an inertial sensor used in many modern navigation and control systems. It is used in military and civilian airplanes, including the Boeing 767.

Library of Congress Cataloging-in-Publication Data

Lawrence, Anthony, 1935-

Modern inertial technology : navigation, guidance, and control /
Anthony Lawrence.

p. cm.

Includes bibliographical references.

ISBN 0-387-97868-2

1. Inertial navigation (Aeronautics) I. Title.

TL588.5.L38 1992

629.135'1—dc20

92-16602

Printed on acid-free paper.

✓
© 1993 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Ellen Seham; manufacturing supervised by Vincent Scelta.

Camera-ready copy prepared using the author's WordPerfect files.

Printed and bound by Hamilton Printing Co., Rennselaer, NY.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-97868-2 Springer-Verlag New York Berlin Heidelberg

ISBN 3-540-97868-2 Springer-Verlag Berlin Heidelberg New York

Contents

Introduction	1
1. An Outline of Inertial Navigation	4
Navigation's Beginnings	4
Inertial Navigation	6
Maps and Reference Frames	6
The Inertial Navigation Process	8
Inertial Platforms	9
Heading and Attitude Reference Systems	10
Schuler Tuning	12
Gimbal Lock	12
Strapdown Systems	13
System Alignment	14
Gyrocompassing	15
Transfer Alignment	16
Advantages and Disadvantages of Platform Systems	16
Advantages	16
Disadvantages	17
Advantages and Disadvantages of Strapdown Systems	17
Advantages	17
Disadvantages	18
Star Trackers	19
The Global Positioning System	19
Applications of Inertial Navigation	21
Conclusions	21
References	22
2. Gyro and Accelerometer Errors and Their Consequences	24
Effect of System Heading Error	24
Scale Factor	25

Non-linearity and Composite Error	26
System Error from Gyro Scale Factor	26
Asymmetry	27
Bias	27
System Error from Accelerometer Bias	27
Tilt Misalignment	29
System Error from Accelerometer Scale Factor Error	29
System Error from Gyro Bias	29
Random Drift	30
Random Walk	31
Dead Band, Threshold, and Resolution	31
Hysteresis	31
Day-to-Day Uncertainty	32
Gyro Acceleration Sensitivities	32
g-Sensitivity	34
Anisoelasticity	34
Rotation Induced Errors	35
Angular Acceleration Sensitivity	35
Anisoinertia	37
Angular Accelerometers	38
Angular Accelerometer Threshold Error	38
The Statistics of Instrument Performance	39
Typical Instrument Specifications	40
References	41
3. The Principles of Accelerometers	42
The Parts of an Accelerometer	42
The Spring-Mass System	43
Q Factor	46
Bandwidth	47
Open-Loop Pendulous Sensors	47
Cross-Coupling and Vibropendulous Errors	47
Pickoff Linearity	49
Closed-Loop Accelerometers	49
Open-Loop Versus Closed-Loop Sensors	50
Sensor Rebalance Servos	50
Binary Feedback	51
Ternary Feedback	53
Pulse Feedback and Sensors	54
The Voltage Reference Problem	54
Novel Accelerometer Principles	54
Surface Acoustic Wave Accelerometer	55
Fiber Optic Accelerometers	55
References	56

4. The Pendulous Accelerometer	57
A Generic Pendulous Accelerometer	57
Mass and Pendulum Length	57
Scale Factor	58
The Hinge	59
The Pickoff	59
The Forcer and Servo	60
The IEEE Model Equations	60
The Sundstrand "Q-Flex" Accelerometer	61
The Capacitive Pickoff	63
The Forcer	64
Other Electromagnetic Pendulous Accelerometers	65
Moving Magnet Forcers	66
Electrostatic Forcers	66
The Silicon Accelerometer	67
References	70
5. Vibrating Beam Accelerometers	72
The Vibration Equation	72
The Resolution of a Vibrating Element Accelerometer	73
The Quartz Resonator	76
VBAs in General	76
The Sundstrand Design	77
Accelerex Signal Processing	78
The Kearfott Design	79
Comparison of Free and Constrained Accelerometers	81
General Comparison of the SPA and VBA	81
Comparison of Performance Ranges	82
Conclusion	83
References	83
6. The Principles of Mechanical Gyroscopes	84
Angular Momentum	84
The Law of Gyroscopics	85
Parasitic Torque Level	86
The Advantage of Angular Momentum	86
The Spinning Top – Nutation	87
Equations of Spinning Body Motion	88
Coriolis Acceleration	89
Gyroscopes with One and Two Degrees of Freedom	91
Conclusion	92
References	92

7. Single Degree of Freedom Gyroscopes	93
The Rate Gyro	93
The Scale Factor	94
The Spin Motor	95
The Ball Bearings	96
Damping	96
The Pickoff	96
The Torsion Bar	97
Flexleads	98
Rate Gyro Dynamics	98
The Rate-Integrating Gyro	99
The Torquer	100
The Output Axis Bearing	102
The Principle of Flotation	103
Damping	104
Flotation Fluids	105
Structural Materials	106
A Magnetic Suspension	107
Self-acting Gas Bearings	109
Anisoelasticity in the SDFG	110
Anisoinertia in the SDFG	111
Vibration Rectification	113
The SDFG Model Equation	114
A Digression into Accelerometers	115
The Pendulous-Integrating Gyro Accelerometer	116
Conclusion	117
References	117
 8. Two Degree of Freedom Gyroscopes	 119
The Two Degree of Freedom (Free) Gyro	119
The External Gimbal Type	120
Two-Axis Floated Gyros	121
Spherical Free Rotor Gyros	122
The Electrically Suspended Gyro	123
The Gas Bearing Free Rotor Gyro	125
References	126
 9. The Dynamically Tuned Gyroscope	 127
The DTG Tuning Effect	127
The Tuning Equations	129
DTG Nutation	132
Figure of Merit	132
Damping and Time Constant	132

Biases Due to Damping and Mistuning	133
Quadrature Mass Unbalance	134
Synchronous Vibration Rectification Errors	136
Axial Vibration at 1N	136
Angular Vibration at 2N	137
Wide Band Vibration Rectification Errors	138
Anisoelasticity	138
Aniso inertia	140
Pseudoconing	141
The Pickoff and Torquer for a DTG	142
The DTG Model Equation	146
Conclusion	146
References	147
 10. Vibrating Gyroscopes	 148
The Vibrating String Gyro	149
The Tuning Fork Gyro	149
Vibrating Shell Gyros	152
The Hemispherical Resonator Gyro	153
Scale Factor	154
Asymmetric Damping Error	155
The Vibrating Cylinder (START) Gyro	155
The Advantages of Vibrating Shell Gyros	157
The Multisensor Principle and Its Error Sources	157
Conclusion	161
References	161
 11. The Principles of Optical Rotation Sensing	 163
The Inertial Property of Light	163
The Sagnac Effect	164
Sagnac Sensitivity – The Need for Bias	166
The Shot Noise Fundamental Limit	167
The Optical Resonator	169
The Fabry-Perot Resonator	170
Resonator Finesse	173
The Sagnac Effect in a Resonator	173
Active and Passive Resonators	174
Resonator Figure of Merit	174
Optical Fibers	175
Refraction and Critical Angle	176
Multimode and Single Mode Fibers	177
Polarization	178
Birefringent Fiber for a Sagnac Gyro	178

The Coherence of an Oscillator	179
Types of Optical Gyro	179
Conclusion	180
References	180
12. The Interferometric Fiber-Optic Gyro	182
The History of the Fiber-Optic Gyro	182
The Basic Open-Loop IFOG	183
Biasing the IFOG	184
Non-reciprocal Phase Shifting	185
The Light Source	186
Reciprocity and the "Minimum Configuration"	186
Closing the Loop – Phase-Nulling	187
Acousto-Optic Frequency Shifters	188
Integrated Optics	189
Serrodyne Frequency Shifting	190
Fiber-to-Chip Attachment – The JPL IFOG	191
Drift Due to Coil Temperature Gradients	192
The Effect of Polarization on Gyro Drift	192
The Kerr Electro-Optic Effect	194
The Fundamental Limit of IFOG Performance	195
Conclusions	195
References	196
13. The Ring Laser Gyro	199
The Laser	199
Stimulated Emission	199
The Semiconductor Laser	202
The Ring Laser	202
Lock-in	203
Mechanical Dither	203
The Magnetic Mirror	206
The Multi-oscillator	208
Shared-Mirror RLG Assemblies	210
The Quantum Fundamental Limit	210
Quantization Noise	213
Conclusion	213
References	214
14. Passive Resonant Gyros	216
The Discrete Component Passive Ring Resonator	216
The PARR Fundamental Limit	218

The Resonant Fiber-Optic Gyro	218
The Micro-Optic Gyro	222
The MOG Fundamental Limit	223
IFOG, RFOG, and MOG Size Limits	226
Fundamental Limits for RFOG, IFOG, and RLG	226
Conclusions	227
References	228
15. Testing Inertial Sensors	230
Inertial Sensor Test Labs	230
Performance Test Gear	231
Environmental Test Gear	232
Qualification, Acceptance, and Reliability Tests	233
Accelerometer Testing	234
The Accelerometer Acceptance Test Procedure	234
Centrifuge Tests	237
Gyroscope Testing	238
Testing the SDF Rate Gyro	238
Testing SDF Rate-Integrating Gyros	239
Tombstone Tests	240
The Six-Position Test	241
The Polar-Axis (Equatorial Tumble) Test	242
The Servo Table Scale Factor Test	243
Vibration Tests	244
Testing the Dynamically Tuned Gyro	245
The Eight-Position Test	245
DTG Rate Testing	247
Testing Optical Gyros	247
The Sigma Plot	248
Conclusion	249
References	250
16. Design Choices for Inertial Instruments	251
A Platform or a Strapdown System?	251
Aiding the IMU	252
Choice of Sensor Type	252
Differential Design	253
Using Resonance	253
Mechanical or Optical Gyros?	254
Reliability	254
Redundancy	255
Sensor Design Check Lists	256
Conclusions	257
References	258
Index	259

1

An Outline of Inertial Navigation

Navigation's Beginnings

Pioneers returning from their journeys provided travel instructions for those who wished to repeat their journeys. They wrote descriptions of their routes and made charts or maps pointing out landmarks and hazards like rivers and mountains on land, or shoals and rocks at sea. Map makers devised a global coordinate system using a grid of latitude and longitude circles, by which the position of any place on earth could be defined.

Buoys and lighthouses provided route markers for sailors close to land, but once they ventured into the featureless seas, their pathfinding became much more difficult. Magnetic compasses were known as early as the 11th century, but were not very accurate. The 14th- century navigator used the stars to find latitude, and by the 18th century, technology had provided instruments like the sextant for more accurately measuring the positions of the celestial bodies.

Finding longitude was more difficult, for the positions of the celestial bodies depend on the rotational position of the earth relative to the stars, that is, the time of day. Longitude is measured from an arbitrarily chosen zero at Greenwich, near London, England; to determine longitude, navigators need to know the time where they are and the time at Greenwich (Greenwich Mean Time, GMT). They can find local time from observations of the sun, but finding GMT means observing the satellites of Jupiter, or lunar distances [1], which are uniquely positioned at any time. But these observations require complicated calculations to yield time, so it is much better to have a clock set to GMT on board. It took the invention of the spring escapement clock in the mid-18th century to make accurate timekeeping available at sea. How the 18th-century mariner would envy today's \$10 quartz watch, accurate beyond all his wildest dreams! Nowadays, GMT is broadcast by radio, worldwide.

This process of following a path on a map through predetermined latitudes and longitudes is called *navigation*, and the process of pointing the vehicle to follow a chosen path is called *guidance*. Navigation at sea uses the process of *deduced reckoning* (called "dead" reckoning), illustrated in Figure 1.1. If a ship starts from a known latitude and longitude and travels in a known direction for a known time, its position on the previously charted sea is known and can be

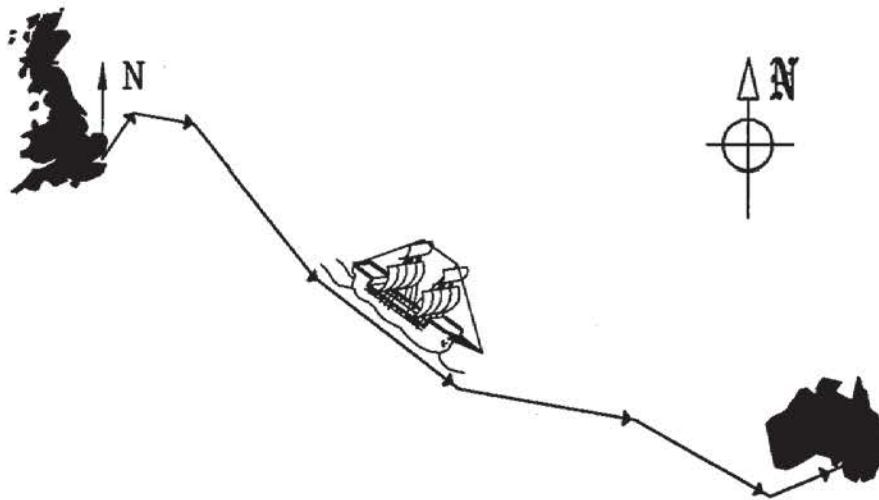


Figure 1.1. Inertial navigation is dead reckoning.

verified by sun (or other celestial body) sightings. Provided the ship's direction and speed are known accurately, the sailors will land in the correct harbor.

When aircraft began to fly over the sea, they got direction and speed from the magnetic compass and airspeed indicator. As radar evolved, the Doppler set could give more accurate ground speed. Dead reckoned courses tried to follow planned routes, corrected by celestial navigation, but airplanes fly so fast that the navigator was hard pressed to make celestial measurements quickly enough. More accurate, safer, navigation was possible when land-based radio beacons and systems like Loran and Omega became available.

By 1908, the magnetic compass had been replaced by the gyrocompass. Gyrocompasses are important marine navigation instruments, but they are not generally useful for aircraft, spacecraft and guided missiles. Because our focus is on the instruments for inertial navigators, we are not going to consider the gyrocompass in this book; they are well described in general navigation books such as Dutton's [2], by Wrigley in the *Encyclopedia Britannica* under "Gyrocompass," and by others [3,4].

Military aircraft prefer not to depend on radio beacons in time of war, for an enemy can jam them, and, conversely, aircraft can be detected by the enemy if they transmit Doppler radar speed measurement signals. They need a self-contained navigation system. If one wants to guide unmanned vehicles (such as a space probe or an intercontinental missile), one must also have a self-contained navigation system. Civil aircraft need to have reliable navigation at the lowest possible cost, and human navigators are expensive. The solution to all these needs is to automate navigation, which means that one must find a way to keep track of position and attitude continuously, in all weather conditions, in space, in the sky, and under the sea. Inertial navigation serves this purpose.

Inertial Navigation

Gyroscopes and accelerometers can provide the necessary signals for automatic navigation. Gyroscopes measure rotation, and accelerometers measure acceleration. Integrating the output from an accelerometer gives speed, and integrating speed gives distance traveled. The gyroscopes provide information on where the accelerations are directed, and therefore heading and distance are known, the essential ingredients for dead reckoning. As these instruments use the inertial properties of matter (or of light) for their operation, dead reckoning with gyros and accelerometers is called *inertial navigation*.

The first inertial navigators were used in the German V1 and V2 weapons in World War II. After the war, a group of German scientists, under Werner Von Braun, developed this technology at Redstone Arsenal, Huntsville, Alabama, for ICBMs and spacecraft, building inertial navigation systems for U.S. Army missiles.

Other U.S. groups developed IN systems, notably one under Draper at the Massachusetts Institute of Technology [5,6]. Their first aircraft inertial navigators were flown in 1949, followed in 1954 by the Navy Ship's Inertial Navigation System. The Draper Laboratory made inertial systems for Polaris submarines; one of them, the *Nautilus*, successfully navigated under the polar ice past the North Pole. In the 1960s the Apollo program took inertial guidance into space, and now inertial systems are being used in "smart" munitions. For more on the history of inertial navigation, see Wrigley's summary [7].

Maps and Reference Frames

Before we go anywhere, we need to orient ourselves in the space in which we will navigate, which we do with *maps* and *reference frames*. A map or chart is drawn to some scale so that the user can calculate distances between places. Over the distance covered by a town map we assume that the flat sheet represents a flat area of the earth, and although there may be hills and valleys in the town, the town is represented as if it were a model on a flat board. But for larger areas, the earth, being roughly a sphere, cannot be precisely represented by flat maps. Many different projections have been developed for making flat navigation maps [2], but we will leave those readers who are interested to pursue this field elsewhere.

Maps are oriented in a known direction, usually with North up the map. Navigation requires us to define frames of reference (or coordinate sets) so that we can orient ourselves in the mapped space, and in this instance "north" specifies the map's reference frame. There is a frame fixed in the "fixed stars," the *inertial frame*, that Newton postulated for his laws of inertia in 1687, and inertial navigation is navigation in this frame. This is a reference frame that is independent of the motion of the vehicle.

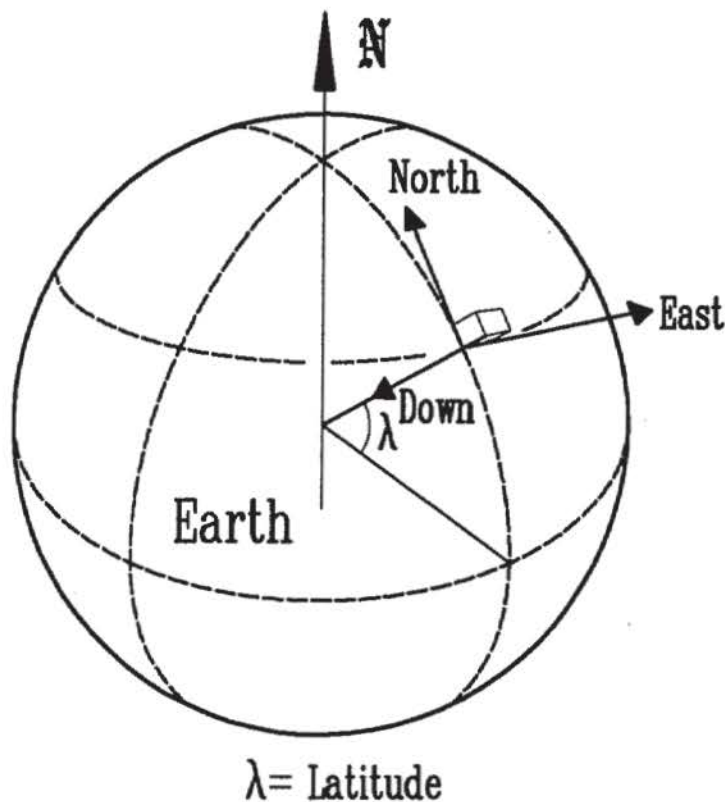


Figure 1.2. Earth-fixed axes.

On (or close to) the earth's surface it is more useful to work in a reference frame oriented to surrounding objects, a *local level* frame with North, East, and Vertically-Down as axes. In the simple (spherical homogeneous) earth local level frame, Figure 1.2, Down is a vector pointing at the center of the earth at the angle of latitude λ , North is the horizontal vector in the plane of Down and the earth's spin axis (True North), and East is the horizontal normal to that plane at the observer's site. The earth rotates once in 24 hours, so its rotation rate is 15 deg/h; the horizontal and vertical earth's rate components Ω_{eh} and Ω_{ev} in local level axes are

$$\begin{aligned}\Omega_{eh} &= 15 \cos \lambda \\ \Omega_{ev} &= 15 \sin \lambda \text{ deg/h}\end{aligned}$$

where λ = latitude.

East is the direction in which there is no earth's rate, a fact that is used to align inertial navigation systems because it is much easier to find a null than the peak of a maximum. Gyroscopic north-finding systems find East!

Vehicles have their own axis set, Roll, Pitch, and Yaw, shown in Figure 1.3, corresponding to the conventional use of the terms.

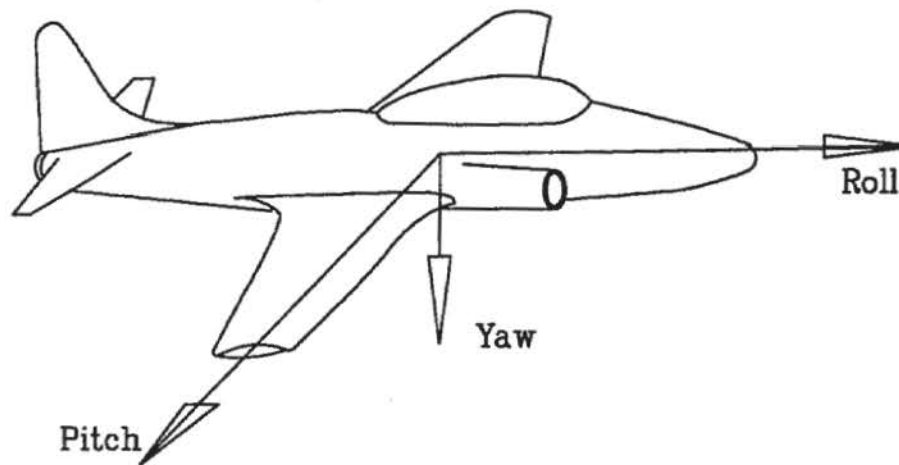


Figure 1.3. Vehicle axes set.

The Inertial Navigation Process

The inertial navigation system, INS (or unit, INU), is made from a navigation computer and a set of gyroscopes and accelerometers that measure in Newton's inertial axes, generally called *inertial sensors*. The group of inertial sensors is commonly called an *inertial measurement unit* (IMU) or an *inertial reference unit* (IRU). Once aligned to a set of reference axes (such as the "North, East, Down" set), the sensors provide distance measurements and the navigation computer carries out the continuous dead reckoning calculations.

The inertial sensors might be mounted in a set of gimbals so that they stay level and head in a fixed direction whatever the vehicle's motion. This construction is called a *navigation platform*. Alternatively, the instruments might be attached to the vehicle, in which case they measure its motion components in the vehicle axes set, and the system computes direction traveled in the reference axes by transforming the measurements from the vehicle axes to the reference axes. This is called a *strapdown* system, jargon for instruments "strapped down" to the vehicle.

To navigate inertially, we first measure the accelerations in the directions of the navigation axes, and if our instruments are not perfect, we might compensate their readings by removing biases or scale factor errors (defined in Chapter 2), perhaps known as a function of the system's measured temperature.

Second, to find the vehicle's vertical acceleration, we subtract gravity from the "Down" accelerometer output, perhaps using a gravity model to allow for the variation of gravity with latitude and longitude. The earth's rotation causes a centripetal acceleration that is greatest (about 3 mg) at the equator and zero at the poles. In addition, the earth is not a sphere; rotation has flattened it, causing

a mass concentration at the equator, so that the value of g varies with latitude. Precise navigators, like those in nuclear submarines, carry gravity gradiometers to correct for gravity vector errors; because of this unevenness in mass distribution, the vector direction is not necessarily along an earth's radius [8,9]. In space we can use Newton's Law of Gravitation to compute the local gravity, but once on a foreign planet, we lose accuracy until we can map its gravity fields.

Third, we integrate the accelerations over a known time, once to get the velocity, twice to get the distance traveled. As integration is the process of summing the outputs at frequent, known intervals, we must know the time interval accurately for it enters as a squared term in the distance computation.

Fourth, we measure the rotation rates either from gimbal motions in a stabilized platform, or directly with gyroscopes in a strapdown system. We then compensate for gyro bias and possibly scale factor errors, and determine a new heading.

Fifth, we compensate for earth's rotation if we are in local level axes, for otherwise the platform would be space stabilized and would seem to tilt in the vehicle axes set.

Finally, the combined distance and heading data give us an updated dead reckoned position to display. Then we go back to the beginning and do it all over again, until the end of the journey.

Inertial Platforms

An inertial platform uses gyros to maintain the accelerometers in a fixed attitude, i.e., the gimballed platform serves to define the directions for the measurements of acceleration. A single-axis platform, Figure 1.4, consists of a gyro mounted so that its sensing axis or *input axis* (IA) is along the axis of the platform, which is set into the vehicle in bearings. Electrical power for the gyro comes in, and the gyro's output goes out, through slip rings. The platform can be driven around its axis by an electric torque motor. The gyro output, which indicates an unwanted platform rotation, drives the torque motor through a servo amplifier; it basically provides the torque to overcome the friction in the slip rings and bearings.

If the platform had no gyro and the vehicle rotated around the platform axis, the platform's inertia would tend to keep the platform aligned in space. But the platform would slowly accelerate due to the torque transmitted through the bearings and slip rings, and the inertial reference would be lost. By adding the gyro, we provide a means of sensing any platform rotation with respect to inertial space, down to the limit of the gyro's resolution. The gyro does not **measure** anything in the inertial platform; it is only used to maintain a fixed position (null operation), and so does not need to be able to measure large rotation rates.

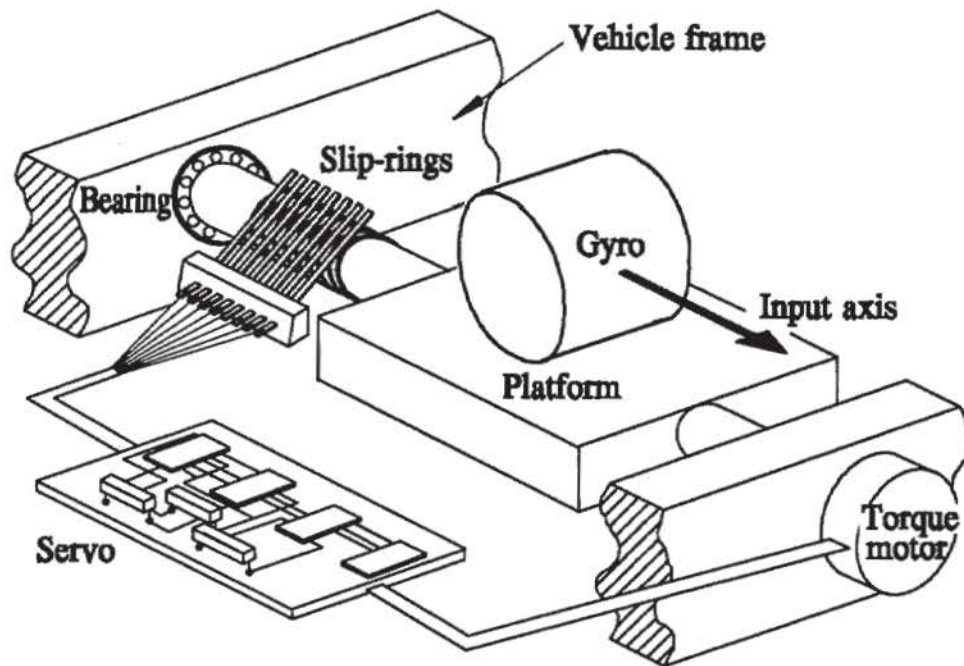


Figure 1.4. A single-axis platform.

For navigation in three dimensions we must expand our platform to three axes by adding gimbals, Figure 1.5. Here we have shown gyros with two sensing axes each (*two degree of freedom* gyros) and also schematically show a group of three accelerometers. We have also provided a *pickoff* on each gimbal axis, a device that measures the angle between two gimbals; the torquers, slip rings, and three servo loops are there but are not shown. If the gimbals are aligned with the vehicle axes before setting out, then the gimbals, being gyro stabilized, define the attitude and heading of the vehicle during the journey.

The gyros alone will try to maintain the platform aligned in inertial space. If the platform is operating in local level coordinates, the navigation computer must keep the platform horizontal. It does this by sending command signals to the gyros which otherwise would fight the gimbal motion. The roll and pitch accelerometers can be used to level the platform if we know that it is not accelerating. The relationship between the sensors, the gimbals and the navigation computer is shown schematically in Figure 1.6.

Heading and Attitude Reference Systems

Rather than use an expensive inertial navigator, many aircraft use a simplified version that indicates the direction of flight (the heading) and the attitude in roll and pitch. This Heading and Attitude Reference System (HARS) is a platform with three stabilized axes and two horizontal accelerometers. The accelerometers

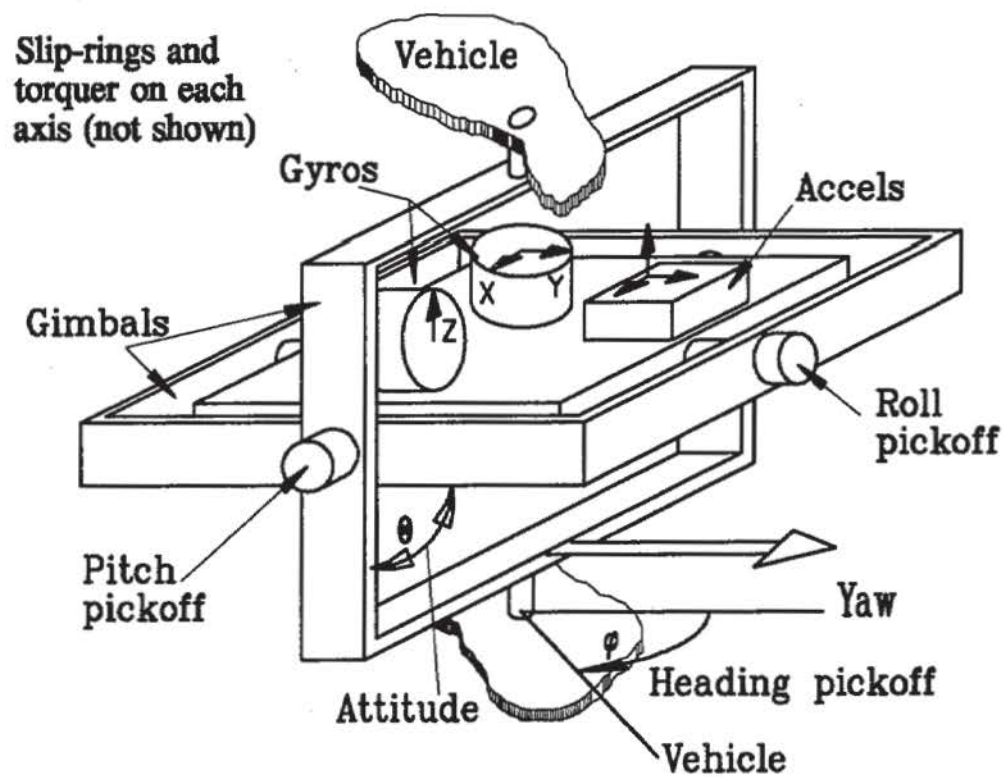


Figure 1.5. A three-axis platform.

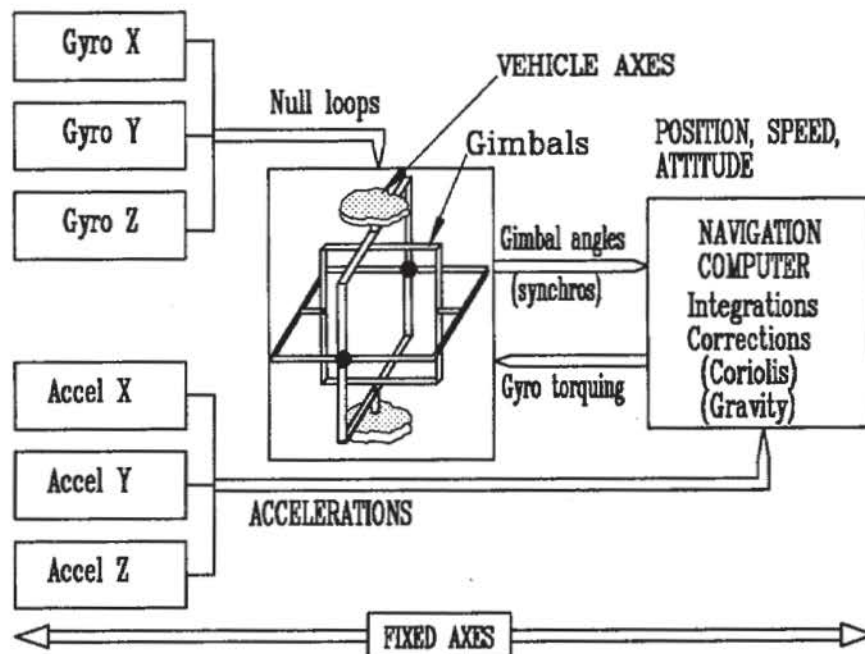


Figure 1.6. The basic platform system.

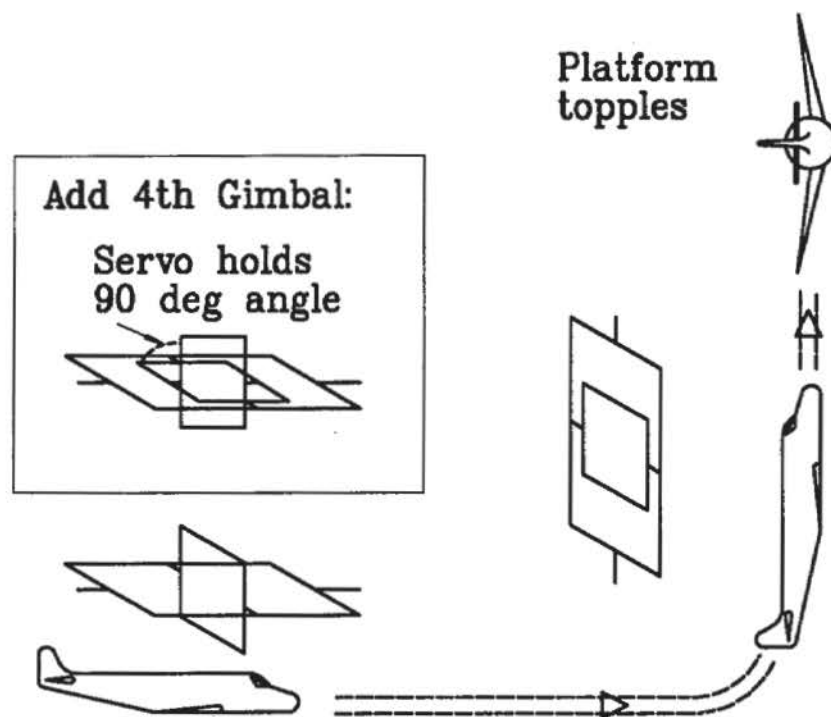


Figure 1.7. Gimbal lock.

provide a vertical reference to set the initial level before takeoff, whereas a flux valve (a kind of magnetic compass) can be used to set the magnetic heading. This heading can be corrected to True North using the local value of the magnetic variation. Alternatively, the heading can be set by gyrocompassing (page 15). Once under way, the HARS provides flight information under all visibility conditions.

Schuler Tuning

Imagine that we have a pendulum hanging in the vehicle that we are navigating, intended to provide a vertical reference. But as we accelerate, the pendulum tilts, giving a false vertical indication. Schuler showed that this would not occur with a pendulum of 84 minutes period, and we can make a compact compound pendulum with this period using an ordinary accelerometer and a servo loop with the correct characteristics. Correcting an inertial system so that it does not tilt when accelerated is known as *Schuler tuning* it.

Gimbal Lock

Three axis platforms have a limitation that can cause them to "lock up" during aerobatic maneuvers. This *gimbal lock* can be caused by a set of aircraft movements that cause two of the gimbal axes to become aligned, followed by a

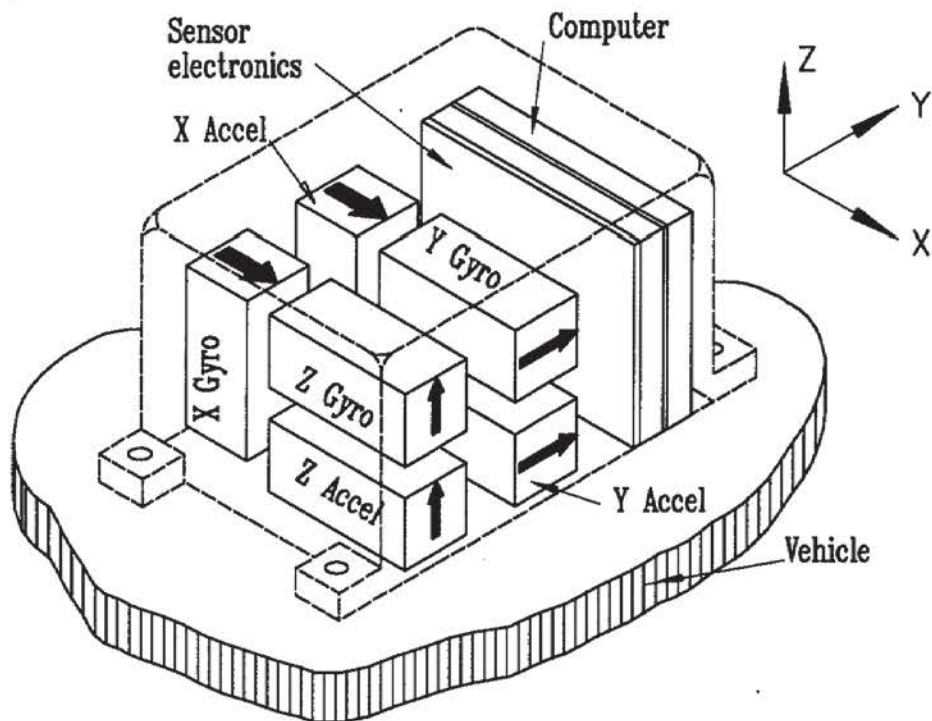


Figure 1.8. The strapdown system

rotation in the plane of the aligned axes. How this happens is much easier to see than to read, so Figure 1.7 shows a sequence that tumbles the platform. A plane flying straight and level climbs vertically. Now the roll gimbal is aligned with the pitch gimbal, and a turn over the wing, around the aircraft yaw axis, forces the pitch gimbal over and makes the system lose alignment. A fourth gimbal [10], a duplicate roll gimbal, can be added to overcome this limitation, although it adds considerably to the system's cost and size.

For further reading, books by Markey and Hovorka [11] and Savant [12] describe inertial systems in more detail.

Strapdown Systems

The platform system's complex mechanical construction goes against the trend in machine design. Since the 1960s, technical advances in electronics and optics have reduced their size and cost, while there has been little progress in reducing the cost of mechanisms, apart from robotic fabrication. Generally, lower costs and higher reliability have come from replacing mechanisms with electronics.

Navigation systems have benefitted from this technological culture; the strapdown system is the outcome. The strapdown system replaces gimbals with a computer that simulates their presence electronically. In the strapdown system the gyroscopes and accelerometers are rigidly mounted to the vehicle structure so that they move with the vehicle, as shown in Figure 1.8. Now, unlike the plat-

form gyros, the strapdown gyros must measure the angles turned, up to the maximum rotation rate expected. Airplanes can experience short-term rates up to 400 deg/s, so a gyro with 0.01 deg/h performance has a dynamic range of 10^8 - two orders above the platform gyro.

As the vehicle travels, the gyros measure the yaw, pitch, and roll angles turned in a short time (say, 0.01 s), and pass them to a computer that uses them to resolve the accelerometer outputs into the navigation axes set. Simplifying to two dimensions, we can easily express the coordinate transformation between the navigation axis set (x_{nav}, y_{nav}) and the body axis set (x_{body}, y_{body}), at a time when the z-axis gyro has measured that there is an angle θ between them. The accelerometers fixed in the body axes record accelerations a_x, a_y . Converting to navigation axes we get

$$(a_x)_{nav} = a_x \cos \theta - a_y \sin \theta \quad (1.1)$$

$$(a_y)_{nav} = a_x \sin \theta + a_y \cos \theta$$

When adding the third (z) axis we must use more complicated transformations as the rotations are non-commutative. Typical systems use direction cosines, and the transformation between the inertial set [X_i, Y_i, Z_i] and the body axes set [X_b, Y_b, Z_b] is expressed as follows:

$$\mathbf{r}^i = \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} = \mathbf{C}_b^i \mathbf{r}^b$$

$$\mathbf{C}_b^i = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

where c_{ij} are the direction cosines between the j^{th} axis in the inertial frame and the i^{th} axis in the body frame; Britting [13] describes the necessary computations. The sensors and computer in a strapdown system are arranged as shown schematically in Figure 1.9.

System Alignment

The first step in inertial navigation is aligning the platform to the navigation axes or, equivalently, initializing the axes in the strapdown system's computer. We can level the platform or measure the attitude of the vehicle for a strapdown system, using accelerometers. We can find North using a magnetic compass (which is not very accurate), by astronomical sightings (which means we must be able to see stars), or by *gyrocompassing*. In some cases, such as that of missiles suspended under the wing of a fighter plane, we can align the missiles' strapdown systems

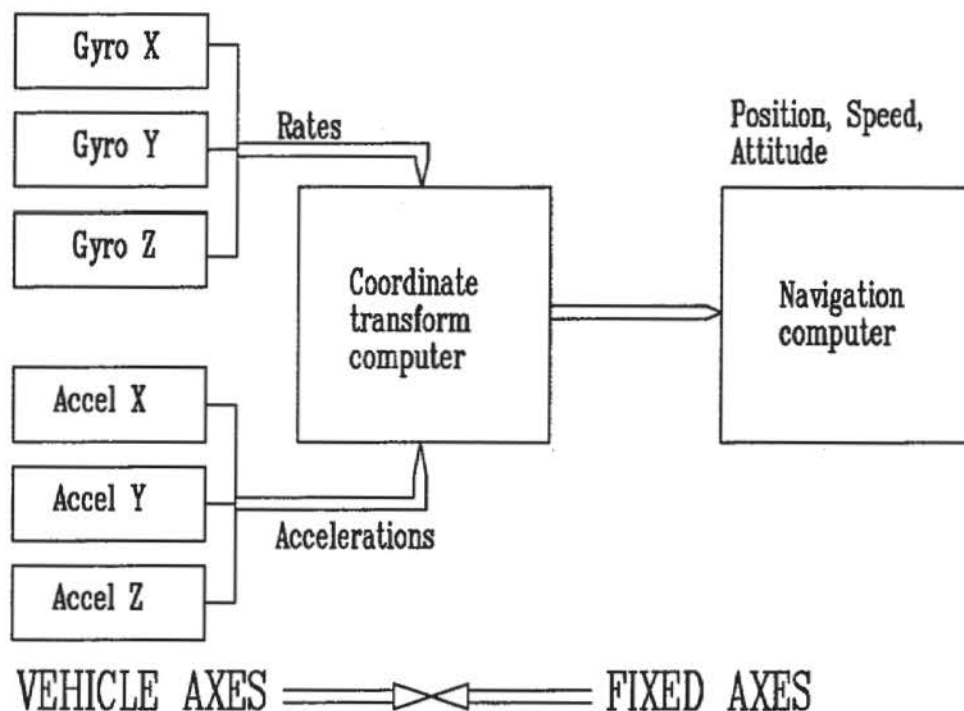


Figure 1.9. The basic strapdown system.

by referring their outputs to those of a master navigator in the fighter, called *transfer alignment*.

Gyrocompassing

To gyrocompass with a platform, a gyro with its input axis horizontal is connected by a servo to a platform torquer that will rotate the gyro about the vertical (yaw) axis. The servo is designed to drive the sensed rotation to zero, so the gimbal will swing around until the gyro's axis points East. But if the gyro has a bias - an output signal when there is no rotation input - it will point a little off East, enough so that the earth's rotation rate component cancels the bias. Then the platform is slewed 180°, and again servoed to null. This time the bias has not changed sign but the earth's rotation component has, so, by combining the East and West readings of platform heading, both North and the gyro bias can be deduced, a big advantage for the platform system. (In passing, note that some systems rotate continuously, the "rotating azimuth" mechanization, so that instrument biases are modulated at the rotation rate and filtered out. This improves performance for given instrument errors and alignment accuracy.)

A strapdown system cannot rotate the gyros except by swinging the entire vehicle, usually impossible to do over the half-circle necessary for East-West gyrocompassing. All one can do is measure the three components of earth's rotation with the gyros and compute where North must be, assuming that the gyro bias has not changed since it was last measured. To find North to an accu-

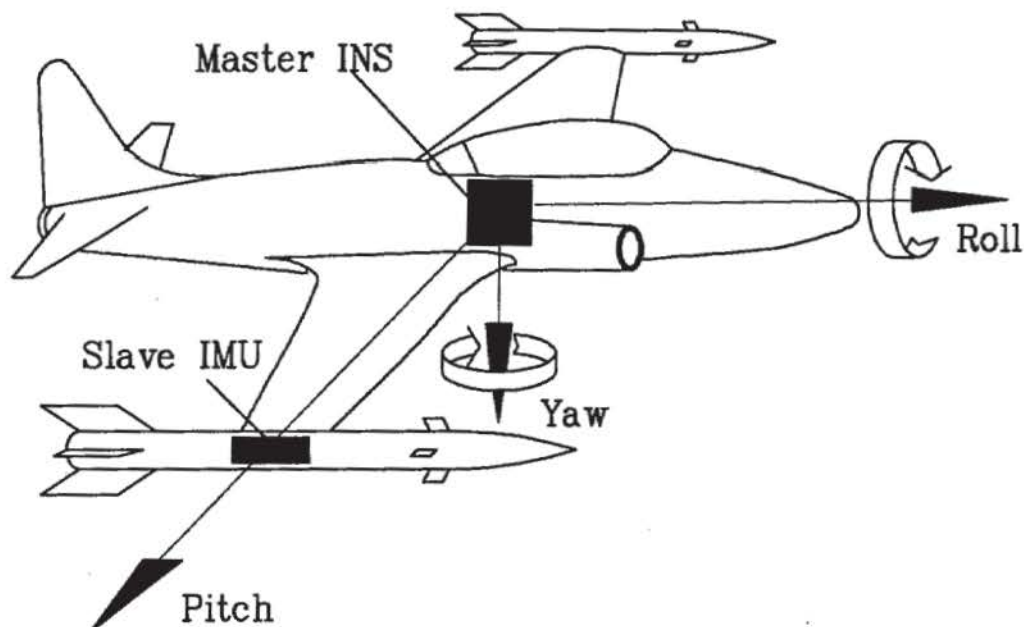


Figure 1.10. Transfer alignment.

racy of 1 mr, one must measure one thousandth of the horizontal earth's rate. At Boston, Massachusetts, for example, $HER = 10 \text{ deg/h}$, so for 1 mr alignment accuracy we must know the gyro bias to 0.01 deg/h , a demanding requirement met only in expensive gyros.

Transfer Alignment

Another way to align a strapdown system is to transfer alignment from a master system. A fighter aircraft can initialize wing-carried missiles, for example, by carrying out maneuvers designed to allow the fighter's master navigator to send alignment information to the missile, as shown in Figure 1.10. If the fighter flies straight and level, the master and slave velocity outputs can be matched by adjusting the accelerometer bias, and controlled rolls or turns can calibrate the slave system's gyros.

Advantages and Disadvantages of Platform Systems

Advantages:

1. Simpler gyros. Because the sensor platform rotates only at the small rates needed to keep it level, the gyros need only a small dynamic range. A

maximum rate of 3 deg/s would suffice for a gyro of 0.01 deg/h performance (e.g., for an aircraft navigator), a range of 10^6 . Further, gyro torquer errors do not lead to attitude error. The lack of gyro rotations means that there are no anisoinertia and output axis angular acceleration errors to minimize in the design.

2. Higher accuracy. Because the accelerometer axes are always well defined, the platform navigator can be very accurate; the North and East accelerometers see no component of gravity and measure only the vehicle accelerations. The vertical accelerometer, though, measures the vehicle's vertical motion in the presence of 1g, therefore less accurately. In an aircraft this causes altitude errors, which can be compensated with a barometer signal.
3. Self-alignment by gyrocompassing.
4. Sensor calibration by platform rotations. The other sensor biases are obtained by orienting the platform with each major axis vertical in turn, provided there is enough time.

Disadvantages:

1. Complexity and cost. The gimbal structure and its bearings must be stiff so that the accelerometer axes remain defined even under vehicle vibrations, but the bearings and slip rings must have as little friction as possible. As a result, the gimbal structure is an elaborate, precisely made mechanism which cannot avoid being expensive.
2. Gimbal magnetics. Each gimbal must have a pickoff and torquer. The pickoffs (synchros) measure the intergimbal angles with arc-sec resolution over a full revolution, a range of 10^6 . When the system is first switched on, the torquers need to provide enough torque to accelerate the platform inertia so that it can gyrocompass and align itself to Level and North in a reasonable time. The torquers also must not leak magnetic flux, for that could upset the sensors as the gimbals move around them.
3. Reliability. The bearings and slip rings tend to wear, degrading alignment and performance.

Advantages and Disadvantages of Strapdown Systems

Advantages:

1. Simple structure, low cost. Strapdown systems are lighter, simpler, cheaper, and easily configured for odd-shaped spaces. As it is only necessary to mount the sensors so that their sensing axes point in known

directions (usually orthogonal) in the vehicle, they can be placed so that they best use the space available.

2. Ruggedness. The simpler structure better withstands shock and vibration, and, being lighter, is easier to shock mount than a platform.
3. Reliability. There are no gimbal magnetics, no slip rings, and no bearings. The electronics that replace them are inherently more reliable.

Disadvantages:

1. Alignment. Strapdown systems are difficult to align because they cannot be easily moved. Transfer alignment is suitable for tactical systems.
2. Sensor calibration. Again, the immobility means that the sensors cannot be calibrated in the system. Therefore, they must be stable, a burden on the sensor design. Strapdown systems rely on sensor models, using real-time compensation of inertial errors and thermal effects.
3. Motion induced errors. The body motions induce unique sensor errors (torquer errors, anisoinertia, output axis angular acceleration), which can be compensated to some degree.
4. Accelerometer errors. Bias errors accumulate, and strapdown accelerometers may be subjected to components of gravity as the vehicle rolls and pitches, reducing the accuracy of the vehicle acceleration measurement and exciting cross-axis errors.
5. The strapdown computer. Not needed in the platform system, the computer must be fast enough to do all the strapdown calculations in a few milliseconds. In a typical tactical system, a bandwidth of 100 Hz demands that sensor compensation and coordinate transformation must be done in less than 0.01 s. This requires well-crafted program code.

The particular benefits of strapdown systems are well illustrated in the following story. One of the first strapdown systems fielded was used in the Apollo Lunar Module (LM) moon lander, as a backup to the primary platform system. The Apollo 13 spaceship had just left the earth when a fuel cell in the main ship (the service module) exploded, drastically reducing the amount of electrical power available. The ship was committed to a trajectory to the moon; it could not change direction much once in space. Mission Control decided to let the ship pass around the moon and come back to earth without landing on the moon. Because of the shortage of electrical power, the astronauts switched off the command module and used the LM. Everything possible was switched off, so that the three astronauts could stay warm enough; even so, the cabin got unpleasantly cold, down to 4°C (40°F). The platform navigator was turned off, and the crew relied entirely on the lower-power strapdown system, operating below its design temperature, for navigation around the moon and back to earth. They then left the LM and used the command module's navigator through re-entry to a safe landing [14,15].

Star Trackers

In applications where a missile may not have time for an accurate ground alignment before firing, or in cases where flight time is very long, the guidance system may be provided with a star tracker to update its alignment. This is a type of telescope with an optical detector at its focus, which can be precisely pointed at a spot in the heavens where a star is known to be. Once flying high enough that the atmospheric turbulence, smoke, and haze do not obscure the view, the tracker locks onto the pre-selected star and uses the star's known position to correct for misalignment in the inertial navigator.

As star trackers are expensive and require a window in the vehicle, they tend to be used only in long range missiles, reconnaissance planes, and long range bombers. They also need a clear view of the sky. Radio systems are all-weather navigation aids. Omega gives 2-4-nautical mile position accuracy but no velocity information, whereas satellite systems provide 15-m position accuracy and velocity data.

Satellite systems have the advantage over ground-based transmitters that they are radiating almost vertically down, so that satellite signals are much less affected by hills and the curvature of the earth. The newest United States satellite system is the Navstar *Global Positioning System* (GPS) and is accompanied by the similar Russian GLONASS system.

The Global Positioning System

The Navstar Global Positioning System (GPS) is a space-based position and navigation system that can provide three-dimensional position to an accuracy a few meters, anywhere on or near the earth, to those with the proper receivers. Although it is primarily a military system and provides the highest accuracy only to those who know its secret codes, it provides a lower accuracy (15-30 m) signal to all users. The precise encrypted "P" code and the coarser clear/acquisition "C/A" signal are transmitted on separate frequencies.

When it is complete, GPS will consist of 18 satellites, 3 each in 6 orbital planes. They will orbit at a height of 20,200 km (10,900 nautical miles) with a 12-h period; the orbits are inclined at 55°, as illustrated in Figure 1.11. The satellites already in orbit are demonstrating that the system will easily meet its design objectives.

In operation, each satellite broadcasts both the P and C/A signals at precisely known times, timed by each satellite's on-board atomic clock. Each receiver has a clock, too, so it can measure the time at which it receives the signal from the satellite. As the speed at which the radio wave travels is known, the distance between the satellite and the receiver is found by multiplying that speed by the time difference between satellite transmission and user reception. Each satellite's

Satellite location is not
accurately drawn

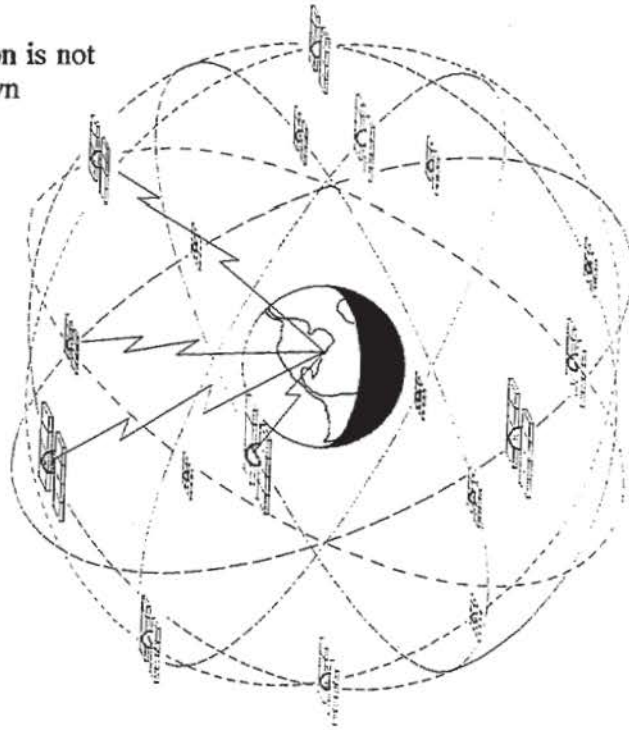


Figure 1.11. The Global Positioning System satellites.

position is known, so the user now knows position along the line of sight to the satellite. Users take data from three satellites at different angles in the sky to fix their position in three dimensions.

Now the user's clock might not be very accurate, and it is to everyone's advantage to allow it to be inexpensive, so each satellite also broadcasts the precise time in its message. As we now need a separate piece of data to decode time, we read data from a fourth satellite and compute position and accurate time. The system has been designed so that four satellites will always be visible from anywhere on the earth.

The satellites' positions can drift slightly in orbit, from irregularities in gravity, from solar pressure and so on, so fixed ground stations observe them and send orbit data to a master control center in Colorado Springs. From there a message is sent to each satellite, telling it how much it is deviating from its predicted orbit, and providing small time corrections to its clock. The satellite incorporates these correction messages (*ephemeris* data) into its transmissions. All of this is naturally done in a computer in the user's receiver, and some companies make receivers small enough to carry in the hand [16,17]. It is possible to make a receiver for both GPS and GLONASS [18,19]. The C/A code is used as a coarse acquisition signal for those privileged with access to the P code, but it is the only signal available to commercial users. However, the C/A code gives 30 m accuracy, so it is still a valuable aid.

GPS will allow more accurate navigation, but will existing charts and maps be accurate enough to realize the benefit? Roeber [20] thinks not. However,

GPS can itself be used for map making and geodetic surveys. It will be used on test ranges to check out inertial navigation systems and will facilitate air traffic control and search and rescue missions. It will change the inertial navigation field for all except strategic weapon guidance, for such a device is exactly what an inertial navigation system needs for a partner.

The INS gives very accurate attitude and distance measurements for a short time, but gyro and accelerometer errors accumulate after a while and navigation errors grow. The GPS system has exactly the opposite characteristic, for it can give accurate position fixes periodically to correct the INS, but cannot tell you quickly if you're turning or rolling; the two together complement one another [21]. Should the vehicle lose the satellite signals for a while, perhaps because the vehicle has banked and a wing hides the antenna from the satellite or because a ground vehicle has entered a tunnel, the INS can carry on until the receiver re-acquires the satellite signal. For military, use the INS can keep a missile on target even if the GPS signal is obliterated by powerful jammers.

Applications of Inertial Navigation

The most accurate systems, with heading error less than 0.001 deg/h (one revolution in 40 years!) are found in submarines and ICBMs, which need navigation with errors measured in tens of feet over long mission times. We might call these "1 mile a day" systems. Their accelerometers resolve gravity to a few μg while measuring vehicle accelerations of the order of 10g, a dynamic range of 10^7 .

Next, aircraft systems, with errors of less than 1 mile an hour, are found in all civil aircraft that cross the oceans, and in long range military aircraft. While all the 1-mile-a-day systems are platform systems using mechanical sensors, laser gyro strapdown systems are dominating the 1-mile-an-hour market now. Less expensive heading and attitude reference systems (HARS) are used in short range military aircraft, and, frequently, in long range planes, to back up the INS. HARS often use mechanical gyros, but laser gyro versions are appearing as laser gyros get cheaper.

Tactical weapons such as short range missiles, Harpoon, Tomahawk, and Phoenix being U.S. examples, exclusively use strapdown systems. They presently use only mechanical sensors, but that will change as the new optical (laser) gyros and micromachined accelerometers complete development. These are described in later chapters.

Conclusions

For a deeper study of inertial systems, the books by Broxmeyer [22], Draper et al. [23], Leondes [24], McClure [25], O'Donnell [26], and Parvin [27] could be

consulted. The guidance of spacecraft uses special techniques, described by Hynoff [28].

GPS will have a tremendous impact on the navigation field, for GPS and strapdown systems go together well. The strapdown processor can support the GPS, and the INS can tell the GPS receiver where to look for satellites and maintain navigation while satellites are obscured.

There is one very important exception, though, and that is the guidance of strategic weapons (including aircraft and submarines). In a war serious enough for strategic weapons to be used, one could not rely on the constellation of GPS satellites to survive. They would either be destroyed or their signals would be jammed. So strategic systems must remain self-contained (perhaps aided with a star-tracker), and the accuracy demanded of these systems is difficult to meet with strapdown gyros.

In the 1960s, when inertial navigation systems began to be used, their instruments were electro-mechanical and very sophisticated; they were expensive, large, and fairly fragile [29]. In the 1980s the gyros were more likely to be optical, using lasers; they were less expensive, smaller, and more rugged. In the 1990s gyros and accelerometers will continue to get cheaper, smaller, and more robust, based on technologies described in the later chapters of this book.

References

1. Duncombe, R.L., R.F. Haupt, "Time and navigation", *Navigation, J. Inst. Nav.*, 17, 4, pp. 381-386, Winter 1970-71.
2. Dunlap, G.D., H.H. Shufelt, *Dutton's Navigation and Piloting*, United States Naval Institute, Annapolis, MD, 1970.
3. Savet, P.H. (Ed.), *Gyroscopes: Theory and Design*, McGraw-Hill, New York, 1961.
4. Allington, P.J.S., "The Sperry Mk 19 gyro-compass", *Symposium on Gyros, Proc. Inst. Mech. Eng. (London)*, 1964-65, Vol. 179, 3E.
5. Draper, C.S., "Origins of Inertial Navigation", *AIAA J. Guidance and Control*, 4, 5, pp. 449-463, Sept.-Oct. 1981.
6. Draper, C.S., "Guidance is forever", *Navigation, J. Inst. Nav.*, 18, 1, pp. 26-50, Spring 1971.
7. Wrigley, W., "History of inertial navigation", *Navigation, J. Inst. Nav.*, 24, 1, pp. 1-6, Spring 1977 (37 references).
8. Hildebrant, R.R., K.R. Britting, S.J. Madden, "The effects of gravitational uncertainties on the errors of inertial navigation systems", *Navigation, J. Inst. Nav.*, 21, 4, pp. 357-363, Winter 1974-75.
9. Paik, H.J., J-S. Leung, S.H. Morgan, J. Parker, "Global gravity survey by an orbiting gravity gradiometer", *Eos*, 69, 48, pp. 1601, 1610-1611, 1988.
10. Fernandez, M., G.R. Macomber, *Inertial Guidance Engineering*, Prentice Hall International, London, 1962.

11. Markey, W.R., J. Hovorka, *The Mechanics of Inertial Position and Heading Indication*, John Wiley and Sons, New York, 1961.
12. Savant, C.R. Jr., R.C. Howard, C.B. Solloway, C.A. Savant, *Principles of Inertial Navigation*, McGraw-Hill, New York, 1961.
13. Britting, K.R., *Inertial Navigation Systems Analysis*, Wiley-Interscience, New York, 1971.
14. Lewis, R.S., *The Voyages of Apollo*, Quadrangle, New York, 1974. Pages 166-168 relate to Apollo 13's return.
15. Kayton, M., "Avionics for manned spacecraft", IEEE Trans. on Aerospace and Electronic Systems, 25, 6, p. 802, Nov. 1989.
16. "Collins demonstrates first hand-held Global Positioning System receiver", *Aviation Week and Space Technology*, 19 June 1989, p. 153.
17. Magellan Hand-held GPS Receiver, *Defense News*, 4 Sept. 1989, p. 21; also *Aviation Week and Space Technology*, 30 Oct. 1989, p. 51.
18. Eastwood, R.E., "An integrated GPS/GLONASS receiver", *Navigation, J. Inst. Nav.*, 37, 2, pp. 141-151, Summer 1990.
19. Klass, P.J., "GPS, Glonass and Glasnost", *Aviation Week and Space Technology*, 5 Oct. 1987, p. 11.
20. Roeber, J.F., "Where in the world are we?", *Navigation, J. Inst. Nav.*, 33, 4, Winter 1986-87.
21. Buechler, D., M. Foss, "Integration of GPS and strapdown inertial subsystems into a single unit", *Navigation, J. Inst. Nav.*, 34, 2, pp. 140-159, Summer 1987.
22. Broxmeyer, C., *Inertial Navigation Systems*, McGraw-Hill, New York, 1964. Good for IN systems mathematical background.
23. Draper, C.S., W. Wrigley, J. Hovorka, *Inertial Guidance*, Pergamon Press, New York, 1960.
24. Leondes, C.T. (Ed.), *Guidance and Control of Aerospace Vehicles*, McGraw-Hill, New York, 1963.
25. McClure, C.L., *Theory of Inertial Guidance*, Prentice Hall, Englewood Cliffs, N.J., 1960. Mathematical treatment, covers Euler's equations, etc.
26. O'Donnell, C.F. (Ed.) *Inertial Navigation Analysis and Design*, McGraw-Hill, New York, 1964.
27. Parvin, R.H., *Inertial Navigation*, Van Nostrand, New York, 1962. Good, readable account of the basics.
28. Hynoff, E., *Guidance and Control of Spacecraft*, Holt, Reinhart and Winston, New York, 1966.
29. Slater, J.M., *Inertial Guidance Sensors*, Reinhold, New York, 1964.

2

Gyro and Accelerometer Errors and Their Consequences

Inertial navigation is an "initial value" process in which a vehicle's location is deduced by adding distances moved in known directions to the known position of the starting point. Errors in the deduced location come from imperfect knowledge of the starting conditions, from errors in the strapdown computation, and from errors in the gyros and accelerometers. In Chapter 1 we alluded to gyro bias when describing gyrocompassing, without defining it; here we will consider sensor performance more rigorously. For the most part we will follow the terminology used in IEEE Standard 528 [1].

The errors caused by a misaligned system, or one with uncompensated instrument biases, depend on the configuration of the inertial system, be it strapdown or platform, the number and type of gyros and accelerometers, and how the gimbals are arranged. Britting [2] has provided a unified error analysis for terrestrial inertial systems and used it to compare performances under different conditions. However, it is useful to have some rules of thumb to give approximate values for system errors caused by deficient instruments or improper setup. We often find it necessary to "think on our feet" and bound a problem with such simplifications, so we will now develop a few of these rules, leaving those who need precision to refer to Britting's book.

Effect of System Heading Error

Given a misalignment α in heading, at speed v and travel time t the cross-track distance error, e_c , is

$$e_c = \alpha vt$$

For 1 mr error (about 3 arc-min) and a speed of 500 mph, this error will be 0.5 mph. In an aircraft this same error in pitch attitude would make the plane climb or dive, which would be indicated by the barometric altimeter and corrected by

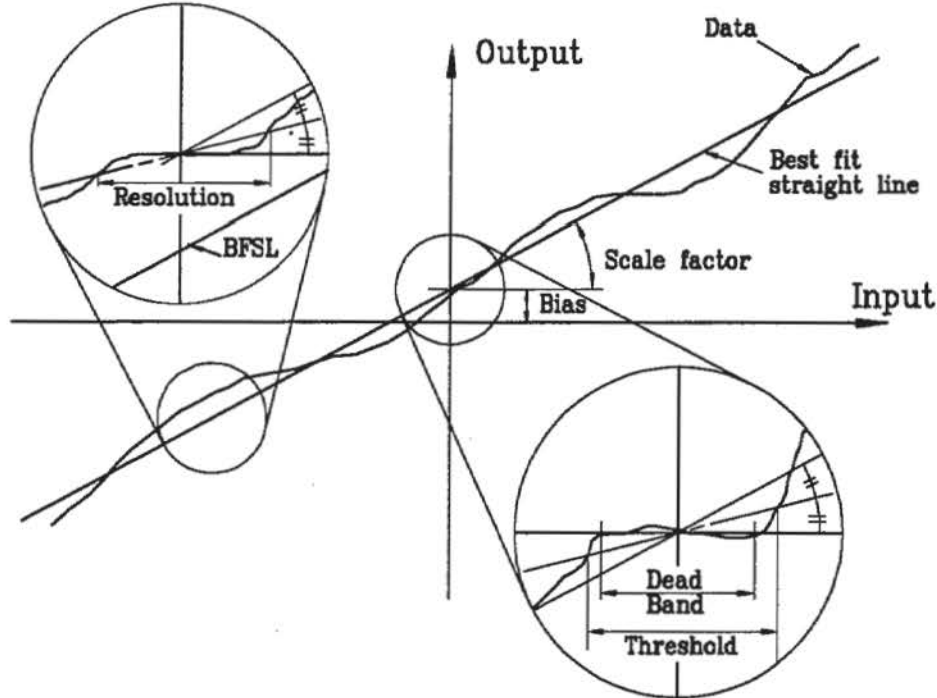


Figure 2.1. Sensor output errors.

the autopilot. As 1 deg is approximately 17 mr, a gyro with a 60-deg/h drift rate allows heading to drift by 17 mr/min.

Scale Factor

An inertial sensor provides an output signal in response to either rotation or acceleration. The signal might be analog, perhaps a voltage proportional to the input, or digital, like frequency or a binary number. As most systems are now controlled by digital computers, digital output is preferred, and outputs from analog sensors are digitized by A-to-D converters.

The scale factor is the ratio between a change in the output signal and the change in input. As most sensors provide an output signal directly proportional to the input, the scale factor is a single number, defined as the slope of the best straight line fitted by the method of least squares to the data obtained by varying the input over a specified range (Figure 2.1). The scale factor K is

$$K = S/I \quad (2.1)$$

where

S = output signal

I = true input

Digital sensors' output is described by $1/K$, the *pulse weight*; it might have units of arc-sec/pulse. However, the inverse of the scale factor is sometimes quoted in specifications, even called "scale factor," using units of deg/h/mA, deg/h/Hz, g/Hz, and so on. In this book we will use K defined by Equation 2.1 (and IEEE STD 528) as the scale factor. Sometimes an instrument will have a different scale factor for positive and negative inputs, known as scale factor *asymmetry*.

Non-linearity and Composite Error

The scale factor may not be exactly linear but may have second or higher order terms relating signal to input. To see these it is usual to run a test in which the acceleration or rotation rate is varied in steps, and the sensor output measured. The data are fitted to a straight line by the least-squares method, and the residuals plotted. Residuals are the differences between the actual outputs at a given rate (acceleration) and the value which would be predicted using the calculated scale factor. If the residuals are randomly scattered about the axis, no further modeling is possible, but if they form curves, second or third order curve fits can provide a better model for the scale factor. The standard deviation of the residuals, called the *standard error*, can be used to specify instrument quality. The *composite error* is the ratio of the largest residual to the full scale range, and combines errors due to hysteresis and resolution, among others.

System Error from Gyro Scale Factor

Gyro scale factor errors are less important in platform systems than in strap-down systems. In platforms the gyro commanding system heading requires no more than 0.1% accuracy. But in strapdown systems the gyro scale factor determines system accuracy directly. For example, if a mission involves an airplane circling for long periods in the same direction over a point, as it might if it were tracking a submarine at sea, the Down axis gyro scale factor controls heading error. The heading error after time t , while circling at rate Ω , due to scale factor error $\Delta K/K$ is

$$e_h = (\Delta K/K)\Omega t$$

Let $\Omega = 1 \text{ deg/s}$ (one circuit in 6 min), and let $\Delta K/K = 100 \text{ ppm}$. Then

$$e_h = 0.36 \text{ deg/h}$$

Asymmetry

Small values of asymmetry, of the order of 100 ppm, are usually acceptable, except in applications where the instrument spends a lot of the mission oscillating about null, for then the system can rectify the asymmetry into a bias. The effective scale factor is

$$K = K^+ \zeta^+ + K^- \zeta^-$$

where

K^+ = scale factor for $\Omega > 0$

K^- = scale factor for $\Omega < 0$

$\zeta^+ = 1$ and $\zeta^- = 0$ for $\Omega > 0$

$\zeta^+ = 0$ and $\zeta^- = 1$ for $\Omega < 0$

Bias

Generally, when there is no input, the signal is not zero; there is some offset caused by manufacturing imperfections, called the *bias* or *zero offset*. When we introduce bias B we write

$$S/K = B + I \quad (2.2)$$

For mechanical gyros, bias is sometimes called the *acceleration insensitive drift rate*. For example, bias can be due to the torque from the mechanical null of the instrument suspension (e.g., the hinge in an accelerometer) not coinciding with the pickoff null, or to magnetic contamination of a sensing element interacting with a magnetic field. If the system must quickly find North, the horizontal gyros and the leveling accelerometers must have repeatable biases from turn-on to turn-on.

System Error from Accelerometer Bias

Accelerometer errors accumulate into distance errors according to the equation of motion

$$s = ut + \frac{1}{2}at^2 \quad (2.3)$$

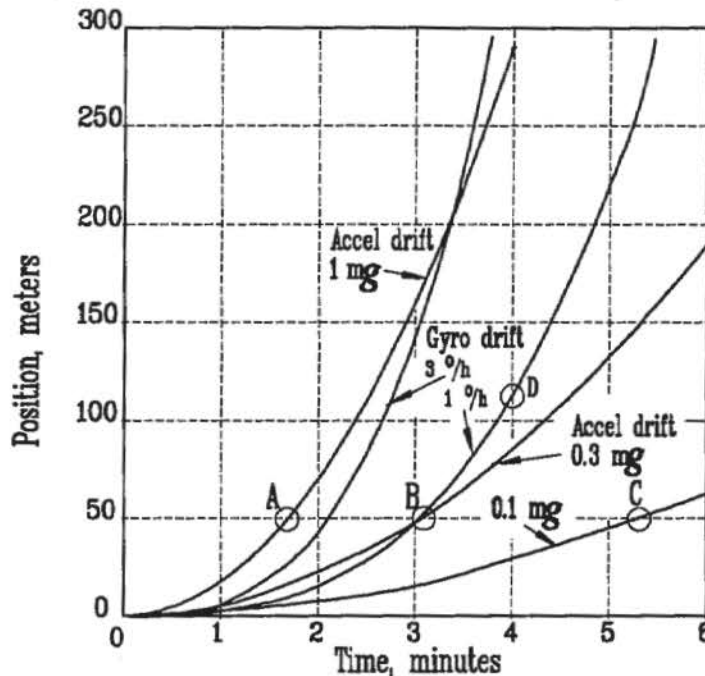


Figure 2.2. Sensor-induced position error.

where

u = initial velocity

a = acceleration

s = distance traveled in time t

An uncompensated accelerometer bias of 1 mg in a journey of 1 h (starting from rest, $u = 0$) gives a distance error $s = 65$ km.

A transoceanic aircraft navigator must have much better accelerometer compensation than that to be useful. This level of performance is closer to the needs of tactical weapons navigators, or of GPS aided navigators.

Figure 2.2 shows the position errors caused by some specific accelerometer and gyro errors. Point A shows the position error of 50 m due to a 1 mg accelerometer bias error over a time of 100 s, calculated from (2.3), and points B and C show the times taken to accumulate a 50-m error with better accelerometers, with bias errors of 0.3 and 0.1 mg.

Generalizing these calculations we can see that a bias error gives a velocity error e_v and a position error e_p

$$e_v = 0.589 \text{ m/s/mg/min}$$

$$e_p = 17.66 \text{ m/mg/min}^2$$

Note that a 10-mg accelerometer error causes a 6-m/s velocity error in 1 min, so external speed measuring devices like ships' logs, airspeed indicators, and Doppler radars are valuable for bounding this error source.

Tilt Misalignment

A system tilt misalignment β in roll or pitch will be indistinguishable from an accelerometer bias error, and the equivalent bias is given by

$$B_t = g\beta$$

Therefore, the position error e_t due to level tilt about North and East is

$$e_t = \frac{1}{2}B_t t^2 = \frac{1}{2}g\beta t^2$$

System Error from Accelerometer Scale Factor Error

Accelerometer scale factor error causes navigation errors only when the vehicle is accelerating (or sensing gravity, of course). Assume that for a mission the acceleration is less than $2g$ most of the time. A 1% scale factor error will give an acceleration uncertainty of 20 mg, which is indistinguishable from a bias, and from the bias calculations above the distance error is 1.3 km/h. It is clear that most missions would need better than 1% scale factor accuracy, by a factor of 10 at least.

System Error from Gyro Bias

A gyro bias is a change in angle over time, a time-varying misalignment which cross-couples into the accelerometer channel. Figure 2.3 illustrates a situation where the z-axis gyro bias B_z causes a misalignment and couples an acceleration along x into the y-accelerometer. Then the z-gyro error angle, ϵ_z is

$$\epsilon_z = B_z t$$

and the y-accelerometer error is

$$e_y = a_x \epsilon_z = a_x B_z t$$

In this expression the gyro error angle is in radians and is assumed to be small. Integrating the y-acceleration over time, the y-velocity error is

$$e_{vy} = \frac{1}{2}a_x B_z t^2$$

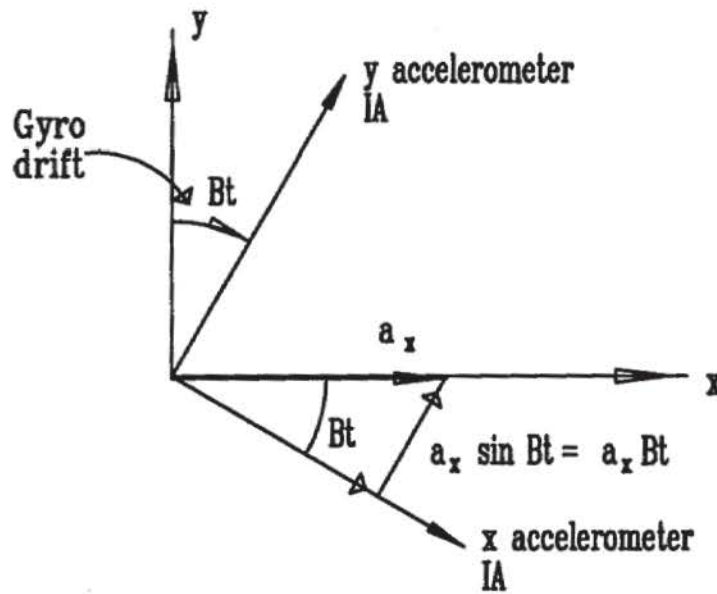


Figure 2.3. Gyro bias causes acceleration error.

Integrating again over time, the y-distance error is

$$s_y = a_x B_z t^3 / 6$$

Because of the t^3 dependence, gyro drift quickly builds up distance errors. As an example let us calculate the position error caused by a 1-deg/h gyro bias error over 4 min, when $a_x = 1g$.

$$s_y = 9.8 \text{ m/s}^2 \times 1 \text{ deg/h} \times (4 \times 60)^3 / 6 \quad (\text{irrational units})$$

Now 1 deg/h = 1 arc-sec/s, and 1 arc-sec = 5×10^{-6} radian, so

$$s_y = (9.8/6) \times 240^4 \times 5 \times 10^{-6} = 113 \text{ m}$$

This point is shown as D in Figure 2.2.

Random Drift

If the sensor is allowed to run on a stable base, sometimes referred to as a *tombstone test*, its output will wander some small amount due to disturbances inside

the sensor. Such a disturbance could be ball bearing noise in a mechanical gyro; it is called *random drift* and can be characterized by the standard deviation of the output measured periodically for some specified time. Mechanical sensors can have characteristic random drifts caused by their construction; the noise may have peaks at particular frequencies. Mechanical gyros with ball bearing wheels have output noise characterized by the ball size, the number of balls, and the operating speed, and by low frequency beats due to the slight differences between the almost identical bearings at each end of the spindle.

These are usually called *in-run* drifts, to distinguish them from turn-on uncertainties. They may depend on temperature and may change as the sensor ages (particularly in ball bearing gyros).

Random Walk

Optical gyros characteristically have an approximately white noise rate output, which causes a long term growth in angle error, called *random walk*. Their root mean square random drift will decrease as the square root of the periodic sample time, as described in Chapter 11.

Dead Band, Threshold, and Resolution

Sensors have a lower limit below which they cannot detect small changes in input, which can be regarded as a dead band around null. Perhaps they rely on movement between parts which stick together until a critical load is exceeded, colloquially known as *stiction*; mechanical sensors can suffer from stiction. Or perhaps the sensor output is noisy, and as the sensor noise output represents an uncertainty in input, small changes in input will be masked by the noise; optical gyros suffer from this deficiency.

No matter what the cause, we define the *threshold* as the largest value of the minimum input that produces an output of at least half the expected value. We call the minimum measurable input the *resolution*, and we define it as the largest value of the minimum input which produces an output of a specified proportion of the value expected using the scale factor. These characteristics are shown in Figure 2.1. Whereas threshold applies only around zero input, resolution affects reading accuracy at all input levels.

Hysteresis

Figure 2.4 shows a sensor's output with an exaggerated separation between the trace taken with input always increasing, and that with it always decreasing. The

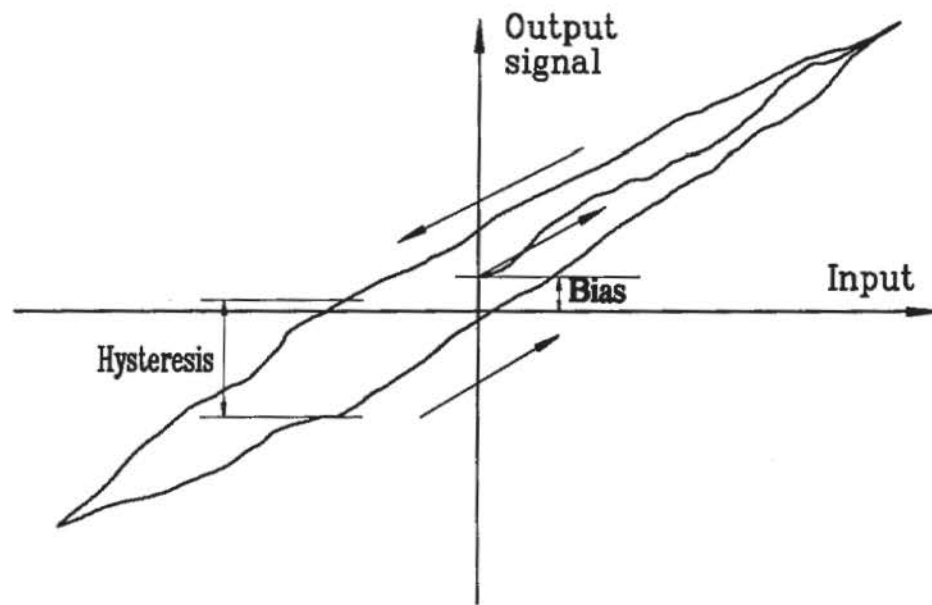


Figure 2.4. Output hysteresis.

vertical separation between the traces is due to hysteresis and the largest separation is defined as the hysteresis error.

Day-to-Day Uncertainty

The scale factor and bias may vary from day to day by as much as 10 times more than their in-run random drifts. They may slowly change over time, perhaps from the aging of internal components such as magnets and bearings, from laser gas contamination, or from mirror or optical fiber aging. They may depend on the temperatures experienced in storage, on storage vibration and shock, on magnetic fields, and so on. Consequently a user will specify the storage and operating conditions, and the sensor supplier will be expected to demonstrate that the sensor family will maintain its calibrated scale factor and bias over a long period when exposed to these conditions.

The foregoing errors are summarized graphically in Figures 2.5 and 2.6, which illustrate bias and scale factor sensitivities to time and temperature.

Gyro Acceleration Sensitivities

Gyros measure rate, and accelerometers measure acceleration. Unfortunately, gyros also sometimes respond to acceleration, and accelerometers sometimes

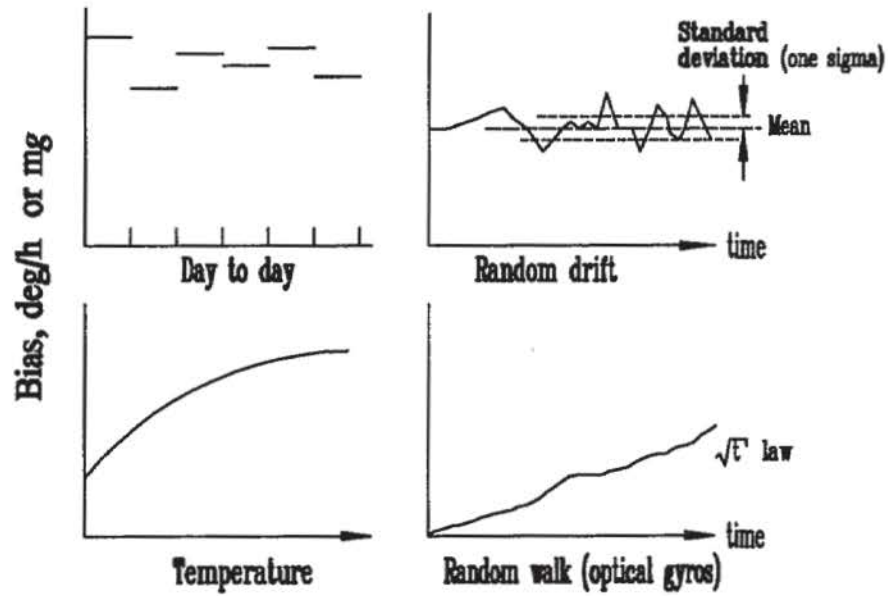


Figure 2.5. Gyro and accelerometer bias.

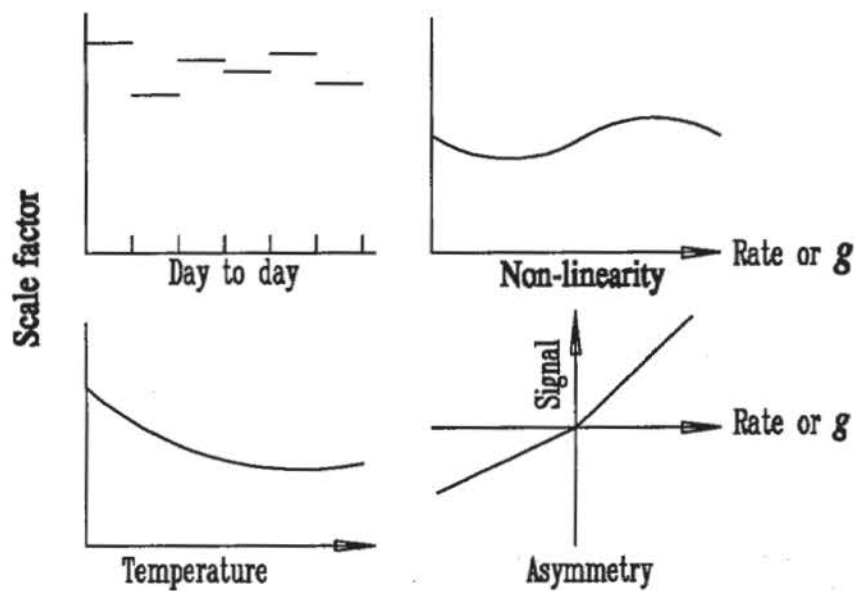


Figure 2.6. Gyro and accelerometer scale factor.

respond to rotation. The principal gyro g -sensitivities, *mass unbalance* and *anisoelectricity*, are described in this section, and rotation sensitivities are described in the next section.

g-Sensitivity

The gyro sensitivity, called colloquially "mass unbalance" (despite attempts to rename it *imbalance*), is due to imperfections in fabrication caused by assembly tolerances. Mass unbalance is linear in acceleration, with units of deg/h/g, but there are other first order g -sensitive terms which are not mass unbalances. (We will describe one, the DTG quadrature term, in Chapter 9.) Provided the value is small enough and stable, g -sensitive terms can be compensated in a navigation system, because we know the accelerations acting along the axes.

Anisoelectricity

Mechanical gyros also output a signal component proportional to the square of the acceleration, called the *anisoelectricity* or the *acceleration-squared drift rate*, caused by mismatches in the stiffnesses of the members supporting the sensing element. A simple model comprising a mass supported on two springs illustrates anisoelectricity (Figure 2.7). If the model is accelerated at an oblique angle, and if the springs are not equally stiff, the mass will move off the line of action of the acceleration and there will be a torque about the normal to the plane of the figure. Referring to Figure 2.7, the torque magnitude is

$$T = \frac{1}{2}m^2a^2[1/K_x - 1/K_y] \sin 2\alpha \quad (2.4)$$

where

m = supported mass

a = acceleration

K_x, K_y = support stiffnesses in the x and y directions

α = angle between the acceleration and the x -axis

Equation (2.4) has a maximum value when $\alpha = 45^\circ$. This equation is sometimes written in terms of the spring *compliances*, the reciprocals of the stiffnesses.

If the acceleration is a vibration at a frequency well below the system natural frequency, the displacement d (Figure 2.7) changes sign in phase with the sign change in the inertia force F , so that the torque is rectified.

Assume $\alpha = 45^\circ$, and let $a = a_0 \sin \omega t$. Then

$$T(\omega) = \frac{1}{2}m^2a_0^2 \sin^2 \omega t [1/K_x - 1/K_y]$$

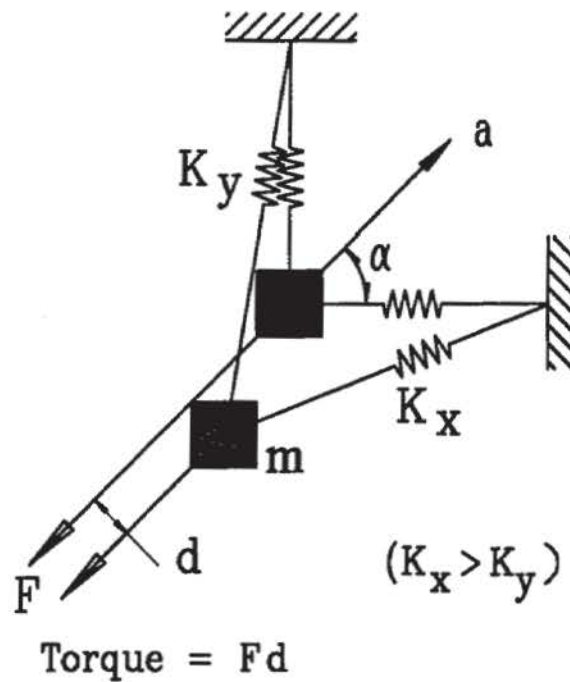


Figure 2.7. Anisoelastic torque.

which has an average value of

$$T(\omega)_{\text{ave}} = \frac{1}{4}m^2a_0^2[1/K_x - 1/K_y]$$

Rotation Induced Errors

These are the angular acceleration sensitivity and the aniso inertia error. Angular acceleration can cause errors in strapdown systems, where the gyros and accelerometers experience the vehicle rotations. Mechanical sensors, being made of parts with finite mass, experience inertia forces when they are accelerated, which, if they act at some distance from a pivot, produce a torque which can be indistinguishable from a rotation or acceleration signal. This rate-induced error is caused by a net centrifugal torque arising from asymmetries in gyro or accelerometer component inertia. A rotating part tries to turn itself so that it rotates about the axis of greatest moment of inertia.

Angular Acceleration Sensitivity

Figure 2.8 shows how a pendulous accelerometer senses angular acceleration about its hinge axis. While an instrument is sensing a steady, non-accelerating rotation it gives the correct output, but under angular acceleration it gives a tran-

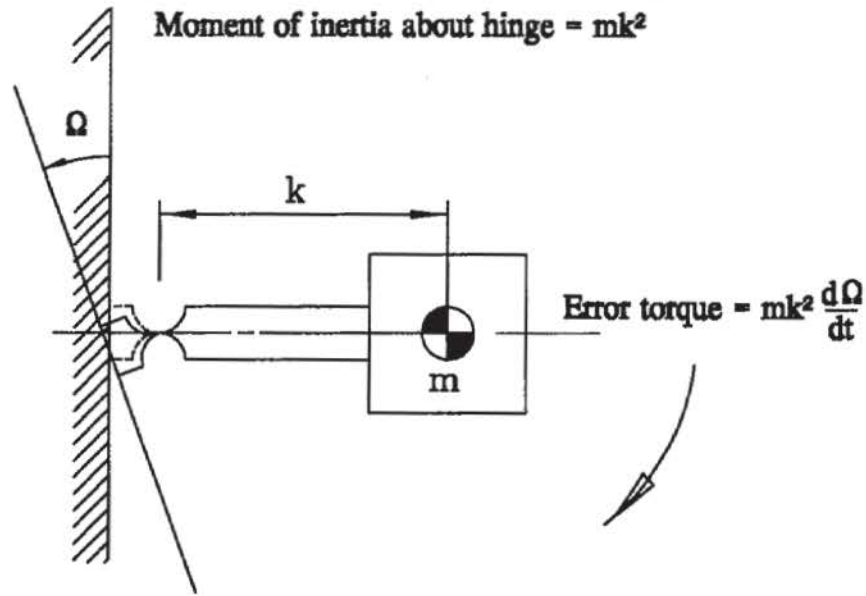


Figure 2.8. Accelerometer angular acceleration error.

sient rate or acceleration error. Defining the *pendulosity* as the product of the sensing mass (m) and the distance from the center of mass to the pivot or hinge (k),

$$p = mk \quad (2.5)$$

we can express the angular acceleration error, e_a as

$$\begin{aligned} e_a &= (I_o/p)(d\Omega/dt) \\ &= (I_o/mk)(d\Omega/dt) = (mk^2/mk)(d\Omega/dt) = k(d\Omega/dt) \end{aligned} \quad (2.6)$$

where

I_o = moment of inertia about the output axis = mk^2

Ω = rate of turn about the output axis

Equation (2.6) tells us that the shorter the pendulum, the better. Of course, a non-pendulous accelerometer—one where the mass translates rather than pivots—will not have an angular acceleration sensitivity.

In an inertial measurement unit we know the rotation rates and accelerations about each axis, and we know the moments of inertia of the sensor components, so we can estimate compensations for angular acceleration error, even though the indicated angular accelerations are corrupted by the accelerations themselves.

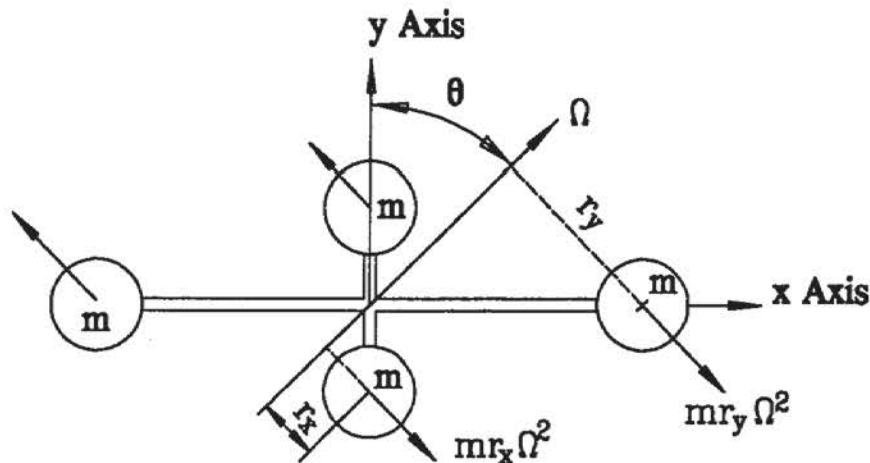


Figure 2.9. Anisoinertia torque.

Anisoinertia

Figure 2.9 shows a structure with unequal moments of inertia about the x- and y-axes being rotated about an axis in its plane. The centrifugal accelerations $mr_x \Omega^2$ and $mr_y \Omega^2$ are unequal, giving rise to a net torque T trying to drive θ to zero

$$T = \frac{1}{2}(I_x - I_y)\Omega^2 \sin 2\theta \quad (2.7)$$

where

I_x, I_y = moments of inertia about x- and y-axes

Ω = rotation rate

θ = angle between rotation and x-axis

Equation (2.7) is sometimes written in terms of the x and y rotation components

$$T = (I_x - I_y)\Omega_x \Omega_y$$

As with the anisoelasticity, the fact that both Ω_x and Ω_y change sign together leads to the rectification of oscillations lower than the resonant frequency.

Angular Accelerometers

If the double integration of linear acceleration gives distance, why not use angular accelerometers instead of gyros, and integrate their output twice to get the angle turned? Every few years this question arises, particularly when someone is brooding over the complexity and cost of gyros, and yearning for the mechanical simplicity of an angular accelerometer [3].

One problem with angular accelerometers is that they have a lower limit to the frequency that they can measure. At high frequencies they measure well, and so are used in autopilot stabilization systems, for example, where vehicle attitude control comes from some other reference. They are used in camera and gun stabilization systems, but they generally cannot be used for navigation. The problem is that they must have an extremely low sensing threshold, for if a vehicle makes a long, large radius turn, or worse, starts a very slow roll, the angular acceleration goes unmeasured if it is below the sensor threshold, meaning that the vehicle's heading can drift severely, at an increasing rate. The vehicle could even roll over, the roll gradually getting faster, quite unknown to the angular accelerometers! Whereas this same problem can occur if a gyro sticks so that its threshold rises, the angular velocity will be bounded, and anyway, the same friction torque will have thousands of times less effect in a gyro than in an angular accelerometer.

Angular Accelerometer Threshold Error

A vehicle using an angular accelerometer could be accelerating around an axis at a value just below threshold and the vehicle's guidance system would be oblivious to it. In a time t the angular velocity ω would build to

$$\omega = \alpha_t t \text{ rad/s}$$

where α_t = threshold, rad/s^2 . The angle error or drift is then

$$\theta = \frac{1}{2}\alpha_t t^2$$

So while the gyro drift angle builds linearly with time, that of the angular accelerometer builds as its square. The cross-track error e_c due to the angular accelerometer's threshold is

$$de_c = \theta v \, dt = \frac{1}{2}\alpha_t v t^2 \, dt$$

where v = vehicle velocity. Integrating over time,

$$e = \alpha_t v t^3 / 6$$

This expression can be used to judge the feasibility of using an angular accelerometer in a navigator, by substituting threshold and mission parameters.

The Statistics of Instrument Performance

Experience has shown that the characteristics of a gyro or accelerometer are not stable; they vary from turn-on to turn-on, and during operation. These errors are bounded, and the *variance* describes the likelihood of a parameter falling within a given range. The day-to-day repeatability parameter might be determined by taking a set of measurements on a given instrument at the same temperature first thing each day, the overnight storage conditions having been specified. The distribution of the measurements will be described by the square root of the variance, the standard deviation, s

$$s = \left[\frac{\sum_{i=1}^n (X_i - X_m)^2}{n-1} \right]^{1/2}$$

where

n = number of measurements

X_i = value of the i^{th} measurement

X_m = mean value of the X_i

Often the data are assumed to be normally distributed, and the standard deviation is identified with the 1σ value of the normal distribution, implying that there is a 68% chance that the bias will be within the band $X' \pm \sigma$ at any turn-on. The 3σ value is often taken as the absolute limit of the parameter, for there is a 99.7% chance that the bias (in our example) will be within $X' \pm 3\sigma$ at any turn-on.

For economic reasons, the instrument's characteristics must be determined from a small sample of tests, so the statistics are applied to verify that the instrument is truly a member of the population with the desired performance. Thus there is reasonable certainty that it will not exhibit larger errors during its life, unless it is wearing out. This is where reliability tests are valuable, for they will characterize the sensor life or Mean Time Between Failures (MTBF), and MTBFs for all the inertial system components can be combined to predict the time probability of system failure. The sequence of tests carried out to verify that the instrument is typical of the population is commonly known as the Acceptance Test Procedure (ATP), and it will include tests over time, temperature, vibration, shock, magnetic fields, and anything else which might make trouble. We will describe accelerometer and gyro testing in Chapter 15.

Table 2.1. Sample Inertial Sensor Specification.

Accelerometer		
Maximum acceleration	<i>g</i>	50
Bias stability		
Day-to-day	mg r.m.s.	1
In-run random	mg r.m.s.	0.1
Scale factor stability		
Day-to-day	%FS	0.05
Nonlinearity	ppm	100
Asymmetry	ppm	100
Cross-axis sensitivity	mg/g ²	1
IA stability	mr r.m.s.	0.5
Bandwidth	Hz	100
Gyroscope		
Maximum rate	deg/s	400
Bias stability		
Day-to-day	deg/h r.m.s.	10
In-run random	deg/h r.m.s.	1
Scale factor stability		
Day-to-day	%FS	0.05
Nonlinearity	ppm	500
Asymmetry	ppm	100
<i>g</i> -sensitivity	deg/h/ <i>g</i>	1
<i>g</i> ² sensitivity	deg/h/ <i>g</i> ²	1
IA stability	mr r.m.s.	0.5
Bandwidth	Hz	100
Environments		
Operating temperature	C	-54 to +85
Vibration level	<i>g</i> r.m.s.	30
Frequency spectrum	Hz	20-2000
Shock level	<i>g</i>	300
Duration	ms	3
(Shape, half-sine)		

Typical Instrument Specifications

By way of example, Table 2.1 gives specifications which describe instruments that could be used in a GPS aided navigator, or in a mid-course guidance system for a missile with a flight time of a couple of minutes. They do not apply to any particular system, nor are they complete. Typically an instrument specification

will run to 30 pages to cover all the operating and storage conditions, as, for example, described in the specification format for a single-axis rate gyro [4]. In a real application there would be more detail about the environment, possibly calling up a military standard [5]. Limits to some cross-coupling, anisoinertia, and higher order terms might be specified, particularly if the chosen sensor type has unusually high response to a particular input. For example, it is necessary to specify acceleration sensitivities for mechanical gyros but not for laser gyros, and laser gyros exhibit random walk while mechanical gyros have different noise characteristics, perhaps caused by the wheel bearings.

In subsequent chapters we will combine the error terms defined in this chapter into individual error models for the different kinds of accelerometers and gyroscopes which we will be considering, and then we will describe the procedures used to measure their magnitudes. While we have covered those error terms which are generally important, in special applications and with individual instrument designs other errors may become significant. Some instruments are sensitive to magnetic fields, and some are not; some generate high frequency noise and some low. We will indicate the characteristics of each instrument type as we discuss them.

References

1. ANSI/IEEE STD 528-1984. Standard Inertial Sensor Terminology.
2. Britting, K.R., *Inertial Navigation Systems Analysis*, Wiley-Interscience, New York, 1971.
3. Hartzell, R.E., "A high reliability, low cost, fluid inertia angular accelerometer and its application in real time inertial coordinate frame of reference systems," DGON Symposium Gyro Technology, Stuttgart, 1985.
4. IEEE STD 292-1969. Specification Format for Single Degree of Freedom Spring Restrained Rate Gyros.
5. MIL-STD-810. Environmental Testing, Aeronautical and Associated Equipment.

3

The Principles of Accelerometers

In this chapter we consider acceleration measurement and examine the dynamic behavior of a common accelerometer, the mass-spring second order model, describing its responses to an impulse and to a sustained periodic driving force. We will describe open- and closed-loop (servoed) instruments and the types of servos which they can use. We will also mention the principles of two new accelerometers, the surface wave and the fiber optic types.

A body's mass is the quantity of matter comprising it; it does not depend on the reference frame [1]. Accelerometers operate by measuring the inertia force generated when a mass accelerates. The inertia force might deflect a spring, whose deflection can be measured; it might change the tension in a string and change its vibrating frequency; or it might generate a torque which will precess a gyro, the measure of acceleration being the gyro's precession rate (as in the pendulous integrating gyro accelerometer, Chapter 7).

The Parts of an Accelerometer

Accelerometers are made up of at least three elements:

1. a mass, often called the "proof mass,"
2. a suspension, which locates the mass, and
3. a pickoff, which puts out a signal related to the acceleration.

It is not always obvious where these elements are. Take a piezoelectric block with two opposite faces plated, for example, and measure the voltage across the plates when it is accelerated. The mass is obvious; it is the block. The spring is the piezoelectric's bulk elasticity, and the pickoff is the piezoelectric effect itself. (Such instruments are used to measure vibration; they are not sensitive enough for most navigation systems.)

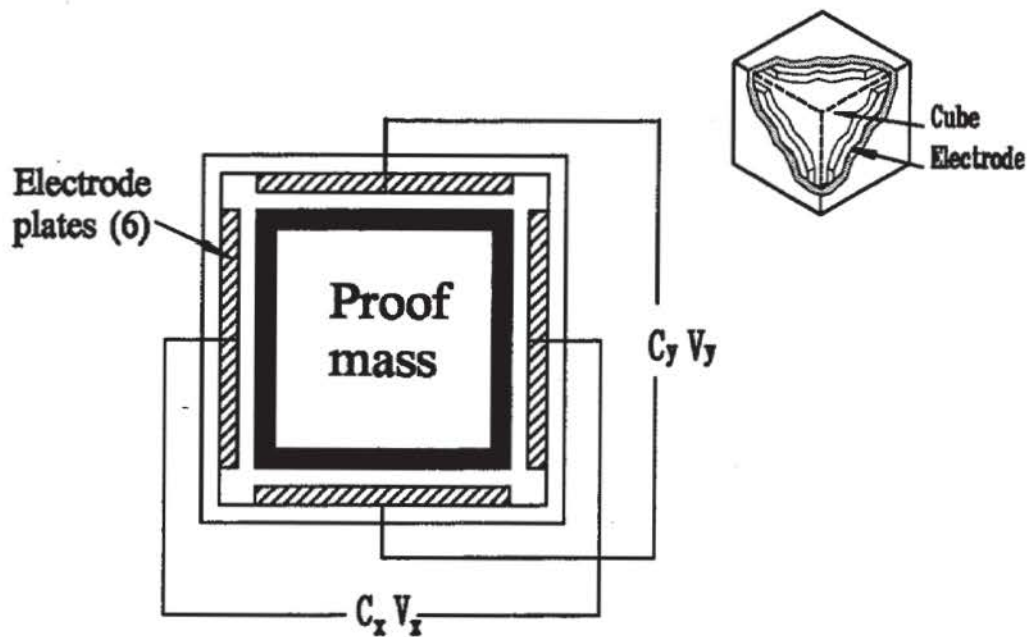


Figure 3.1. An electrostatic accelerometer.

Some instruments are servo controlled and will need, in addition:

4. a forcer, an electric or magnetic force generator designed to oppose the inertia force, and
5. an electronic servo loop.

As another example, Figure 3.1 shows a metallic cube inside a box, with six electrode plates each spaced a small distance from a cube face. The cube may be suspended electrostatically by charging the plates with voltages V_x , V_y , and V_z . The capacitance of each plate to the cube (C_x , C_y , C_z) is a measure of the gap between the plate and the cube, and serves as a pickoff. Three servos, one per pair of opposite plates, can adjust the plate voltages so that the cube is supported with equal gaps on all faces, and the voltages are a measure of the acceleration. The cube is the proof mass, and the servo-controlled electrostatic field is both the suspension and the forcer; this is a *closed-loop* sensor. A French group [2] used such an instrument to study small acceleration effects in spacecraft.

The Spring-Mass System

The spring-supported mass shown in Figure 3.2 is a basic single degree of freedom accelerometer, and the relationships between the size of the proof mass, the

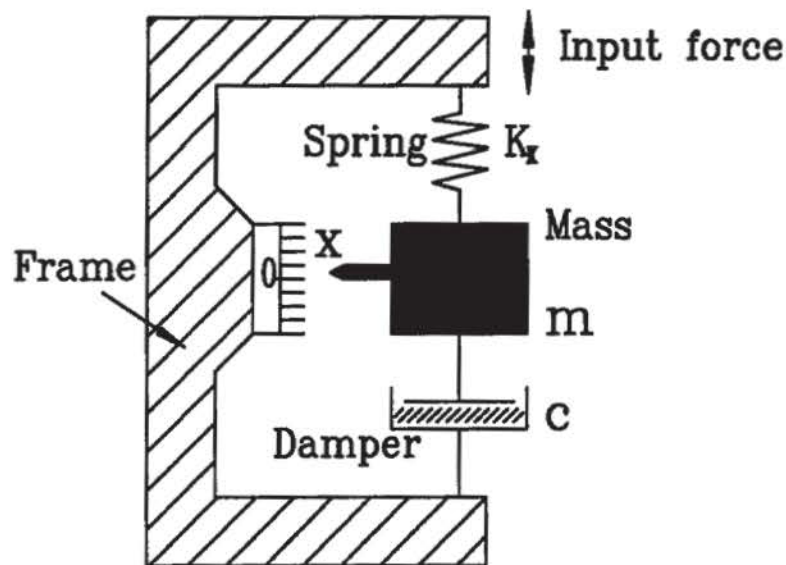


Figure 3.2. The spring-mass system.

damping, and the stiffness of the suspension determine its characteristics. We can consider the response of such a system to a force applied to the frame along the spring's axis by summing the forces from inertia, fluid damping, and spring displacement and equating them to the applied force

$$F = m(d^2x/dt^2) + c(dx/dt) + K_x x \quad (3.1)$$

where

x = displacement from the mass' rest position

c = damping coefficient

K_x = spring stiffness

In some instruments the damping is so high that the damping term outweighs the inertia term, so that the highest derivative in (3.1) is dx/dt and the equation is first order. It is more usual for the inertia term to dominate, and then the equation is second order.

If the acceleration is steady and the mass displacement is steady (any initial transient oscillation having died away),

$$m(d^2x/dt^2) = -K_x x \quad (3.2)$$

that is, the inertia force is balanced by the opposing spring force, and x is a measure of the acceleration. Such instruments can be purchased; the Setra

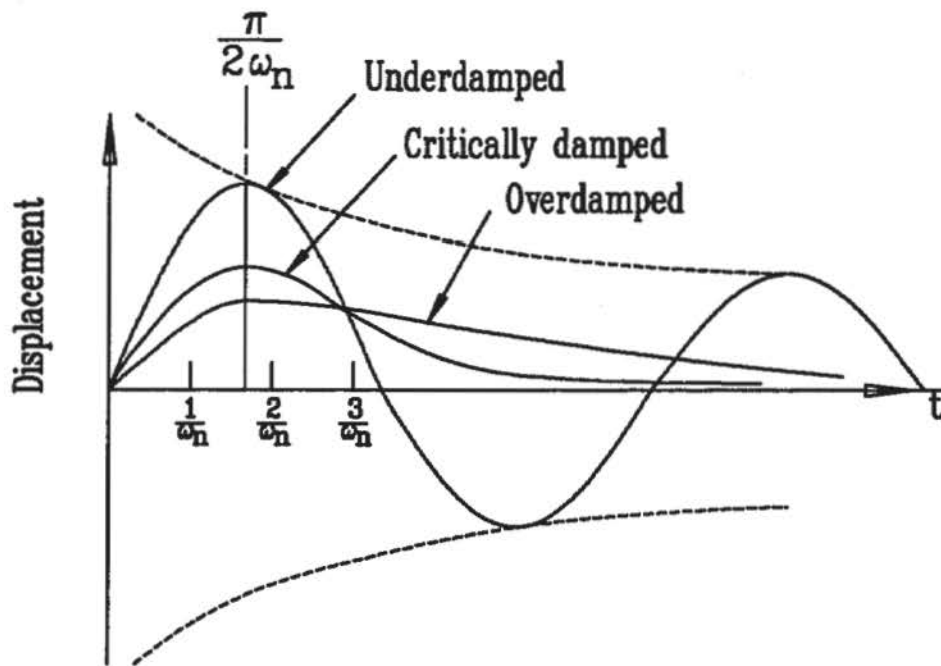


Figure 3.3. Second order system impulse response.

Company makes one with a metal diaphragm as the spring and the capacitance between the diaphragm and the case as a pickoff. Accelerometers where the displacement indicates the acceleration are *open-loop* instruments.

The solution of (3.1) is given [3] in terms of the undamped natural frequency $\omega_n (= 2\pi f_n)$ and the damping ratio ζ

$$\omega_n = (K_x/m)^{1/2} \quad (3.3)$$

$$\zeta = 1/2c/(K_x m)^{1/2}$$

The actual resonant frequency differs a little from f_n depending on the damping ratio

$$\omega = 2\pi f = \omega_n(1 - \zeta^2)^{1/2}$$

The impulse response of the spring-mass system, Figure 3.3, shows the mass displacement over time for different damping ratios. Three curves represent the conditions:

Condition	Damping	Response
Underdamped	$0 < \zeta < 1$	Overshoots $x = 0$ and oscillates
Overdamped	$\zeta > 1$	Sluggish, no overshoot
Critical damping	$\zeta = 1$	Shortest rise time without overshoot

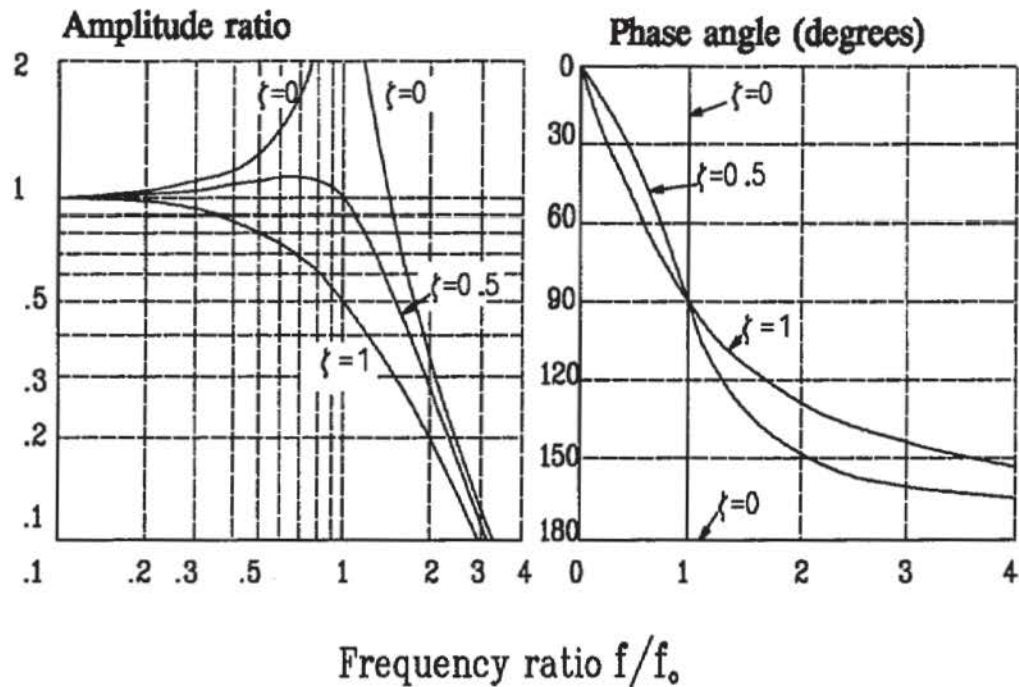


Figure 3.4. Mass-spring response.

Note that the first peak of the underdamped response occurs at time $t = \pi/2\omega_n$. If the damping ratio is set at 0.7, there is a small overshoot but a fast response, and mechanical sensors are often designed to work at "0.7 of critical damping."

When the acceleration in Equation (3.1) is time varying, the pickoff output (the measure of x) varies with frequency and damping. At the resonant frequency, f_n , the output will peak. Figure 3.4 shows how the output varies in amplitude and phase with frequency, for various values of damping. When there is no damping at all, the resonant response is infinite, but this does not happen in practice as there is always some damping in a system.

Q Factor

The Q factor describes the sharpness of the resonant response of a single degree of freedom, lightly damped system. Q is approximately related to ζ by

$$Q = 1/2\zeta \quad (3.4)$$

If the sharpness is represented by the linewidth Γ (the frequency difference between the two half-power points on the resonance curve), then

$$Q = f_n/\Gamma \quad (3.5)$$

Q also describes the amplitude of resonant response; it is roughly equal to the resonant amplitude divided by the driving amplitude (the transmissibility) for lightly damped systems. Finally, the Q factor relates the energy lost per vibration cycle (ΔW) to the amount stored (W), damping being the cause of energy loss

$$Q = 2\pi W / \Delta W$$

Bandwidth

The accelerometer output should be described by a simple scale factor, constant over a useful frequency range called the *bandwidth*, defined as the frequency band from zero to the frequency at which the phase of the response lags by 90° (second order system; 45° for a first order system).

Alternatively, the upper limit of the bandwidth can be defined by the frequency at which the amplitude ratio response is 0.5, the 3 dB point. In a true second order system both bandwidths should be the same. Commonly a system designer will specify the minimum instrument natural frequency and a range within which the damping constant must lie (e.g., 0.5 to 0.9).

Open-Loop Pendulous Sensors

The simple open-loop pendulous accelerometer will have a proof mass, a spring hinge, and a pickoff, and will use the angular deflection of the sensing element as its output. Because of this deflection, it can suffer from cross-coupling and can rectify vibrations, the so-called *vibropendulous error*, which we will now consider.

Cross-Coupling and Vibropendulous Errors

Figure 3.5 shows a schematic pendulous accelerometer. If the pendulum is displaced by an acceleration normal to it (along the *input axis*), or if the pendulum's center of mass is offset from the hinge axis, the instrument is sensitive to accelerations along the pendulous axis. Similarly if the hinge is twisted or misaligned in its plane, the instrument can be sensitive to acceleration along the hinge axis. Any of these sensitivities to steady-state acceleration along an axis other than the input axis is called a cross-coupling error.

In Figure 3.5, the pendulum is accelerated at an angle β to the input axis. Due to finite stiffness K_H the pendulum deflects a small angle α . Therefore the deflected pendulum senses a component of acceleration along the pendulous axis, giving rise to a steady-state cross-coupling error. Also, under vibration at frequencies well below its natural frequency, the pendulum reverses its position as the acceleration changes

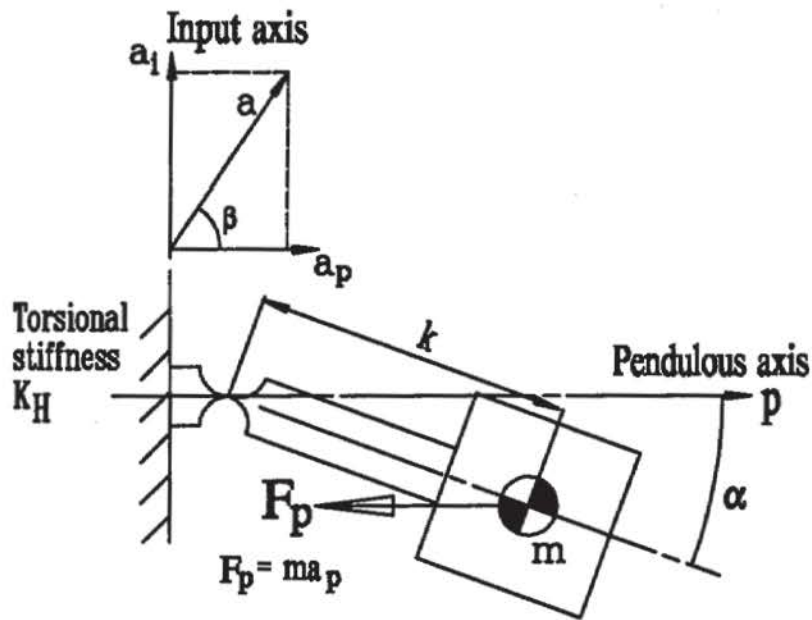


Figure 3.5. Pendulous accelerometer cross-coupling.

direction (although acceleration and displacement may not be exactly in phase), so that the net torque about the hinge has a finite value, the vibropendulous error. In Figure 3.5, the error torque about the hinge is

$$T = F_p \alpha k$$

where

F_p = inertia force along the pendulous axis

α = pendulum angular displacement = $(mak/K_H) \sin \beta$

k = length from hinge to proof mass CG

m = proof mass

a = acceleration

K_H = rotational stiffness of the hinge

β = angle between acceleration and pendulous axis

Thus the steady-state cross-coupling error torque is

$$T = \frac{1}{2}(m^2 a^2 k^2 / K_H) \sin 2\beta \quad (3.6)$$

which has a maximum value at $\beta = 45^\circ$. Under sine vibration at frequencies well below the natural frequency, $a = a_0 \sin \omega t$, the maximum value is

$$T(\omega) = (m^2 k^2 / 2K_H) a_0^2 \sin^2 \omega t$$

which has an average value

$$T(\omega)_{\text{ave}} = (m^2 k^2 / 4K_H) a_0^2$$

The larger K_H (a stiff hinge) and the smaller α and k , the lower these errors. However, with small α the pickoff must be extremely sensitive, and the resolution of the sensor will be set by the electrical noise in the pickoff circuit.

Pickoff Linearity

Open-loop mechanical sensors rely on the linearity of the pickoff and the spring which opposes the input force (or torque) for their scale factor linearity. Mechanical springs usually suffer from hysteresis, although springs of diamond and silicon are exceptions. Because the spring must oppose the full scale input, it must be stiff; therefore, its mounting must be extremely stable to avoid null shifts, as very small motions represent substantial inputs.

Closed-Loop Accelerometers

We can improve instrument performance by operating closed loop, for then K_H in (3.6) becomes the servo stiffness and can be many orders higher than the open-loop value. The linearity problem has been transferred from the spring and pickoff to the forcer, and forcers can be made with better linearity and low hysteresis. Now the hinge spring can be made very weak, just strong enough so that it does not bend or break under shock and vibration, so its hysteresis and temperature effects are negligible. Also, pickoff linearity does not matter much, as the instrument always operates near null, although a highly non-linear pickoff can upset the servo stability because the loop gain will change with angle. All that matters is that the mechanical and electrical null positions are stable over time and temperature.

The servo loop is intended to keep the pickoff at null all the time, regardless of acceleration, which implies a servo with "infinite" stiffness. But practical designs have infinite (very high) stiffness only at low frequencies. Practical servos have limited bandwidths by choice, for the higher the bandwidth, the higher the noise in the system; most strapdown sensors have bandwidths in the 50- to 100-Hz range. Aircraft and missile structures vibrate in flight from engine noise, air turbulence and so on, in a band of frequencies from 20 to 2000 Hz or so. Therefore there are accelerations in the system well beyond the instrument bandwidth which cause the pendulum to deflect from null. These can lead to vibropendulous errors.

Open-Loop Versus Closed-Loop Sensors

The closed-loop mechanical sensor relies on the feedback system to restrain the sensitive element under high accelerations and rotations; therefore when there is no power to the instrument the sensitive element is unconstrained. If there is not much damping in the sensor, as in a dry pendulous accelerometer or in a dynamically tuned gyro, the instrument can be damaged by non-operating environments. This problem occurs in airplane-carried missiles, which may be carried at the wing tip for hundreds of hours, during which the airplane may make supersonic dashes and vibrate the missile severely (so-called "under wing carriage" environments). Unpowered closed-loop sensors have been known to suffer metal fatigue and fail under this treatment. But the open loop instrument makes a virtue of necessity; it is designed so that its sensitive element reaches full scale deflection under the maximum input; therefore it is inherently more rugged and can withstand shock and vibration whether or not it is powered.

Generally speaking, open-loop mechanical instruments are satisfactory where dynamic range does not exceed 5000 to 1, and where scale factor error can be 0.1%. Where higher performance is needed, it is better to use feedback systems to null any sensing element motion, and the output is the feedback required to maintain the instrument at null. But note that the vibrating beam accelerometer and ring laser gyro are open-loop instruments with very good performance, as we shall see in later chapters.

Sensor Rebalance Servos

Accelerometer servos share some common characteristics with gyro servos, so we can take a general look at servos relevant to both sensors. Servo design is a science usually practiced by a specialist, so we will limit this look to an overview. Oppenheim [3] provides an introduction to linear feedback systems, and there are many other good books on servo design. We will consider gyro and accelerometer loops from the instrument designer's point of view, concerning ourselves with the demands that the rebalance loop makes of the sensor.

Servos may be analog or digital. Analog servos continuously vary the feedback signal to a forcer¹ so as to null the sensor, and the output is usually a voltage proportional to input. Digital servos pulse the signal (often a fixed current) to the forcer, and the output is the count of the number of pulses per unit time. Analog servos require the instrument scale factor to be well known over the entire operating range, whereas digital schemes operate at only one level, reducing demands on the instrument scale factor.

¹ Generally a forcer for an accelerometer, and a torquer for a gyro.

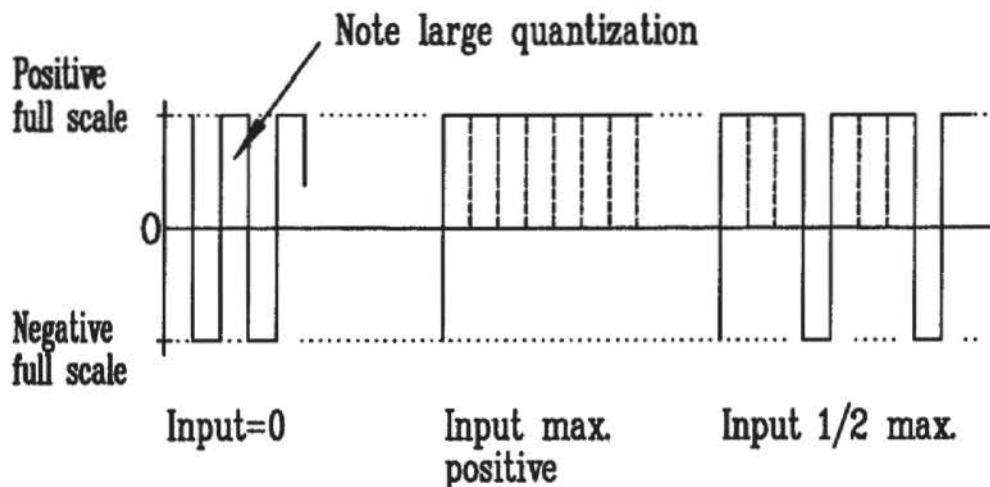


Figure 3.6. Binary feedback signals.

Binary Feedback

Binary servos provide pulses having only two states, one corresponding to full scale positive input and the other to full scale negative. With no sensor input, positive and negative pulses alternate and no net force is generated. In Figure 3.6 each pulse represents either a "quantum" of velocity (if an accelerometer) or of angle (if a gyro). The measure of this quantization, the *pulse weight*, may be in arc-sec/pulse for a gyro or meter/pulse for an accelerometer. The sensor output is the difference between the number of "up" pulses and "down" pulses in a sample period.

Assume, for example, that the servo delivers current pulses. Each pulse has finite rise and fall times due to the circuit time constants, and these time constants are likely to vary with temperature. Therefore the pulse weight can vary with temperature, which is minimized by making the duration of the peak (constant) current large compared with the rise and fall times. This tends to give the pure binary system coarse quantization (large pulse weight).

To avoid this problem one can use a width modulated binary loop. As in the simple binary loop the current has only two states (plus and minus full scale) but the switching point between the states can be varied in small steps, shown in Figure 3.7. Thus the quantization - these small steps - is fine and the pulse weight small.

Binary loops reverse the drive current with a bridge of four switches, Figure 3.8, an "H-switch." The switching speed of the output transistors sets

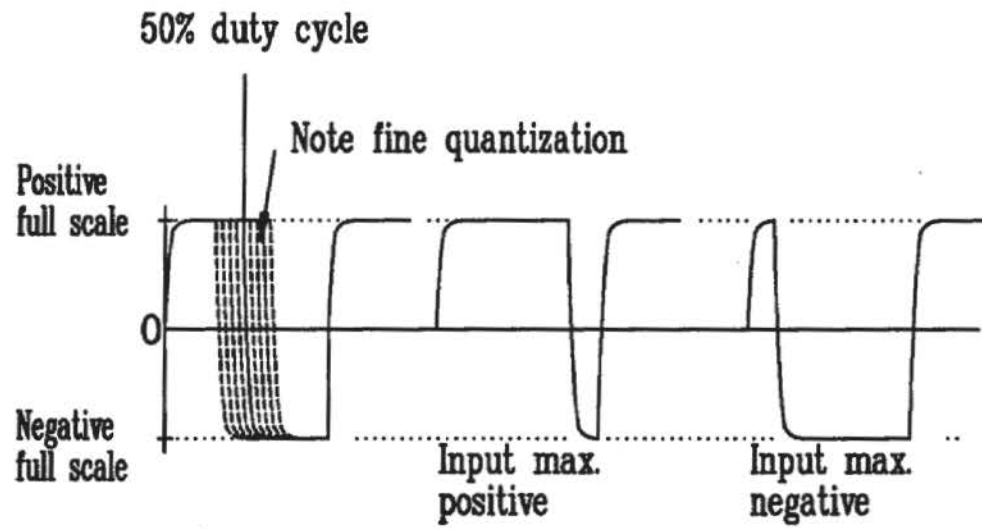


Figure 3.7. Width modulated binary feedback signals.

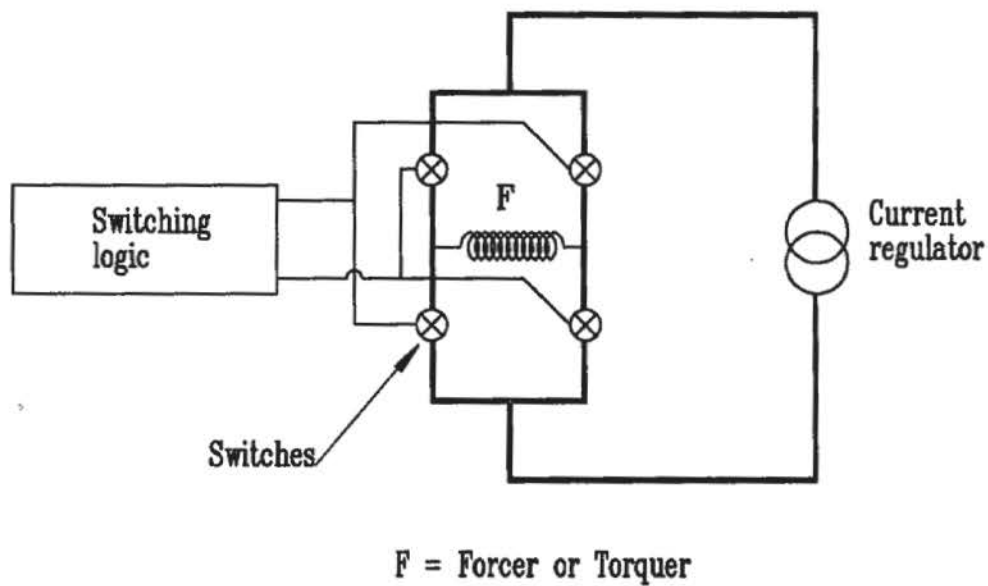


Figure 3.8. Binary switching (H-switch).

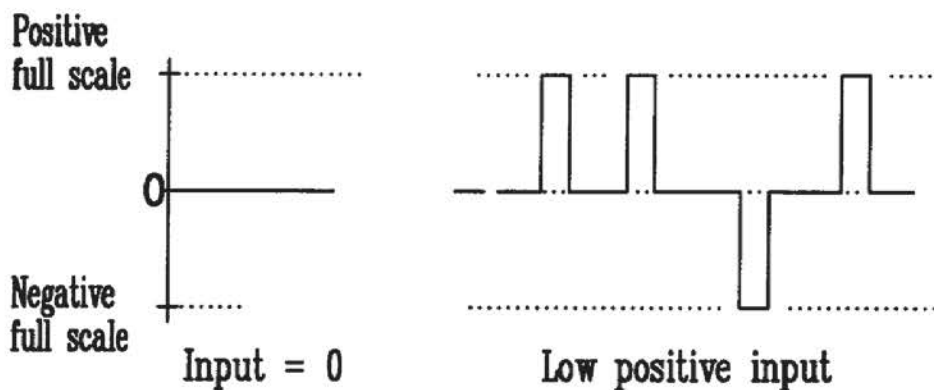


Figure 3.9. Ternary (pulse-on-demand) feedback signals.

the upper limit to the quantization in width modulated loops. If the switches have high leakage currents they will contribute bias errors to the sensor loop as current will flow in the forcer even when the switch is "off." Timing errors, which cause the switch openings and closings to be staggered, allow current to bypass the forcer and also give bias errors. If the current regulator is unstable, the scale factor will be unstable. However, with good engineering these error can be limited to 10 ppm.

Either binary system applies the full scale rebalance power to the sensor all the time, which has the advantage that the sensor heat dissipation is constant and therefore its temperature gradients are steady. Whereas this improves the performance of temperature controlled instruments, it is of little value for other (unheated) tactical sensors. It can be a disadvantage in the dynamically tuned gyro (DTG) where it is difficult to dissipate heat from the torquing coils. One way around this is to use multi-range loops where the peak current is switched up as the input rises. Multi-range loops are more expensive because they add the hardware cost of the level switching logic and the more complex current controller, and the cost of the system software needed to keep track of the changes in scale factor.

Ternary Feedback

The analog servo supplies only the current necessary to rebalance the input, so it dissipates much lower power than the binary schemes. However, as noted earlier, it demands better scale factor linearity in the sensor. The ternary loop offers a compromise; it supplies fixed-height pulses upon demand. When there is no input, there are no pulses, as shown in Figure 3.9, so the power is not

constant. Width modulation is possible to avoid large quantization, mechanized by growing the pulse in width as more torque or force is called for.

Pulse Feedback and Sensors

Pulse feedback hits the sensor with an impulse, so the pulse repetition rate must not match a sensor mechanical resonance or a nutation frequency, or the sensor will become unstable. Even non-resonant motion can become large if mechanical or electronic damping is not provided, so that designing servos for the DTG is difficult because the DTG must have as little mechanical damping as possible (see Chapter 9).

Permanent magnet forcers or torquers must have as low a resistance (R) as possible to reduce Joulean heating (i^2R), and as low an inductance (L) as possible to keep the time constant (L/R) low. This means that the number of turns on the forcer/torquer coil must be kept low. Tuning the circuit by adding a capacitor across the coil can reduce the effective time constant. However in width modulated systems the tuning capacitor can cause non-linearity errors because it acts as a bypass to the higher frequency Fourier harmonics which arise as the pulse width varies from 50% duty cycle.

The Voltage Reference Problem

The current which rebalances the mechanical sensor is measured by passing it through a high-quality resistor of precisely known value and comparing the voltage across the resistor to a standard. This standard reference voltage ultimately determines the scale factor, and therefore the navigation accuracy. Providing a very precise voltage reference, unaffected by nuclear radiation, has turned out to be a challenging part of accurate navigation system design and is a first order problem for force-feedback sensors.

Novel Accelerometer Principles

We have discussed the general principles of accelerometers above using the pendulous sensor as an example. Other phenomena are being applied to accelerometry, notably the vibrating beam (which we will cover in detail in Chapter 5), and two new technologies offer promise, the surface acoustic wave and optical fiber sensors, so their principles are worth noting.

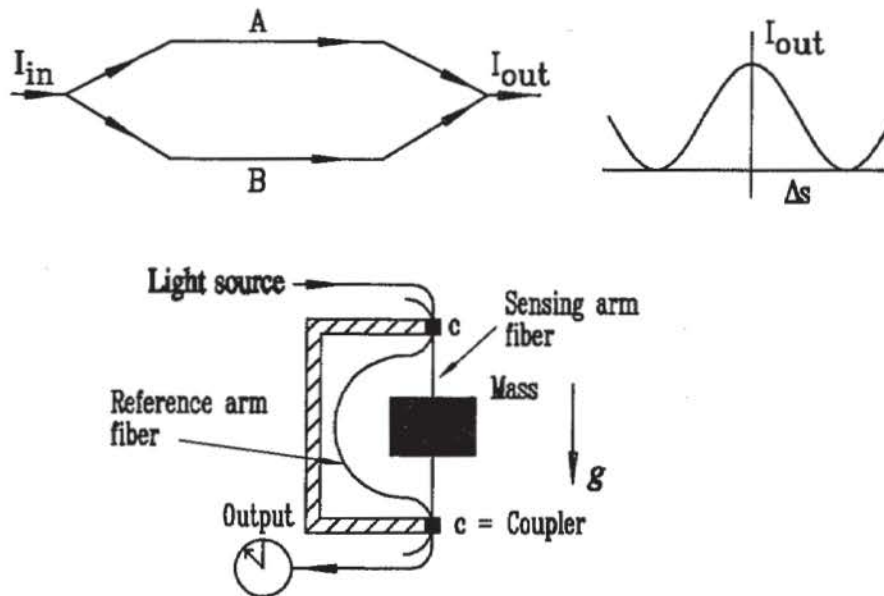


Figure 3.10. The Mach-Zehnder accelerometer principle.

Surface Acoustic Wave Accelerometer

The surface acoustic wave (SAW) accelerometer uses very high frequency (about 100 MHz) surface acoustic waves to measure the deflection of the hinge supporting a proof mass. There are different ways of using the SAW, described, for example, in [4,5]. One approach uses an acoustic interferometer. Surface waves travel on both sides of the hinge, axially along it, launched in a piezoelectric material on each side of the hinge. After a centimeter or so the waves are picked up by another transducer, and the phases compared on the two sides. When the hinge is not bent, the two paths are equal and there is no phase difference, but when the hinge bends, the path on the outside of the bend is longer than that on the inside and there is a phase difference between the two waves. By summing the waves, one gets an intensity signal depending on the bend curvature.

This sensor is analogous to the interferometric optical gyro, Chapter 12, except that the acoustic wave travels with the hinge medium, whereas optical waves travel in inertial space.

Fiber Optic Accelerometers

Fiber optic accelerometers are sensors which use guided wave optical effects to measure the displacement of a proof mass against an elastic support. They can be connected to the indicator by optical fibers, which might be useful in situations where it is dangerous to bring electricity to the point where the acceleration is to be sensed, in fuel tanks for instance.

The optical effects employed may be microbend losses, interference, or photoelasticity. Microbends are bends in the fiber whose radii approximate the fiber diameter; they cause the fiber loss to jump and therefore reduce the light transmitted. Interference between two optical beams gives a sensitive measure of their path difference, just as in the SAW accelerometer. In the Mach-Zehnder interferometer, Figure 3.10, a light beam is split into two paths (A and B) which travel separately for a short distance and then recombine. If the divided beams undergo a relative phase change (perhaps due to one path being longer, or passing through a medium of different refractive index), they will interfere when recombined and the intensity will be reduced. Should the phase difference on recombination be a multiple of 2π , the two waves will cancel one another and the intensity will be zero. The Mach-Zehnder is analogous to the electrical bridge circuit and can be made from two pieces of fiber separated by fused couplers. A mass attached to one fiber of such an interferometer will strain when accelerated and will give a measure of acceleration. By attaching one mass to both fibers so that it increases the tension in one at the same time as it reduces the tension in the other gives push-pull operation with better sensitivity. Resolution of $1\mu\text{g}$ has been reported [6]. The presence of the common mode path reduces thermal and acoustic sensitivity by many orders.

References

1. Okun, L.B., "The concept of mass", *Physics Today*, pp. 31-36, June 1989.
2. Bernard, A., B. Foulon, G.M. Le Clerc, "Three axis electrostatic accelerometer", DGON Symposium Gyro Technology, Stuttgart, 1985.
3. Oppenheim, A.V., A.S. Willsky, I.T. Young, *Signals and Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
4. Hartemann, P., P-L Meunier, A. Jacobelli, "Elastic surface wave accelerometers", U.S. Patent 4 515 016, 7 May 1985.
5. Bower, D., M. Cracknell, A. Harrison, "A high linearity SAW accelerometer", *Proc. IEEE 41st Annual Frequency Control Symposium*, pp. 544-547, 1987.
6. Tveten, A.B., A. Dandridge, C.M. Davis, T.G. Giallorenzi, "Fiber optic accelerometer", *Electronic Letters*, 16, 22, pp. 854-856, 23 Oct. 80.

6

The Principles of Mechanical Gyroscopes

In this chapter we present the base for describing, in subsequent chapters, different mechanizations of mechanical gyro. We will derive the Law of Gyroscopics and the expression for the Coriolis acceleration, the phenomena underlying spinning wheel and vibratory gyroscopes.

The French scientist Leon Foucault used a large pendulum (67 m long), which he built in the Panthéon in Paris, to study the earth's rotation. Its iron bob weighed 28 kg, and it swung with a period of about 15 s. As the earth rotated under the swinging pendulum, the plane of the swing appeared to rotate clockwise (because he was in the northern hemisphere) at vertical earth's rate.

The pendulum motion was perturbed by draughts, so it soon becomes inaccurate as an inertial reference. So, building on the experiences of Bohnenberger (1810, Germany) and Johnson (1832, United States), Foucault devised a more compact and accurate instrument in 1852, a gimbaled wheel which, because of its angular momentum, would stay fixed in space while his laboratory rotated around it [1-4]. (Onlookers see the *gyro* rotating, being oblivious to their motion with the earth.) He joined the Greek words which mean "to view" and "rotation," *gyros* and *skopein*, and coined the word "gyroscope" for his suspended wheel. While some writers reserve this word for rotation sensors with spinning wheels, in this book we use it for any instrument which signals or measures rotation. To begin with, let us consider the inertial property arising from the angular momentum of a spinning wheel.

Angular Momentum

Angular momentum stabilizes. The familiar rifle is a gun with a spiral groove down the inside of its barrel, which spins the bullet to stabilize its path.

From Newton's second law of motion we know that the angular momentum of a body will remain unchanged unless it is acted upon by a torque, and that the rate of change of angular momentum is equal to the magnitude of the torque, T :

$$T = dH/dt \quad (6.1)$$

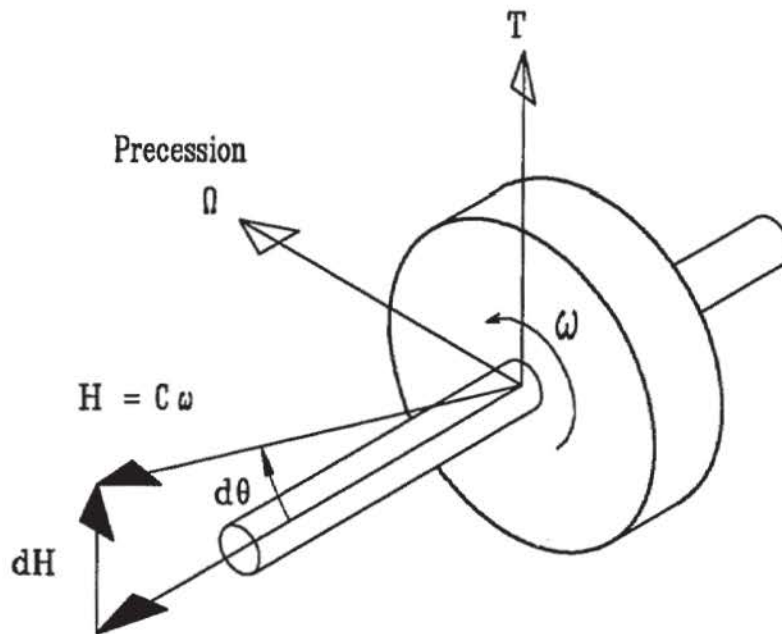


Figure 6.1. The law of gyroscopics.

where $H = \text{angular momentum} = I\omega$.

If the torque acts about the axis of rotation, its effect is to increase the angular velocity:

$$T = C \, d\omega/dt = C\alpha$$

where

$C = I$ about the spin axis (polar moment)

$\alpha = \text{angular acceleration}$

In passing, note that the moment of inertia (which describes the *distribution* of the mass in a body) need not be constant. For example, pirouetting ice skaters spin faster and faster as they move their arms from horizontal (a higher moment of inertia) to vertical (a lower moment of inertia). As there are practically no torques acting on the skaters, and Newton's law requires that angular momentum must be conserved, as the skaters' moments of inertia fall their spin speeds must rise.

The Law of Gyroscopics

If a torque is orthogonal to an object's spin axis it cannot change the angular velocity vector magnitude, but it can change its direction, shown in Figure 6.1.

The angular momentum H lies along the spin axis, and the torque T acts upwards, that is, it is a motion which is trying to turn the wheel in the direction that a right-handed screw would turn if it were moving along the same direction as T . The (small) angular momentum change dH is in the same direction as T , and its magnitude is

$$dH = H d\theta$$

where $d\theta$ = angle turned through. Combining this with (6.1) we get the Law of Gyroscopics:

$$T = dH/dt = H d\theta/dt = H\Omega \quad (6.2)$$

where Ω is the *Precession Rate*, the angular velocity of the wheel about the axis normal to the plane of the spin and input torque.

Conversely, if the wheel is turned about the precession axis, a torque appears about the orthogonal axis. For remembering which direction a gyro wheel moves, the rule is that when you try to turn a gyro wheel, it tries to turn so that its spin axis is aligned with your input torque.

Parasitic Torque Level

We are usually concerned with reducing gyroscope drift caused by parasitic torques, so let us visualize the size of a typical torque. A gyro for a 1-m/h navigator could have $H = 10^6$ dyn·cm·s (g·cm²/s), and need to have random drift below 0.01 deg/h. Then

$$\Omega = 0.01 \text{ }^\circ/\text{h} = 4.8 \times 10^{-8} \text{ rad/s}$$

$$T = 4.8 \times 10^{-2} \text{ dyn·cm}$$

A dyne is roughly 1 mg weight; 0.05 mg is the weight of a piece of aluminum oven foil 1 mm square. Such a weight at a radius of 1 cm would cause this level of drift.

The Advantage of Angular Momentum

One could use the inertia of a stationary wheel to measure the rotation of a body attached to its spindle by measuring the angle between the spindle and the wheel. One might try to economize in making a gyro by leaving out the drive motor and its power supply, merely supplying a pickoff or encoder to measure the angle between the shaft and rotor. So now let us emphasize the benefit we get from spinning a gyro wheel. Assume that the friction in the

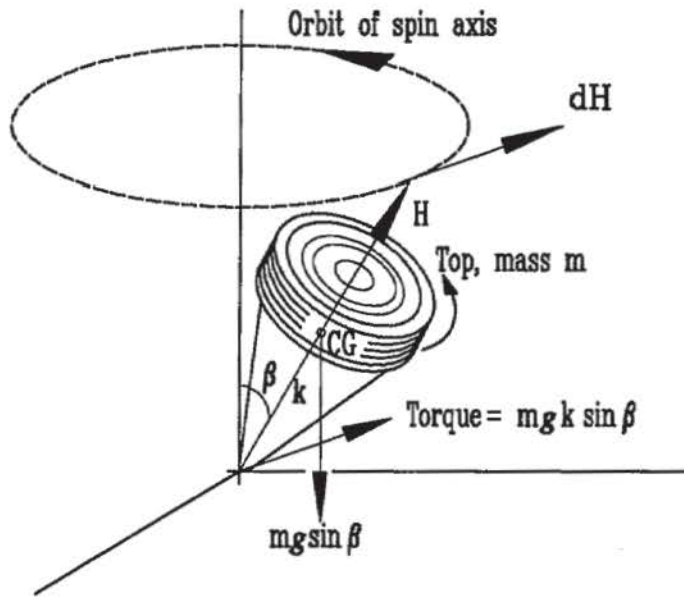


Figure 6.2. Precession of a spinning top.

bearings transmits a torque T to the wheel. In time t , the wheel's acceleration α leads to an angular velocity Ω_1 :

$$\alpha = T/C \Rightarrow \Omega_1 = (T/C)t$$

The same torque applied normal to a wheel spinning at angular velocity ω gives a precession angular velocity Ω_2 :

$$\Omega_2 = T/C\omega \text{ so that } \Omega_1/\Omega_2 = \omega t$$

If T represents an unwanted parasitic torque, its effect is ωt times greater on the unspun wheel, and ω can easily be 2500 rad/s; angular momentum stabilizes far more than stationary inertia.

The Spinning Top - Nutation

A spun top stands upright, its spin axis making a slow coning motion in space about the vertical (it is *precessing*). Figure 6.2 shows such a top, with gravity acting to create a horizontal torque to topple the top. As we saw earlier, the spin axis will move to try and align itself with the torque, that is, it will move in the direction dH (parallel to the torque) in Figure 6.2. But as soon as it moves, the

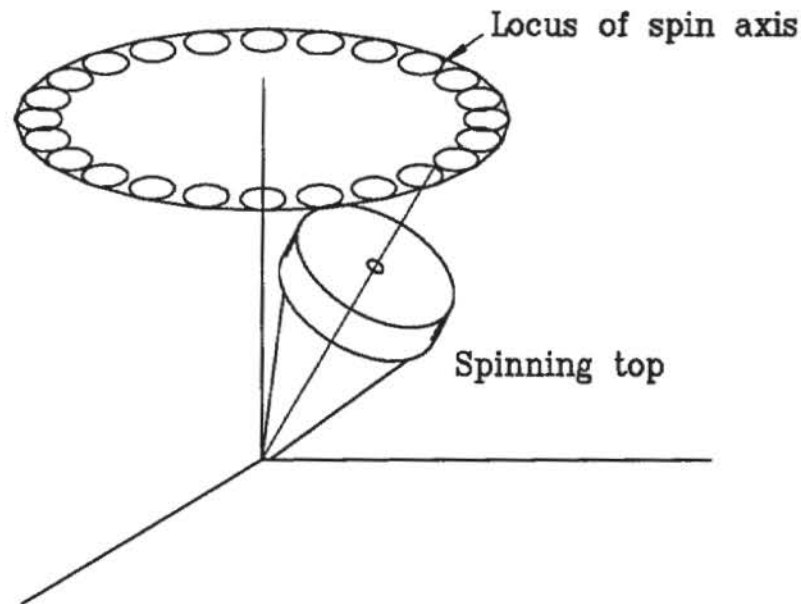


Figure 6.3. Precession and nutation.

direction of the torque moves too, for it is fixed in the top, so the precession direction changes, with the result that the top axis orbits in a circle. As the top slows, H decreases, and the top falls over when dH/dt can no longer balance the gravity torque.

For high spin rates, the precession rate is determined by the gravitational torque rather than by inertial torques. But if you were to tap a spinning top lightly with your finger, its spin axis would execute a rapid cycloidal motion called *nutation*, as sketched in Figure 6.3. Nutation is less influenced by the gravity torque and is determined by the inertia forces acting on the spinning body.

Equations of Spinning Body Motion

To justify these assertions, let us assume a spinning body symmetrical about its spin axis, so that the transverse inertias are equal. If the body spins about the z -axis in a case with a fixed axes set x , y , and z , it has angular momentum $H = C\Omega_z$. We can write the torques about the x - and y -axes as the algebraic sum of the inertia torque and the gyroscopic torque (6.2):

$$T_x = A(d\Omega_x/dt) + H\Omega_y$$

$$T_y = A(d\Omega_y/dt) - H\Omega_x$$

where A = transverse moment of inertia.

This pair of differential equations has solutions which are the steady precession and an oscillation

$$\Omega_x = -T_y/H \quad \text{Precession}$$

$$\Omega_y = T_x/H$$

$$\Omega_x = K_1 \sin \omega_n t + K_2 \cos \omega_n t \quad \text{Nutation}$$

$$\Omega_y = -K_1 \cos \omega_n t + K_2 \sin \omega_n t$$

where

$$\omega_n = H/A = (C/A)\omega_z = \text{nutation frequency} \quad (6.3)$$

K_1 and K_2 are constants found from the initial conditions. Note that for a sphere, $A = C$ and $\omega_n = \text{spin speed}$. For a thin disk, $C = 2A$ and nutation is twice spin speed. Generally, gyro rotors' inertia ratios C/A lie between the two extremes, and nutation occurs at about 1.5 times the spin speed.

The amplitude of nutation depends on the damping. In actual practice two-axis gyros (the only ones able to nutate) have small mechanical damping and need to be provided with nutation damping electronics.

Coriolis Acceleration

Gaspard de Coriolis, a French engineer and mathematician, postulated his acceleration in 1835, as a correction for the earth's rotation in ballistic trajectory calculations. The Coriolis acceleration acts on a body which is moving around a point with a fixed angular velocity, and moving radially as well. We will describe it here because it drives the tuning fork gyro (Chapter 10), but it must also be compensated in navigation systems where the vehicle is moving over the rotating earth.

Figure 6.4 illustrates a mass at point P moving in the yz -plane so that it has a rotation rate ω about the x -axis and its radius (r) is varying. We assume that the angular and linear velocities are constant. The mass velocity has components

$$dx/dt = -r \sin \theta (d\theta/dt) + \cos \theta (dr/dt)$$

$$dy/dt = r \cos \theta (d\theta/dt) + \sin \theta (dr/dt)$$

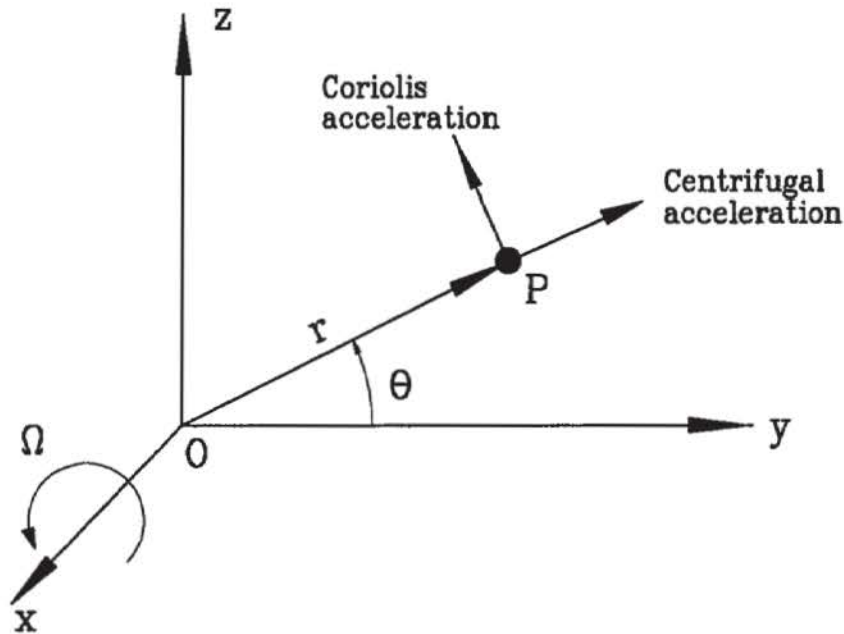


Figure 6.4. Coriolis acceleration.

Differentiating each equation with respect to time and rearranging, we get expressions for the accelerations of P along OP and normal to OP. Because we have constant velocities

$$dr/dt = v \quad \text{and} \quad d^2r/dt^2 = 0$$

$$d\theta/dt = w \quad \text{and} \quad d^2\theta/dt^2 = 0$$

and we have these accelerations:

Radial: $-r\omega^2$	Centrifugal acceleration	(6.4)
----------------------	--------------------------	-------

Normal: $2v\omega$	Coriolis acceleration	(6.5)
--------------------	-----------------------	-------

The centrifugal acceleration (6.4) causes a body to travel in a curved path because there is a force acting on it which is directed towards the path center of curvature. The sensation of centrifugal force which we experience in an automobile rounding a curve is the sense of our bodies attempting to obey Newton's first law to continue in a straight line. The centripetal force (providing the turning) comes from the friction of the tires on the road, transferred to the car seat, and thence to our bodies. When distinguishing between centrifugal and centripetal forces we must consider which reference frame carries the observer. In the car, we carry it with us; when whirling a stone on a string and feeling the outward tension in the string, we are sensing the equal and opposite reaction.

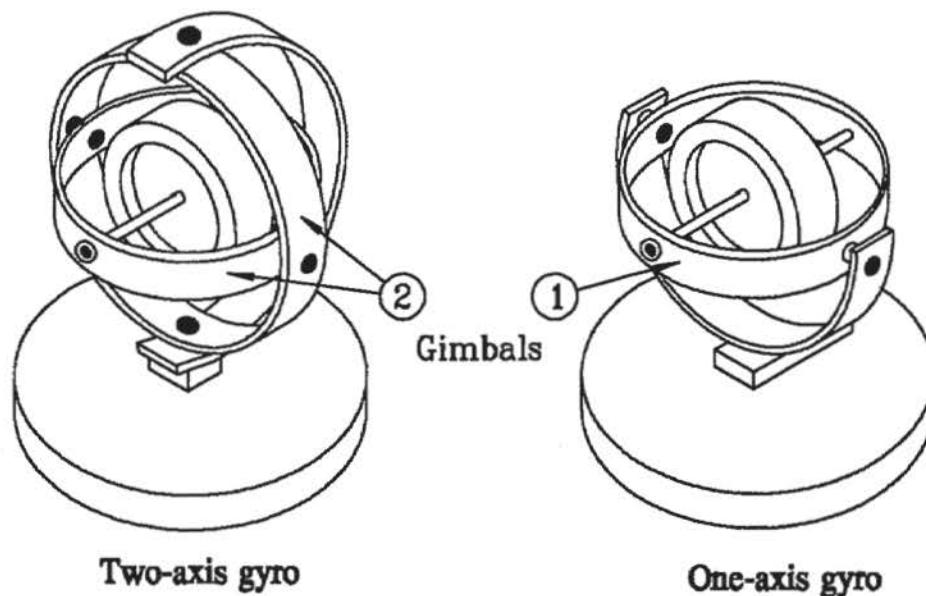


Figure 6.5. One- and two-axis gyros.

Gyroscopes with One and Two Degrees of Freedom

Mechanical gyros can be constructed so that the wheel is free to point to any direction in space by supporting the wheel in two gimbals, Figure 6.5. This is the type of construction that Foucault used in his gyro. As the wheel can rotate about two axes (in addition to its spin axis), we call this a *two degree of freedom* gyro. As we shall see in Chapter 8, this kind of gyro might not have physical gimbals; it may be made by electrostatically suspending a sphere, for example.

Around 1950 there was a school of gyro design at Massachusetts Institute of Technology, led by Draper, that believed that for the highest accuracy, gyros should be restricted to one free axis, what we call a *single degree of freedom* gyro, also shown in Figure 6.5. Nevertheless, engineers on the West Coast at Litton (and other firms) mass produced aircraft navigators with two degree of freedom gyros in the 1950s and 1960s. Later, the less expensive Dynamically Tuned Gyro (DTG), a two-axis gyro, became widely used.

The choice between one-axis and two-axis mechanical gyros is not based on very powerful cost or performance criteria, and in the next chapters we will describe the construction of both types of gyro, so that you can be better informed should you be faced with the task of deciding between them.

Conclusion

John Perry's book [5] on spinning tops is well worth reading.

The essence of spinning wheel gyroscopes is in the Law of Gyroscopics, Equation (6.2)

$$T = H\Omega$$

It tells us that a gyro will precess or drift at a rate Ω under the influence of a torque T . The angular momentum H is therefore a primary design parameter for such a gyro; it should be as large as possible to keep drifts small. Gyro size strongly influences H because the wheel speed is practically limited to the 12,000 to 60,000 rpm range, whereas the polar inertia of a gyro wheel goes roughly as the fourth power of its diameter, making diameter the dominant parameter.

Equation (6.3) states that the free (two-axis) gyro will nutate with frequency

$$\omega_n = H/A = (C/A)\omega_z$$

The Coriolis acceleration [Equation (6.5)]

$$a_c = 2v\omega$$

relates to bodies moving in an arc with varying radius.

References

1. Sorg, H.W., "From Serson to Draper - Two centuries of gyroscopic development", *Navigation, J. Inst. Nav.*, 23, 4, pp. 313-324, Winter 1976-77.
2. Scarborough, J.B., *The Gyroscope - Theory and Applications*, Interscience Inc., New York, 1958.
3. Foucault, L., "Sur une nouvelle demonstration expérimentale mouvement de la terre fondée sur la fixité du plan de rotation", *C. R. Acad. Sci.*, 35, p. 421, 1852.
4. Foucault, L., "Sur les phénomènes d'orientation des corps tournants entraînés par un axe fixe á la surface de la terre", *C. R. Acad. Sci.*, 35, p. 424, 1852.
5. Perry, J., *Spinning Tops and Gyroscopic Motion*, Dover, New York, 1957. Library of Congress Cat. 57-3845. Reprint of lectures first published about 1900.

11

The Principles of Optical Rotation Sensing

So far we have considered inertial navigation sensors which use the mechanics of matter described by Newton's laws of motion, basically concerned with the conservation of momentum in a frame of reference fixed in the stars. In the following chapters we will describe gyroscopes based on the inertial property of light. We will begin by defining that property, then we will review the theory of some optics used in gyros.

The Inertial Property of Light

Maxwell's wave theory and the quantum theory can explain between them all the optical phenomena that we observe. Newton had argued that light acted as if it were corpuscular, and Huygens had proposed wave explanations for refraction and reflection. Nowadays we see both aspects of light's interaction with the rest of the physical world operating in the optical gyro.

When Maxwell published his theory of electromagnetic waves in 1860, he showed that the waves would travel through space with a speed of 3×10^8 m/s. He noted that this was the value that people had measured for the speed of light and suggested that light might be an electromagnetic wave. The *ether* was postulated as the medium which carried light; it had to permeate all space, for we can see the distant stars because the light they emit reaches us, and it had to be very rigid to allow the propagation of such high speed waves. The ether must not have viscosity; there can be no mechanical drag associated with it, for Kepler's laws describe planetary motions without assuming any drag.

In any frame of reference (such as one fixed in the earth) moving with respect to the ether, the velocity of light would depend on the frame velocity relative to the ether, just as sound wave velocity depends on the source and observer's velocities relative to the air. So to test for an ether, Michelson and Morley set out to measure the difference in the speed of light on earth due to the earth's orbital velocity [1]. They concluded (1887) that the earth did not move relative to an ether.

In his special theory of relativity (1905), Einstein postulated that the speed of light is independent of the motion of the source. Accordingly there is no need of an ether frame of reference, and no surprise at the Michelson-Morley result.

Early in the 20th century, other scientists sought the ether by rotating a ring interferometer (one in which the light beams travel around a closed path, enclosing a finite area) in its plane; Haress (circa 1910) used a ring interferometer made of glass prisms. Not all these experimenters came up with the same answer, though; Georges Sagnac reported (1913-1914) that he saw fringe shifts which confirmed the ether's existence [2,3]. Nevertheless, the effect which relates interferometer fringe shift to rotation rate is called the Sagnac effect [4]; it is the basis of all the optical gyros which we will describe.

The Sagnac Effect

Sagnac's fringe shift can be explained without the need for an ether [5,6]. In the ring interferometer, Figure 11.1, two light waves circulate in opposite directions around a path of radius R , beginning at source S . Assume that the source is rotating with velocity Ω , so that the light traveling in the opposite direction to the rotation returns to S sooner than that which travels in the same direction. The wave traveling with the rotation covers a distance L^+ in transit time t^+ , and that against it covers L^- in time t^- . Then

$$L^- = 2\pi R - R\Omega t^-$$

$$L^+ = 2\pi R + R\Omega t^+$$

Assuming that the waves can circle the path N times, the difference in transit times is

$$\Delta t = N(t^+ - t^-) = 4\pi N R^2 \Omega / c^2 = (4NA/c^2)\Omega$$

and the distance represented by this transit time difference is

$$\Delta L = c\Delta t = 4AN\Omega/c = (Ld/c)\Omega \quad (11.1)$$

where

A = area enclosed by path

d = path diameter

L = total length traversed

For a wave of frequency f (wavelength $\lambda=c/f$), the period $T = 1/f$ is the time taken for a 2π phase change, so that the Sagnac phase change ϕ over time Δt is

$$\phi = 2\pi\Delta t/T = 2\pi f\Delta t$$

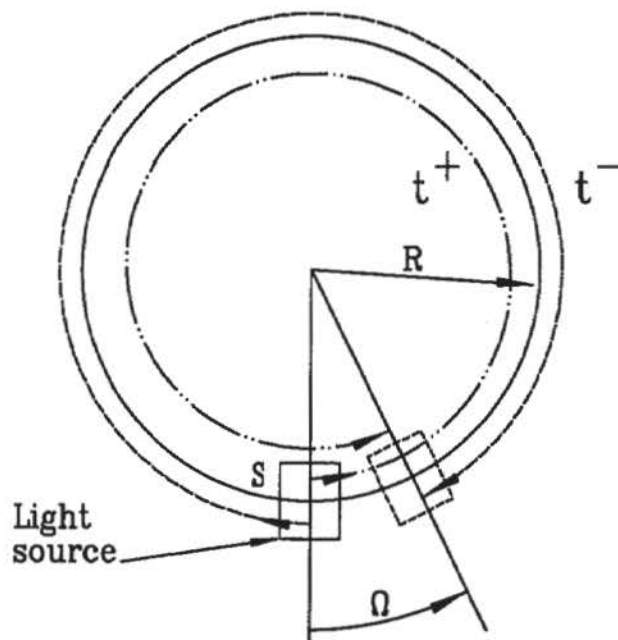


Figure 11.1. Transit time difference due to rotation.

so that, after substituting for Δt ,

$$\phi = (8\pi AN/\lambda c)\Omega \quad (11.2)$$

The fringe shift x is then $\phi/2\pi$, and

$$x = (4AN/\lambda c)\Omega = (4dL/\lambda c)\Omega = K_o\Omega \quad (11.3)$$

where K_o is the open-loop scale factor. The light path can be defined by prisms (Hares), mirrors (Sagnac), or by optical waveguides; it can have any shape, and A is the plan area of the shape normal to the rotation vector.

Figure 11.2 shows Sagnac's interferometer. Light from a source meets a beam splitter where it is divided into beams traveling clockwise (marked with "+" signs in the figure) and counterclockwise (marked "-") around a path defined by four mirrors. On completing the circuit the beams again meet the beam splitter and combine to pass to a camera. The two beams interfere to form a fringe pattern at the camera, which records the fringe position. The interferometer is mounted on a table which rotates it in the plane of the mirrors, and the fringe displacement is measured as a function of rotation rate. If it were to be used as a gyro, the rotation rate would be calculated from (11.2). Nowadays, the beams combining on the camera would meet on a photodetector, in which case the output current is a measure of phase shift.

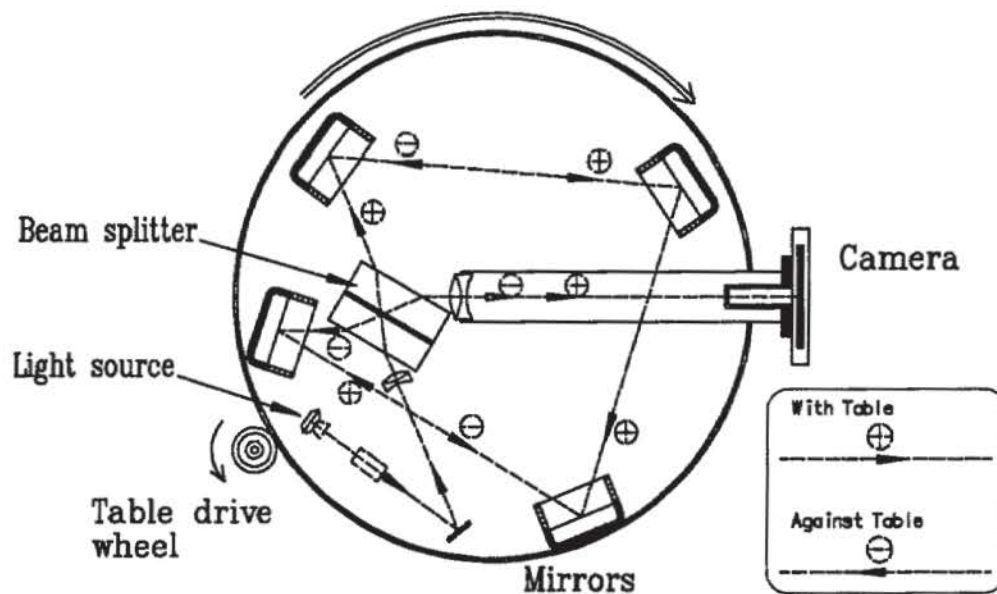


Figure 11.2. Sagnac's interferometer.

Sagnac Sensitivity - The Need for Bias

Assume a lossless Sagnac interferometer with an ideal 3 dB beam splitter. The light amplitude at the detector is the sum of the clockwise and counterclockwise amplitudes, shifted by the Sagnac phase ϕ :

$$E_D = E^+ + E^- e^{i\phi}$$

Assuming $E^+ = E^- = E$,

$$E_D = E(1 + \cos \phi - i \sin \phi)$$

Then the intensity on the detector is

$$I_D = |E_D|^2 = 2I(1 + \cos \phi)$$

where $I = |E|^2$. The sensitivity is

$$|dI/d\phi| = 2I \sin \phi$$

If the interferometer is stationary in inertial space, the intensity at the detector is maximum. As the interferometer rotates faster the intensity drops, reaching zero when the Sagnac phase difference is π (Figure 11.3). Also, because the

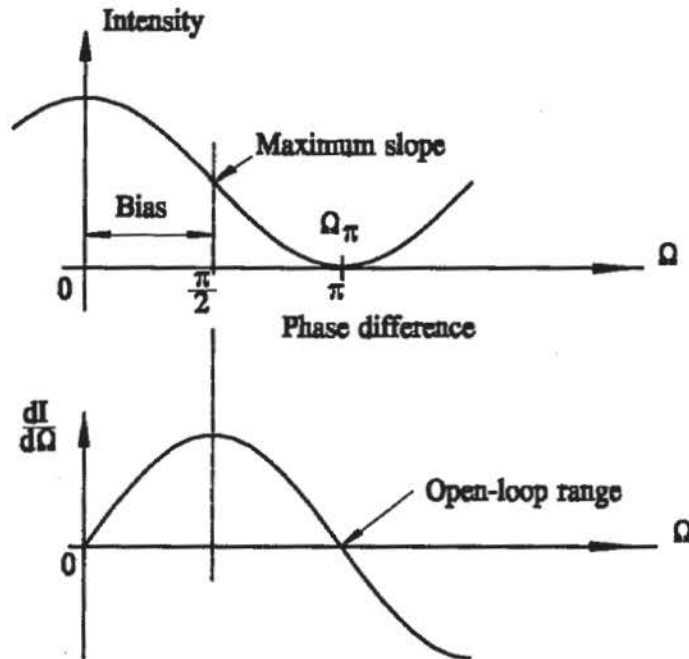


Figure 11.3. Sagnac interferometer sensitivity.

slope of the detector response curve is zero at zero rate, the basic Sagnac interferometer has a high threshold; large changes from zero rate are needed to give a detectable output.

At the rotation rate where the Sagnac phase $\phi = \pi$, $\Omega_\pi = \lambda c / 8AN$, [Equation (11.2)], and the intensity will again fall to zero. Once at this point, any increase in intensity can be due to either an increase or a decrease in rate; the output is ambiguous.

To overcome the first problem, the interferometer can be biased to the point of maximum slope (the maximum sensitivity) at $\phi = \pi/2$. This would give optimum threshold and reasonable linearity; the scale factor would at least be predictable if not constant. The rate range is still limited to that which represents a $\pm\pi/2$ phase shift, where ambiguity sets in. To overcome that problem we need to close a servo on the Sagnac phase, inserting a non-reciprocal phase modulator to oppose it so that the output intensity stays at the maximum-slope bias point. The gyro rate output will then be whatever measure we have of the feedback signal to this phase modulator, usually a voltage. We will describe specific implementations in the next chapter.

The Shot Noise Fundamental Limit

When we studied mechanical gyros we did not consider them as having some theoretical limit to their performance. There is a mechanical noise threshold,

Boltzmann noise, from the random thermal motion of the gyro structural material, which limits mechanical gyro performance, but this theoretical threshold is far below that achieved. In practice, material properties such as microcreep cause mass shifts, and electromagnetic components cause reaction torques, which overwhelm the fundamental limit.

But that is not so with optical gyros. While they have excellent stability over periods of hours, they are quite noisy in the short term (over seconds); indeed, they act as white noise generators. Their theoretical fundamental threshold noise really does set a lower limit to their performance. Mechanical gyros, especially gas bearing single degree of freedom floated gyros, are quieter and preferable if ultra-quiet short term pointing accuracy is needed, in the nanoradian regime.

Light energy is quantized in photons, and when light falls on a detector, it creates electrical noise (*photon shot noise*). Assuming that the detector amplifier electronic noise is small, the shot noise depends on the detector current and the measuring bandwidth [7], its r.m.s. value being

$$i_s = (2ei_d B)^{1/2} = (ei_d/t_s)^{1/2} \quad (11.4)$$

where

e = electron charge

i_d = average detector current

B = measurement bandwidth = $1/2t_s$

t_s = sample time

The gyro scale factor, $K = i_d/\Omega$, is the ratio of the output signal to the input rate, so that the minimum detectable rate change, $d\Omega$, caused by the uncertainty in output current i_s , defines the resolution

$$\frac{d\Omega}{\Omega} = \frac{1}{K} \frac{di_d}{i_d} = \frac{1}{K} \frac{i_s}{i_d}$$

so that, using (11.4),

$$\frac{d\Omega}{\Omega} = \frac{1}{K} \left(\frac{e}{i_d t_s} \right)^{1/2}$$

If the rate noise has a white spectrum with an r.m.s. value σ , it gives rise to a random walk in angle, R_θ , which propagates as the square root of time

$$R_\theta = \sigma \sqrt{t} \quad (11.5)$$

Angle random walk is a commonly specified parameter for optical gyros.

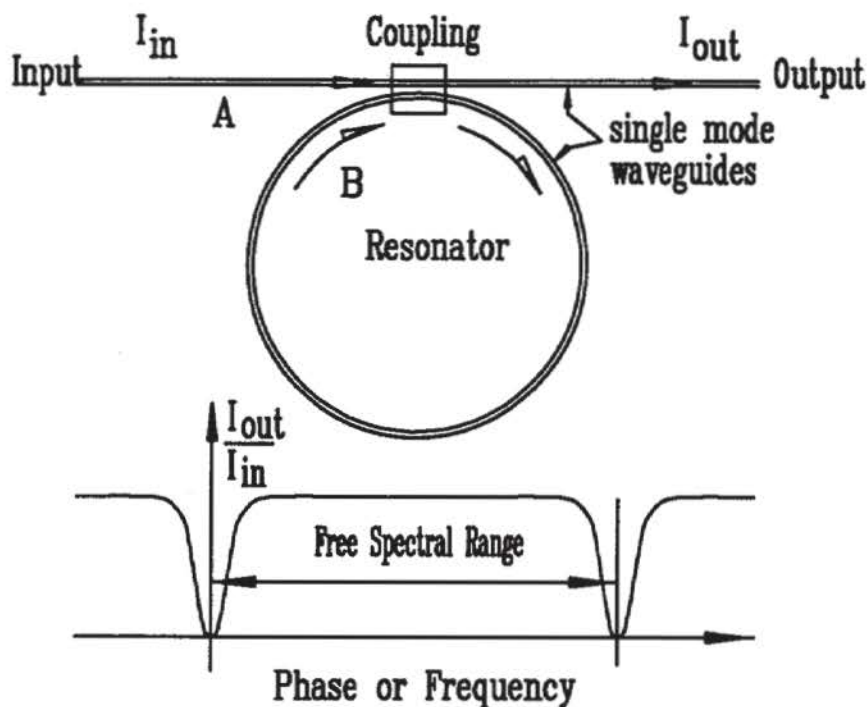


Figure 11.4. The optical resonator.

The Optical Resonator

In the Sagnac interferometer, a light wave splits and passes both ways round a path before being recombined and interfering with itself. The beam travels once over the path, making this a *single beam* interferometer. The Mach-Zehnder (Figure 3.10) is another kind of single beam interferometer; as with the Sagnac, its output varies sinusoidally with phase. The light frequency is constant in a single beam interferometer.

Resonators can operate with different clockwise (CW) and counter-clockwise (CCW) frequencies. If we have a path in which light can travel (it could be a ring of single-mode optical fiber) with some means of coupling light in and out, Figure 11.4, and if the perimeter of the ring is an integral number of wavelengths of the light in the medium, the ring will resonate. If this resonator rotates, the Sagnac phase causes the CW and CCW resonances to change. (Compare this with the serrodyne frequency shifter, Chapter 12.) For resonance the phase condition at a point must be the same for the wave leaving the point as it is for the wave arriving at the point, so that any Sagnac phase shift must be compensated by a shift in the wavelength to preserve phase matching.

Light enters the resonator of Figure 11.4 through the coupler and circulates. When it gets to the coupler again, its field is added to that of the wave entering through the coupler, the resulting amplitude depending on the phase between them. The wave circulates around the same path many times, so this kind of

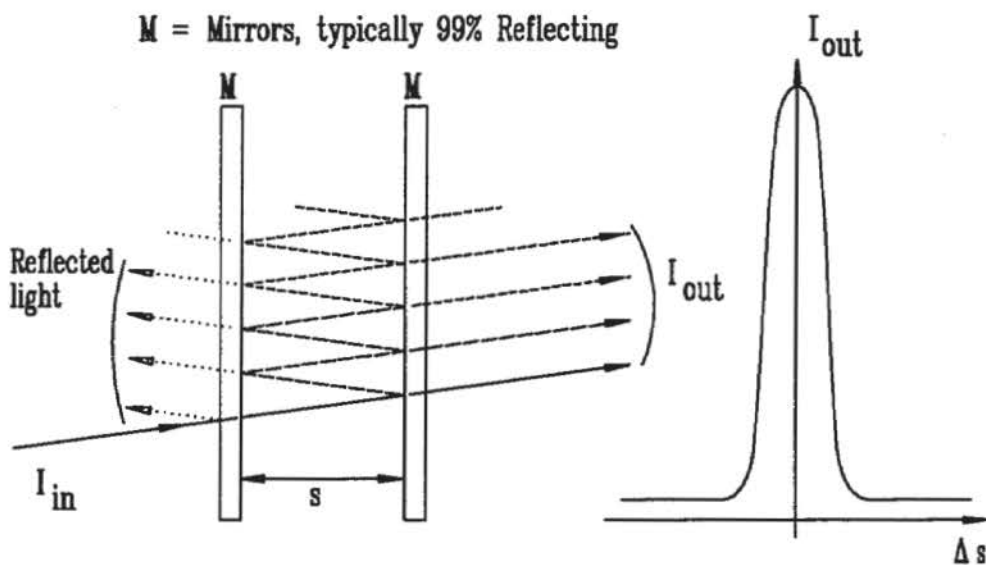


Figure 11.5. The Fabry-Perot interferometer.

interferometer is called a *multiple beam* interferometer. The continuing interference in the resonator sharpens its frequency response, so that it is narrower than the sinusoidal response of the single beam interferometer. Optical losses like absorption and scattering broaden the linewidth, for the waves can only circulate a few times before their energy is dissipated to the level of the background. If the loss is so high that the wave only makes it round the ring once, the frequency response is that of the single beam interferometer - sinusoidal.

The Fabry-Perot Resonator

The classical Fabry-Perot interferometer is a free-space optics device made from a pair of parallel, high reflectance, mirrors, Figure 11.5. It is sometimes called an *etalon* [8]; for the sake of example, assume that the mirrors have 99% reflectivity and are lossless. Light enters through the left mirror (1% transmission), passes across space s to the right mirror where 99% is reflected and 1% transmitted. After reflection at the left mirror, another 1% passes out of the right mirror, interfering with the first. This continues with subsequent reflected and transmitted waves. The phases of the transmitted waves, and thus the net intensity of the output, depend on spacing s .

The same process operates in the waveguide resonator, shown in Figure 11.6 with two couplers analogous to the two mirrors of the etalon. It has diameter d ; perimeter $p = \pi d$. In order to derive the expression for the power transmission

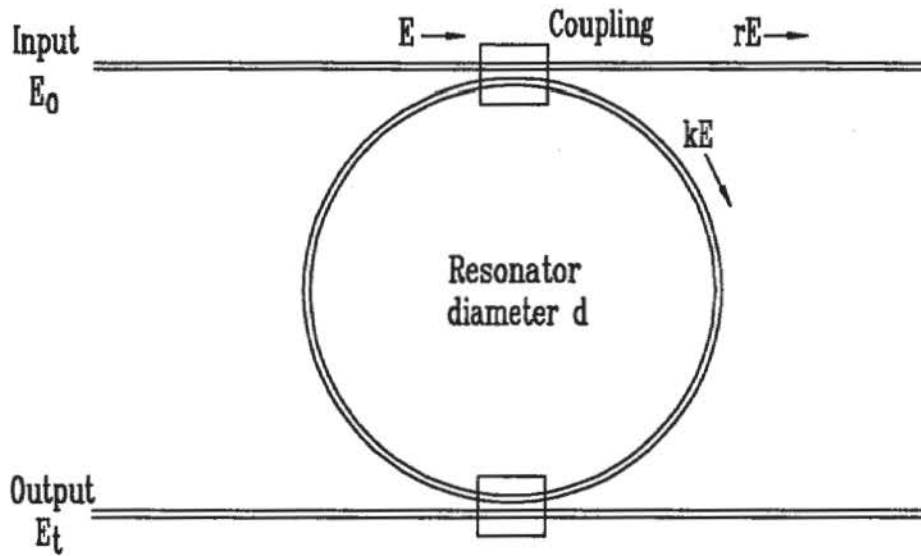


Figure 11.6. A two-port waveguide resonator.

of such a resonator, assume that the waveguide has a field attenuation of α per unit length, so that the field along the guide is

$$E(z) = Ee^{-\alpha z}$$

Assume also that the couplers are lossless and have a field transmission coefficient k . Then if a wave of amplitude E_0 enters the input coupler, the amplitude just inside the resonator is kE_0 . If the output coupler is opposite the input coupler, the wave amplitude at the output coupler is $kE_0e^{-\alpha p/2}$.

The amplitude of the wave transmitted through the output coupler is

$$\Delta E_{t1} = k^2 E_0 e^{-\alpha p/2} \quad (11.6)$$

The amplitude in the resonator just past the output coupler is $rkE_0e^{-\alpha p/2}$, where r is the coupler reflection coefficient. Going around the ring to the input coupler, the amplitudes just before it and just after it are $rkE_0e^{-\alpha p}$ and $r^2kE_0e^{-\alpha p}$. Back at the output coupler, the amplitude to be added to (11.6) is

$$\Delta E_{t2} = r^2 k^2 E_0 e^{-\alpha p} \quad (11.7)$$

and so on for an infinite number of trips around the resonator. Guided waves are described by the *propagation constant* $\beta = 2\pi n/\lambda_0$, where n is the effective index and λ_0 is the free space wavelength; waves having different propagation con-

starts travel with different speeds. If we assume that there is no phase change in the couplers, the phase difference δ between the waves at the output coupler is $\delta = \beta p$.

Summing the transmitted waves whose first terms are (11.6) and (11.7), allowing for the phase $e^{i\delta}$,

$$E_t = k^2 E_o e^{(i\delta - \alpha p)/2} [1 + r^2 e^{(i\delta - \alpha p)} + r^4 e^{2(i\delta - \alpha p)} + \dots]$$

Summing to infinity

$$E_t = \frac{k^2 E_o e^{(i\delta - \alpha p)/2}}{1 - r^2 e^{(i\delta - \alpha p)}}$$

The power transmitted is the square of this complex number:

$$\frac{P_t}{P_o} = \frac{k^4 e^{-\alpha p}}{[1 - r^2 e^{(i\delta - \alpha p)}]^2}$$

If we let $a = r^2 e^{-\alpha p}$, the denominator simplifies to

$$(1-a)^2 + 4a \sin^2 \delta/2$$

The power transmission and reflection coefficients are $R = r^2$ and $K = k^2$, and for lossless couplers $R = 1-K$, so that

$$a = (1-K)e^{-\alpha p}$$

$$\frac{P_t}{P_o} = \frac{K^2 e^{-\alpha p}}{[1 - (1-K)e^{-\alpha p}]^2} \cdot \frac{1}{1 + A^2 \sin^2 \delta/2} \quad (11.8)$$

$$A^2 = \frac{4(1-K)e^{-\alpha p}}{[1 - (1-K)e^{-\alpha p}]^2}$$

So Equation (11.8) gives the phase (or frequency) response of a Fabry-Perot resonator in terms of its waveguide loss and coupling constant.

The peak transmitted power occurs at $\delta = 0, 2\pi, \dots, 2m\pi$, where m is an integer, the *Longitudinal Mode Number*. The peak power ratio (P_t/P_o) is given by

$$P = \frac{K^2 e^{-\alpha p}}{[1 - (1-K)e^{-\alpha p}]^2} \quad (11.9)$$

For a lossless guide ($\alpha = 0$), $P = 1$. The Fabry-Perot has a series of resonances, for each integer m , whose frequency spacing is the *Free Spectral Range* (FSR). For a ring of perimeter p , the FSR is

$$f_{sr} = c/np$$

Resonator Finesse

The quality factor, Q , was defined in Chapter 3 for the frequency response of a mechanical resonator. The same definition holds for an optical resonator

$$Q = f_o/\Gamma$$

where

f_o = resonant frequency

Γ = resonance linewidth (full width at half maximum)

But whereas the Q -factor is a useful parameter when the resonant frequency is in the kHz or MHz range (10^2 to 10^8 Hz), optical frequencies are so high (10^{14} Hz) that optical Q 's would be so large as to be meaningless. In interferometry, instead of using the absolute resonant frequency as a reference, we use the free spectral range, defining a parameter analogous to the Q -factor called the *finesse*, F :

$$F = f_{sr}/\Gamma = c/np\Gamma \quad (11.10)$$

From Equation (11.8),

$$F = \frac{\pi A}{2} = \frac{\pi(1-K)^{1/2}e^{-\alpha p/2}}{1 - (1-K)e^{-\alpha p}} \quad (11.11)$$

For a lossless guide, $F = \pi(1-K)^{1/2}/K$. The finesse represents the number of times a wave circulates around the resonator until its intensity has dropped to the background light level; it is analogous to the number of turns in a fiber Sagnac interferometer of the same dimensions [9].

The Sagnac Effect in a Resonator

If we assume a resonator operating with a longitudinal mode number m so that $L = m\lambda = mc/f$, then $dL/df = -L/f$. The effect of the Sagnac length change can be deduced from Equation (11.1):

$$\Delta L = 4AL\Omega/cp; \quad |\Delta L/L| = \Delta f/f = 4A\Omega/cp$$

In a medium of index n , $c/n = f\lambda$, so

$$\Delta f = \frac{4A}{n\lambda p} \Omega \quad (11.12)$$

If we are using this frequency change upon rotation to make a gyro, Equation (11.12) defines the scale factor $K = \Delta f/\Omega$

$$K = 4A/n\lambda p \quad (11.13)$$

For a ring resonator, $K = d/n\lambda$; compare this with the identical scale factor for the phase-nulled interferometric fiber-optic gyro, Equation (12.1). If the frequency is thought of in pulses/s, then the reciprocal of K represents a pulse weight in arc-sec/pulse, or if each pulse is a digital bit, in arc-sec/bit.

The number of circuits of the path, N , does not affect the scale factor, but does affect the theoretical gyro threshold, or minimum detectable rate, by way of the resonator finesse.

Active and Passive Resonators

The resonators shown in Figures 11.4-11.6 are optical circuits with light entering either through a coupler or through a mirror. It is possible to place a lasing medium such as an excited helium-neon gas mixture inside a free-space resonator, so that light is generated inside the resonator and automatically finds its CW and CCW resonant frequencies. We call this an *active* resonator, Figure 11.7a.

Alternatively we can probe the resonator with laser beams from outside, by placing photodetectors so that they measure the intensities of the CW and CCW beams, Figure 11.7b. We can servo the frequencies of the light sources so that the resonator remains on resonance in both directions; this configuration is called a *passive* resonator. The active resonator emits the CW and CCW resonant frequencies; in the passive resonator we must find them. Active and passive resonator gyros have identical scale factor and fundamental limits for identical parameters, but the achieved resolution of these gyros depends on necessary engineering choices which prevent the ideal from being achieved.

Resonator Figure of Merit

Equation (11.11) expresses the finesse of the resonator of Figure 11.6, with coupling K and distributed waveguide loss of $\alpha \text{ cm}^{-1}$.

$$F = \frac{\pi(1-K)^{1/2}e^{-\alpha p/2}}{1 - (1-K)e^{-\alpha p}}$$

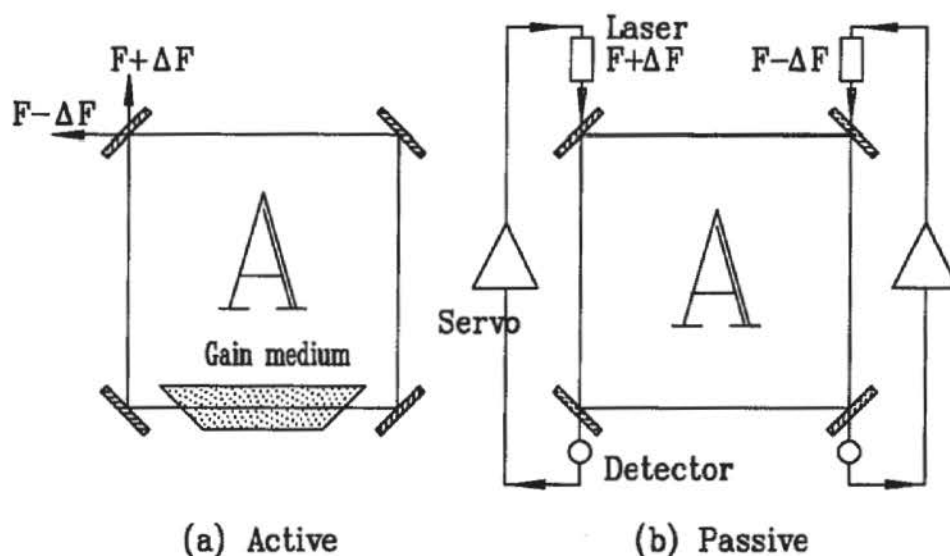


Figure 11.7. Active and passive resonators.

Note that finesse falls monotonically with increasing coupling, for the power coupled out looks like loss to the resonator. The coupling can be reduced to get high finesse, but only at the expense of the power transmitted to the detector. The resonance may then be very sharp, but it cannot be seen (it becomes unobservable). We can define a parameter which combines finesse and observability into a *Figure of Merit*:

$$F_m = F(P/P_o)^{1/2} \quad (11.14)$$

P/P_o is the transmitted power [Equation (11.8)].

If the resonator is over-coupled, the transmitted power is high and the finesse will be reduced. There is an optimum coupling corresponding to the peak F_m ; it occurs when the coupling K is equal to the round trip loss, αp , which we will use when we consider the fundamental noise performance of the passive resonator.

Optical Fibers

The Sagnac interferometer of Figure 11.2 has low sensitivity because the light passes only once round the path [$N=1$ in Equation (11.2)]. One way to make a Sagnac interferometer more sensitive is to send the light around the path many times ($N>1$), by replacing the mirrors with an optical fiber coil. Although mirrors

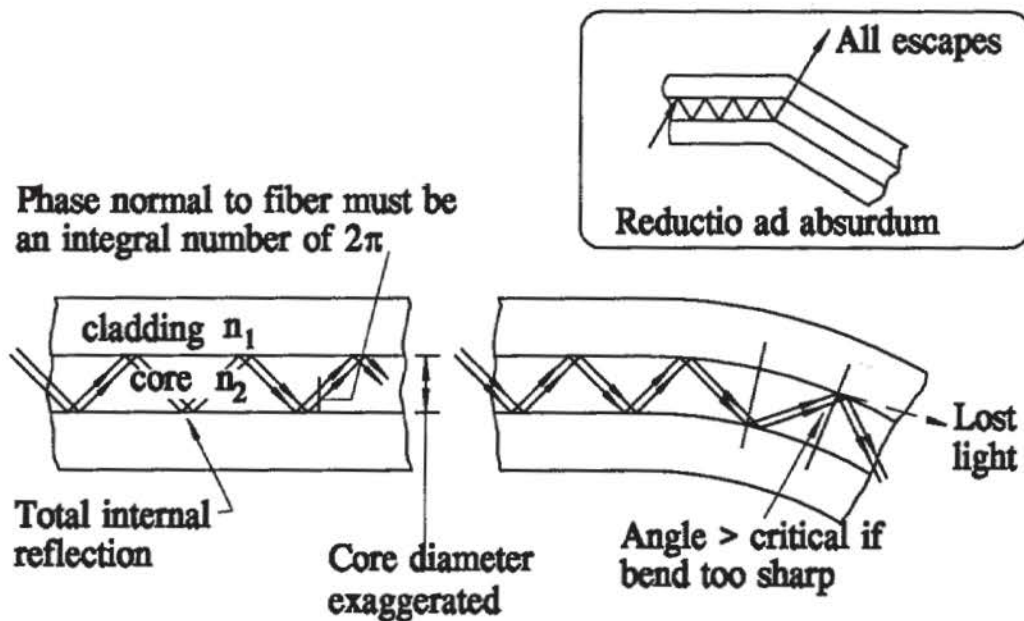


Figure 11.8. Transmission in a fiber.

and free-space optics preceded fiber guided-wave optics by centuries, we will leave free-space optics gyros until later and first describe some of the fundamentals of the optical fiber.

The optical fiber guides light by *total internal reflection*, a consequence of refraction and the *critical angle*. The next section reviews these characteristics of light propagation in a medium.

Refraction and Critical Angle

The refractive index of a transparent medium, n , relates the speed of light in the medium (v) to that in free space (c) by $v = c/n$; the index is usually temperature dependent. The differing speed in different media causes refraction; Snell's law relates the refraction of a wave to the refractive index of the medium

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

where

n_1, n_2 = media refractive indices

θ_1, θ_2 = angles of incidence (to the normal)

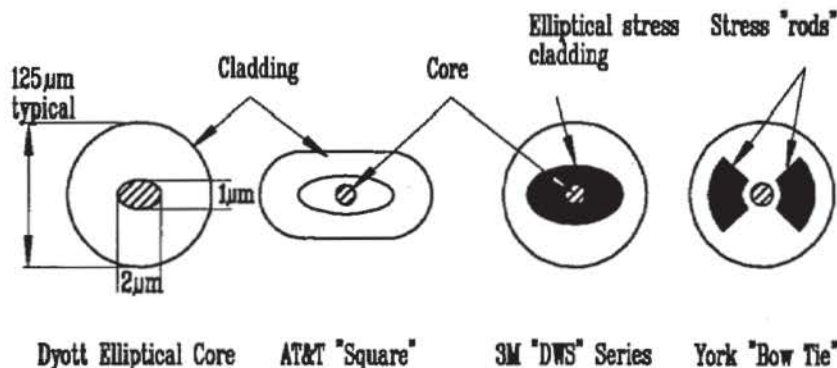


Figure 11.9. Some polarization maintaining fibers.

If $n_1 > n_2$, there will be a value of θ_1 for $\theta_2 = 90^\circ$, called the *critical angle*. If θ_1 is increased further no light passes into medium 2, for it is *totally internally reflected* within medium 1. This property underlies the operation of the optical fiber.

Multimode and Single-Mode Fibers

An optical fiber is a fine thread of glass composed of a core of higher index material in a cladding material, commonly silica. The core diameter is of the order of the wavelength of light, and light passing along it is incident at an angle greater than the critical angle, so that it reflects along the core as illustrated in Figure 11.8. However, not any ray incident at greater than the critical angle can propagate, for it is necessary that the phases of the reflected rays reinforce one another, so the phase differences must be multiples of 2π , illustrated in the figure. Consequently there are angles which allow rays to propagate, and each is a different *mode*. If the fiber size and index are chosen appropriately, there is only one mode allowed at a particular wavelength [10]. Such a fiber is called a *single-mode* fiber in the United States and a *monomode* fiber in Europe. *Multimode* fibers have core diameters much larger than the light wavelength.

Light escapes at a bend if the curvature is too large and/or there is only a small difference between the core and cladding indices. This is because the incident ray no longer meets the outside of the bend at greater than the critical angle, as shown in Figure 11.9; this effect causes light loss at microbends and kinks in the fiber.

Fibers are drawn from a composite preform. A preform is about 5 cm in diameter and 2 m long, and is held vertically at the top of a tower some 10 m

tall. The preform's lower tip is melted in an annular furnace and a fine thread drawn down; the thread has the same cross-sectional composition as the preform on a microscopic scale. The drawing speed determines the fiber diameter; typically, the fiber core is a few micrometers in diameter, and the cladding is of the order of 100 μm in diameter. The fiber is usually polymer coated for protection so that a complete fiber has a diameter of 250 μm .

The fiber behaves as if its core's refractive index is slightly different from that of the bulk material, and we can characterize the fiber as having an *effective index*, which may be dependent on both temperature and light wavelength. The effective length of an optical fiber is its actual length multiplied by its effective index.

Polarization

The *polarization* of an electromagnetic wave is defined as the direction in which the electric field propagates. A fiber drawn from a cylindrical preform has a circular cross section, so it is degenerate in polarization; there is no preferred axis for aligning polarization. But refractive index generally depends on the stress in the material; stress causes the core to act as if it had different refractive indices in different radial directions, called *stress birefringence*. Therefore an actual fiber coil may provide a preferred polarization, although it may not be stable over temperature and time.

A birefringent fiber can be considered as having different propagation constants β_x , β_y along two orthogonal axes x and y normal to the fiber axis, so that the waves traveling along x and y travel at different speeds, and linearly polarized light coupled into both x and y will become elliptically polarized as it moves along the fiber. When it has traveled a certain distance, the *beat length*, L , the phase between the x and y wave components will have changed by 2π and the wave will again be linearly polarized [11]:

$$|\beta_x - \beta_y|L = 2\pi$$

Another measure of fiber birefringence is the amount of power transferred between the cross-polarized states per unit length [12], denoted by the *h-parameter*. If linearly polarized light is launched into the x -axis of 1 km of a fiber with $h = 10^{-4} \text{ m}^{-1}$, 10% of the power will be in the y -axis mode at the other end.

Birefringent Fiber for a Sagnac Gyro

The accurate measurement (or nulling) of the small rotation induced Sagnac phase requires the rejection of the large phase (e.g., $10^6\pi$) common to the CW and CCW paths. There are two ways of ensuring this; either scramble the polarization so that it is random at the detector [12], or ensure that the paths

have identical optical lengths (identical β 's). Practically, that implies that they use the same polarization. Most gyro developers seem to be using the latter approach, employing special polarization maintaining fibers; Figure 11.9 illustrates some of the ways that such are made. The direct way is to make the core elliptical [13], although circular-section cores give the lowest losses. It is better to induce stress birefringence in the core by providing orthogonal asymmetry; successful methods include:

- (a) building-in an elliptical member of a different expansion glass (3M DWS Series),
- (b) building-in two rods, (*stress rods*) on either side of the core (York so-called "Bow-Tie" fiber) [14], and the PANDA fiber [15], for example, and
- (c) flattening the preform by squeezing it while it is hot (AT&T) [16], the so-called "square" fiber.

Square fiber has the useful property that the asymmetry axis is clearly visible from outside, and coils with controlled core position can be readily wound. Stress rods function by setting up an asymmetric stress in the core when the fiber cools down.

The Coherence of an Oscillator

We will find that the performance of optical gyros depends on the characteristics of the light source. In the case of the fiber-optic gyro we are particularly concerned with source *coherence*.

The wave generated by an oscillator is said to be coherent if its phase is continuous. A source can have degrees of coherence depending on the length of time over which the phase is continuous (or the distance the wave propagates with an unbroken wave form). The coherence length of the emission from a light source with a linewidth Γ (the full width at the half power point) is $L_c = c/\Gamma$; the narrower the line, the longer the coherence length.

The domestic electricity supply provides a coherent 60 Hz (in the United States) voltage, whereas the crackling heard in a radio receiver in a thunderstorm comes from the incoherent RF energy emitted by a lightning flash. The laser is a coherent source of light, electromagnetic energy at about 10^{14} Hz, whereas the light from an electric light bulb is incoherent, coming as it does from the random emission of photons from the heated filament.

Types of Optical Gyro

Optical gyroscopes may be characterized as either passive or active, as resonant or non-resonant. In passive sensors the Sagnac phase is measured by some

external means, whereas in active gyros the Sagnac phase causes a frequency change internal to the gyro which signals the rotation directly.

The Sagnac interferometer is the basis of the Interferometric Fiber-Optic Gyro (IFOG). The most widely used optical gyro is the active resonant Ring Laser Gyro (RLG), a very successful product which has dominated the 1-m/h aircraft navigator business since 1980. Two resonant passive gyros, the Resonant Fiber Optic Gyro (RFOG) and the Micro-Optic Gyro (MOG), are potentially lower cost instruments than the RLG. We will study the RLG in Chapter 13, and the two passive resonant gyros in Chapter 14.

Conclusion

Optical gyros have different operating characteristics than mechanical gyros. Instead of the mechanical noise created by ball bearings and the electrical noise from pickoffs, the optical gyros' noise floor is set by quantum effects (shot noise and spontaneous emission). The optical gyro's threshold is relatively high, compared with the stiction in the mechanical gyro, but the optical gyro's long term stability is an order of magnitude better. Rosenthal [17] and Anderson [18] provide good overviews of optical gyro technology.

References

1. Tipler, P.A., *Foundations of Modern Physics*, Worth Publishers, Inc. New York, 1973, Chapter 1.
2. Sagnac, G., "L'éther lumineux démontré par l'effet du vent relatif d'éther dans un interféromètre en rotation uniforme", C. R. Acad. Sci., 95, pp. 708-710, 1913.
3. Sagnac, G., "Sur la preuve de la réalité de l'éther lumineux par l'expérience de l'interférographe tournant", C. R. Acad. Sci., 95, pp. 1410-1413, 1913.
4. Hariharan, P. "Sagnac or Michelson-Sagnac interferometer?", Appl. Opt., 14, 10, pp. 2319-2321, October 1975.
5. Post, E.J., "Sagnac effect", Rev. Mod. Phys., 39, 2, pp. 475-493, April 67.
6. Arditty, H.J., H.C. Lefèvre, "Theoretical basis of Sagnac effect in fiber gyroscopes" in Ezekiel, S., H.J. Arditty (Eds.) *Fiber-Optic Rotation Sensors*, Springer-Verlag, New York, 1982, p. 44.
7. Yariv, A., *Introduction to Optical Electronics*, Holt, Rinehart, and Winston, Inc., New York, 1971, Chapter 7.
8. Fowles, G.R., *Introduction to Modern Optics*, Holt, Rinehart and Winston, Inc., New York, 1968. Fabry-Perot interferometer, p. 86 et seq.
9. Lawrence, A.W., "The Micro-Optic gyro", DGON Symposium Gyro Technology, Stuttgart, 1983.

10. Marcuse, D., *Theory of Dielectric Optical Waveguides*, Academic Press, New York, 1974. Chapter 1 uses a geometric optics approach.
11. Kaminow, I.P., "Polarization maintaining fibers", in Ezekiel, S., H.J. Arditty (Eds.) *Fiber-Optic Rotation Sensors*, Springer-Verlag, New York, 1982, pp. 169-177.
12. Ulrich, R., "Polarization and depolarization in the fiber-optic gyroscope", in Ezekiel, S., H.J. Arditty (Eds.) *Fiber-Optic Rotation Sensors*, Springer-Verlag, New York, 1982, pp. 52-77.
13. Dyott, R.B., "Elliptically cored polarization holding fiber", in Ezekiel, S., H.J. Arditty (Eds.) *Fiber-Optic Rotation Sensors*, Springer-Verlag, New York, 1982, pp. 178-184.
14. Birch, R.D., D.N. Payne, M.P. Varnham, E.J. Tarbox, "Fabrication and development of polarisation-maintaining fibres using gas phase etching", Proc. 1st International Conference on Optical Fibre Sensors, IEE, London, pp. 83-86, April 1983.
15. Himeno, K., Y. Kikuchi, N. Kawakamu, O. Fukuda, K. Inada, "A high extinction ratio and low loss single mode single polarization optical fiber", *Optical Fiber Sensors*, 1988 Technical Digest Series, Vol. 2, Optical Society of America, Washington, D.C., 1988.
16. Stolen, R.H., W. Pleibel, J.R. Simpson, "High-birefringence optical fibers by preform deformation", *J. Lightwave Tech.*, LT-2, 5, pp. 639-641, 1984.
17. Rosenthal, A.H., "Regenerative circulatory multiple-beam interferometry for the study of light propagation effects" *J. Opt. Soc. Am.*, 52, 10, pp. 1143-1148, Oct. 1962.
18. Anderson, D.Z., "Optical Gyroscopes", *Scientific American*, 254, 4, pp. 94-99, April 1986.

16

Design Choices for Inertial Instruments

In this chapter we will summarize the trade-offs between inertial systems designs and sensors, considering performance, cost, and reliability. The technology appropriate for a particular system depends on the performance and reliability needed and the size available; the cost will always need to be the lowest possible. We will also try to predict where the instrument design field is headed, allowing for the full impact of the Global Positioning System (GPS) satellite radio aid.

A Platform or a Strapdown System?

The performance needed will dictate the choice between the platform inertial measuring unit (IMU) and the strapdown type, and the nature of the mission will determine if radio or star-tracker aiding is permissible.

Platforms can have lower errors than strapdown systems. The null operation of their gyros eliminates anisoinertia and angular acceleration errors; spinning wheel gyros do not need to have high power torquers, whose dissipation causes thermal gradients. Missions demanding high accuracy, and providing a benign environment, might allow a gravity gradiometer to be used in a platform to compensate for gravity anomalies (Chapter 1); nuclear submarines typify this class. Platform accelerations can be measured with a Pendulous Integrating Gyro Accelerometer (PIGA, Chapter 7), currently the most precise accelerometer available, but the PIGA cannot be used in a strapdown system.

To summarize the other benefits of gimballed platforms:

- they can operate with vehicle rotation rates over 1000 deg/s,
- they self-align by gyrocompassing,
- they can calibrate the sensors by platform rotations,
- gyro torquer errors do not lead to attitude error.

But they are mechanically complicated, with their slip rings and bearings, synchros and torque motors. Accordingly, platforms are larger, more expensive, and have lower reliability than strapdown systems. The Advanced Inertial Reference System

(AIRS), used in the MX strategic missile, replaces the gimbal set with flotation; the gyros and accelerometers are assembled into a ball which is neutrally buoyant in a fluid (like the gimbal in floated gyros, Chapters 7 and 8). However, the complexity of the usual gimbals and bearings is equalled by that of the fluidic torque motors, fluid pumps, heat exchangers, etc., needed for its operation.

Conversely, strapdown systems are lighter, simpler, cheaper, easily configured for odd-shaped spaces, and due to the absence of moving parts, can have high reliability. Their deficiencies are:

- they have practical limits to maximum rate, in the 400-600 deg/s range,
- they are difficult to align,
- the sensors cannot easily be calibrated so must be stable,
- system rotation induces sensor errors (torquer asymmetry, non-linearity, anisoinertia, etc.),
- their accelerometer bias errors accumulate.

Aiding the IMU

If the IMU can incorporate a star-tracker or a radio aid, the integrated system will have better performance at lower cost. Strategic weapons systems must remain self-contained, perhaps aided with a star-tracker, for in a war serious enough for strategic weapons to be used one could not be sure that radio aids, like the constellation of GPS satellites, would survive. They could either be destroyed, or their signals jammed.

For commercial and tactical military use, the GPS aided guidance systems now being developed will change the nature of the navigation field. Their lower cost will expand automatic navigation, for high cost presently prevents its use in many places where it would be beneficial. These aided systems will invariably be strapdown systems, for GPS and strapdown systems provide complementary performance, and the two can be integrated in a number of ways [1]. In one scheme, a common processor can service both the strapdown and the GPS, and the IN system can tell the GPS receiver where to look for satellites and maintain navigation while satellites are jammed, obscured by buildings, or by passage through a tunnel.

In any case, the GPS receiver updates the IN system regularly, so that the INS can use instruments with lower day-to-day and in-run drift and relaxed scale factor. These lower performance gyros and accelerometers are expected to be considerably less expensive, by as much as an order of magnitude.

Choice of Sensor Type

In earlier chapters we described sensors by case studies of manufactured models. It is instructive to categorize them in other ways; one can distinguish, for

example, between sensors which use *differential design* and those which use resonance. Then again, one can choose between optical and mechanical gyros.

Differential Design

In the open-loop accelerometer or the single-axis rate gyro, the displacement of the sensitive element signals the input, and consequently the sensor suffers from cross-coupling and vibropendulous errors. We avoid this in the closed-loop instrument by operating at null; the pickoff signal is the difference between the acceleration or rate input and the feedback torque and is zero in the steady state.

The Vibrating Beam Accelerometer (VBA, Chapter 5), the Mach-Zehnder, Sagnac, and Fabry-Perot interferometers (Chapter 11) all exhibit differential operation. Oscillators are connected in push-pull, or paths are common and change length opposite to the inertial stimulus. This symmetry or reciprocity provides high rejection of the common mode signal (the unloaded tuning fork frequency, the total path phase) so that the sensor has high resolution. For example, the CW and CCW frequencies of a Ring Laser Gyro (RLG, Chapter 13) are of the order of 10^{14} Hz, yet their difference can be read to 10^{-3} Hz.

Using Resonance

Uncontrolled resonance of mechanical parts can be destructive to systems and sensors. The Dynamically Tuned Gyro's rotor (DTG, Chapter 9) is suspended on elastic flexures on a shaft, in a very low damping environment. Under external vibration at the resonance of the rotor mass M_r on the net flexure bending stiffness K_b , $\omega = (K_b/M_r)^{1/2}$, the flexures stress can exceed the elastic limit. Drift errors from g and g^2 terms are higher at resonance. Clearly, the natural frequency will be designed to be higher than the environmental frequencies; anti-vibration mounts may even be used to support the gyro.

But resonance is used constructively in the DTG and other sensors. In the DTG, the gyroscopic torque from gimbal angular oscillation (the dynamic anti-spring) cancels the positive torque of the flexure twist at the tuned speed. The tuning forks of the VBA and the Quartz Rate Sensor (QRS, Chapter 10), and the resonator of the Hemispherical Resonator Gyro (HRG, Chapter 10) operate at a resonance so that the amplitude of motion is large; the high Q-factor provides high resolution of the output frequency.

The RLG and the other resonant optical sensors, Resonant Fiber Optic Gyro and Micro Optic Gyro (RFOG and MOG, Chapters 13 and 14) accumulate Sagnac phase because resonance causes multiple circuits of the resonator perimeter; the finesse (like the Q-factor) expresses the sharpness of resonance and the resolution. The non-resonant Interferometric Fiber Optic Gyro (Chapter 12) achieves the same resolution by using a long, coiled, fiber path rather than resonance.

Mechanical or Optical Gyros?

The differences between optical and mechanical gyros affect their selection for particular applications. Mechanical gyros provide lower short term noise than optical gyros (but worse long term stability), they have inertial memory and can offer longer life.

Performance: Mechanical gyros with wheels, and the vibrating HRG, have low short term noise, and gas bearing single degree of freedom gyros (SDFG, Chapter 7) have demonstrated output noise as low as a few nanoradians. They are preferable for precise pointing for short times, for telescopes and laser designators.

But optical gyros have no *g*-sensitivity, can measure high rates without an increase in power, have better day-to-day drift stability, and have stable, linear scale factor.

Inertial Memory: Both wheeled and vibrating gyros have inertial memory. If power is lost for a short time (tens of seconds), the sensitive element integrates the input motions so that when power returns, the gyro signals the angle turned in the interval. For gimballed wheels (either SDFG or DTG) this angle must not be so large that the gimbal or rotor hits a limit stop, practically 1-3°. Nor, for the HRG, can the rotation be so large that one or more nodes have passed the pickoff (nodes being indistinguishable from one another).

Optical gyros do not have this memory feature; once the power is off, the gyro loses reference. Optical gyros can also lose memory under nuclear radiation, either from super-radiance in the laser or from blackening of a fiber, the first being short term and the second having a long term effect.

Lifetime: Gas bearing SDFGs have very long lifetimes, particularly if the wheel is kept running, for then there is no material contact in the bearing and no wearing. The RLG depends on the correct amount of helium for its operation, and helium is noted for its ability to diffuse through many materials, which can limit RLG life. The solid-state optical gyros (IFOG, RFOG, MOG) have no such limitation. For the vibrating sensors, the QRS lifetime should be unlimited, although the performance of the HRG depends on maintaining a hard vacuum around the resonator. Gettering can remove outgassed vapors in order to extend life.

Reliability

Reliability is the probability that a system will perform for a given time under given conditions. The system is assumed to have been randomly chosen from a group of identical systems; reliability is a characteristic of the group, not of the individual. The reliability of inertial sensors is often estimated by *similarity* - we have numbers on this old model and the new one looks similar therefore its reliability must be the same - replaced by test data as early as possible.

Experience has shown that some systems tend to fail early, so to weed out early failures it is common to *burn-in* systems for a week or so.

Things often have a finite life. Some gracefully degrade, others catastrophically fail. Bearings and sliprings wear, gas laser gyro cathodes may sputter away, and a DTG flexure or an accelerometer hinge might break. These failures might call for maintenance (the process of returning the system as closely as possible to its as-new condition), if the system was designed to be repairable. When it is not economically feasible to repair a system, it has *worn out*. The mean life of a population of systems is the average time to wearout, and the *Useful Life* of a system is the period between burn-in and the onset of wearout.

If we can repair a failed device such that the repair does not introduce a new failure potential, then the average time that one can expect the device to work before the next failure is the *Mean Time Between Failures* (MTBF). In a test of a large sample the MTBF is the total operating time of all the samples (not just the failed ones) divided by the number of failures. After a long time this test should converge on a constant MTBF. The ratio of MTBF to mission time is the probability that the device will allow the mission to be completed. Statistically, 63% of a population of a device will fail in an operating time equal to the MTBF. *Failure rate* is the reciprocal of the MTBF and is usually expressed in failures per 10^6 hours.

Redundancy

In order to ensure a successful mission, any single system which could cause the mission to fail must have an adequate MTBF. If it does not, or if the consequence of any failure is unacceptable, backup systems will provide *redundancy*. Redundancy can be provided by installing a few independent, identical copies of a system and comparing their outputs. Inertial navigation systems are often used in threes, and a computer correlates their outputs (as well as monitoring obvious failures like a blown fuse). If two of the three consistently differ from the third, the third is considered to have failed and its data are ignored. Deciding between two might be possible if other data from separate instruments are available.

Redundant systems are expensive, and the cost of providing backup systems must include the cost of carrying their extra weight. Rather than installing three copies of a system, redundancy can be provided at the component level. Particular system designs include one in which a fourth single-axis sensor is mounted with its sensitive axis skewed to the three basic sensor axes. This sensor can then act as a check on the basic three and can provide information if one should be determined to have failed.

Table 16.1. Sensor selection factors - performance.

Operating life, MTBF	OA acceleration sensitivity
Activation time	1N and 2N sensitivities
Maximum rate or acceleration	Transfer characteristics:
Hysteresis	bandwidth
Scale factor: stability, long term	damping
linearity and asymmetry	Thermal: temperature sensitivities
Anisoinertia	maximum operating temperature
Axis alignment stability	minimum operating temperature
Cross-coupling	gradients
Threshold and resolution	shocks
Bias, day-to-day repeatability	Magnetic field sensitivity
in-run random	Resonant frequencies
random walk	Power consumption
Noise spectrum	unusual voltages/frequencies
g-sensitivity	Size and mass
Anisoelasticity (g^2 terms)	Self-test
Vibropendulosity	

Sensor Design Check Lists

Much of the information in this book can be summarized into check lists, and Tables 16.1 and 16.2 list many of the performance and environmental factors to be considered in applying inertial sensors. Table 16.3 shows how different performance features vary with gyro type.

Table 16.2. Sensor selection factors - environmental.

Storage life	Shock: wave shape
Temperature extremes	peak value
Linear acceleration, level	duration
duration	(e.g., 50g, half-sine, 15 ms)
directions	Shock mount: transmissibility
Vibration spectra: sine sweep	resonances
frequency range	axis alignment degradation
r.m.s. level	Air pressure/vacuum
random, g^2 /Hz profile	Ionizing radiation
Non-operating resonances (fatigue)	

Table 16.3. Performance dependencies of some gyros.

<i>Feature</i>	<i>SDFG</i>	<i>DTG</i>	<i>FOG/MOG</i>	<i>RLG</i>
Maximum rate	Torquer heat dissipation		Not instrument limited, to first order	
Scale factor stability	Permanent magnet ageing		Wavelength stability	
Scale factor linearity	Quality of magnetic design		Quality of feedback system (serrodyne)	Lock-in rate and means of suppression
Bias stability (long term)	Flexleads Magnetic leakages	Time const. Mistuning Magnetic leakage Bearing noise		Backscatter and mode control
Random drift (short term bias)	Fluid OA bearings Spin bearings	Spin bearings Gas drag pickoff stability	Quantum shot noise	Spontaneous emission Mechanical dither Quantization
<i>g</i> -sensitivity	Gimbal mass stability	Rotor mass stability Flexure quadrature torque	None to the first order	
IA stability	OA bearing Pickoff null Case mounting flange	Pickoff nulls Case mounting flange	FOG coil former MOG chip mounting	Dither motor

Conclusions

There are few choices for inertial sensors for the ultimate performance, with better than 1 mile/day system drift. The field is limited to platforms carrying the SDFG gas bearing model, the Electrically Suspended Gyro (ESG, Chapter 8), the continuously rotated RLG, and the PIGA. Gyro technology does not presently offer new technology for this market; the HRG is close, but the FOGs do not have this potential.

But for low accuracy systems, sensor choices are not clear-cut. The least expensive inertial sensors are typified by the quartz VBA and the QRS tuning

fork gyro, and the silicon-based Micromachined Silicon Accelerometer (MSA) and MOG, all too new to have pedigree. Their selection implies some risk.

For new medium accuracy GPS/IN systems, Heading and Attitude Reference Systems and the like, the IFOG, the RFOG, and the VBA should be chosen; for despite the pedigree and extensive installed base of the SDFG and DTG, they are less attractive because of their high rate power dissipation, lower reliability (from the ball bearings), and higher cost.

References

1. Johannessen, R., M.J.A. Asbury, "Towards a quantitative assessment of benefits INS/GPS integration can offer to civil aviation", *Navigation*, 37, 4, pp. 329-346, Winter 1990-91.